DIRECT AND EM-BASED MAP SEQUENCE ESTIMATION WITH UNKNOWN TIME-VARYING CHANNELS

H. Chen, R. Perry, and K. Buckley

ECE Department, Villanova University, Villanova, PA 19085
buckley(hchen,perry)@ece.villanova.edu
(610)-519-5658, FAX: (610)-519-4436

ABSTRACT

In this paper we address sequence estimation when the InterSymbol Interference (ISI) communication channel is unknown and time varying. We employ a Maximum A Posterior (MAP) approach, in which the unknown channel parameters are assigned a distribution and integrated out. For several channel models of interest we describe both the exact MAP estimator and Viterbi algorithm based implementations. We also present EM algorithms for solving these MAP sequence estimation problems, and we contrast these EM solutions with direct MAP algorithms.

1. INTRODUCTION

Maximum Likelihood Sequence Estimation (MLSE) is widely used in digital communication systems to estimate the transmitted data sequence observed over noisy ISI channels. For an unknown channel, many algorithms have been developed to estimate the sequence and/or identify the channel blindly. In this paper we focus on direct sequence estimation for unknown, fast time-varying channels.

Per-Survivor Processing (PSP) [1] has been proposed for MLSE with fast time-varying channels. In [2], we describe an approach to MLSE based on probabilistic modeling of the fast time-varying channel. We formulate the MAP sequence estimator, assign a prior distribution to the parameters of a time-varying channel model, and marginalize over the channel parameters. For temporally independent Gaussian channels, an optimum Viterbi algorithm is derived. For Gauss-Markov channels, a PSP approach based on the generalized Viterbi algorithm (GVA) [3] is described which retains a fixed number of survivors (L) per trellis state. We show that the proposed method approaches that of ML exhaustive search for reasonably small values of L.

In [4], we employ the EM algorithmic approach to solve the MAP problem described in [2]. This approach is similar to other EM based approaches to direct sequence estimation (e.g. [5, 6, 7, 8]), but is unique in its focus on exploiting prior information on fast time-varying channel models to improve performance. The EM method for iterative solution of Maximum Likelihood (ML) and MAP estimation problems for continuous parameters was introduced into the signal processing community by Dempster, et al. [9]. With EM, convergence and initialization are important considerations. Wu [10] proved convergence to a local minimum or saddle point of the negative log likelihood function for the continuous parameter case. This holds true for the MAP problem posterior density [11]. For the problem of discrete parameter (e.g. sequence) estimation, convergence is less understood. In [11] it is claimed that for certain discrete parameter estimation EM problems, including the one addressed in [4], EM is guaranteed to converge to a stationary point of the posterior density. In this paper we show that this claim is not true, and that EM is not guaranteed to converge to a stationary point. Nonetheless, numerous researchers have shown that EM can be used effectively for sequence estimation.

In this paper we describe and contrast direct and EM iterative algorithms for MAP sequence estimation with unknown, fast time-varying channels.

2. DATA MODEL

In the discrete-time FIR model of a time-varying noisy communications channel with inter-symbol interference, for the received data sequence up to time n, the received data r_k at time k is given by

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

where $\mathbf{a}_k^T$ is a complex row vector containing transmitted data $\{a_{k-i+1}, i = 1, \ldots, M\}$, M is the FIR channel length, $\mathbf{h}_k$ is a complex column vector containing the channel impulse response coefficients, and n_k is the white Gaussian complex noise with variance σ^2. Let $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$ represent the matrix of channel coefficient vectors over time, and let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]^T$ represent the matrix of transmitted data. Also let $\mathbf{r} = [r_1, \ldots, r_n]^T$ and $\mathbf{n} = [n_1, \ldots, n_n]^T$.

THIS RESEARCH WAS SUPPORTED BY ONR UNDER GRANT N00014-98-1-0892.

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

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

To minimize BER, we find $\mathbf{A}$ to maximize $f(\mathbf{A}|\mathbf{r})$ (i.e. the MAP estimator). This is equivalent to 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 $f(\mathbf{r}|\mathbf{A})$ is used. Referring to (2), if the channel is known, the Viterbi algorithm can be used directly to estimate the data sequence. If the channel is unknown, the dependence of $f(\mathbf{r}|\mathbf{A})$ on the random channel coefficients is

$$f(\mathbf{r}|\mathbf{A}) = E[f(\mathbf{r}|\mathbf{H},\mathbf{A})] = \int_{\mathbf{H}} f(\mathbf{r}|\mathbf{H},\mathbf{A})f(\mathbf{H})d\mathbf{H} \quad . \quad (3)$$

3. MAP SEQUENCE ESTIMATION

Here we summarize results from [2]. For channel coefficients which are independent over time, $f(\mathbf{H})$ can be expressed as

$$f(\mathbf{H}) = \prod_{k=1}^{n} f(\mathbf{h}_k). \quad (4)$$

If $f(\mathbf{h}_k)$ is Gaussian with mean $\mathbf{d}_k$ and covariance $\mathbf{C}_k$, then

$$f(\mathbf{h}_k) = \frac{1}{\pi^M|\mathbf{C}_k|}e^{-(\mathbf{h}_k-\mathbf{d}_k)^{*T}\mathbf{C}_k^{-1}(\mathbf{h}_k-\mathbf{d}_k)}, \quad (5)$$

where $*$ denotes conjugate. We assume that $\mathbf{d}_k$ and $\mathbf{C}_k$ are known. Substituting (2) and (4) into (3), taking the negative natural logarithm of the integration result and ignoring the terms that are irrelevant to the minimization, we obtain:

$$-\log(f(\mathbf{r}|\mathbf{A})) \doteq \sum_{k=1}^{n}\frac{|r_k-\mathbf{a}_k^T\mathbf{d}_k|^2}{\sigma_k^2} + \log(\sigma_k^2)\ , \quad (6)$$

where

$$\sigma_k^2 = \sigma^2 + \mathbf{a}_k^T\mathbf{C}_k\mathbf{a}_k^* \quad (7)$$

and $\doteq$ denotes equivalent for optimization purposes. The time-recursive form for (6) is obvious, and since the incremental cost only depends on the state transition at the current time, the Viterbi algorithm can be used directly as an efficient, exact MAP (optimum) algorithm.

For a fast time-varying channel with Gauss-Markov fading parameters, the optimum solution is derived in [2] where a computationally effective suboptimum algorithm based on PSP and GVA is also proposed.

Concerning PSP, for each trellis survivor path the quantities involved in the transition cost are computed by channel model aided estimation as dictated by the MAP formulation. As per GVA, at each state of each stage of the trellis, we keep $L \geq 1$ survivors. We show that for reasonable L the algorithm approaches optimum results.

4. EM ALGORITHMS

Here we summarize results from [4] and discuss EM convergence for sequence estimation. To maximize (3) using EM, define the auxiliary function [9]:

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

where we desire the MAP estimate of $\mathbf{A}$, and $\mathbf{B}$ represents the current estimate of $\mathbf{A}$. In EM terminology, $\mathbf{r}$ 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: (i) initialize $\mathbf{B}$ to an estimate of $\mathbf{A}$; (ii) the E-step (Expectation) (i.e. construct $Q(\mathbf{A}|\mathbf{B})$); (iii) the M-step (Maximization) (i.e. maximize $Q(\mathbf{A}|\mathbf{B})$ with respect to $\mathbf{A}$); and (iv) set $\mathbf{B} = \mathbf{A}$ and repeat steps (ii) and (iii) until convergence.

$Q(\mathbf{A}|\mathbf{B})$ may be written as

$$Q(\mathbf{A}|\mathbf{B}) \doteq \int_{\mathbf{H}} \log(f(\mathbf{r}|\mathbf{H},\mathbf{A}))f(\mathbf{r}|\mathbf{H},\mathbf{B})f(\mathbf{H})d\mathbf{H}. \quad (9)$$

Note that in the above steps, if $Q(\mathbf{A}|\mathbf{B}) \geq Q(\mathbf{B}|\mathbf{B})$ then according to Jensen's inequality, $f(\mathbf{r}|\mathbf{A}) \geq \mathbf{f}(\mathbf{r}|\mathbf{B})$ [9]. However, this does not prove convergence of EM to a local maximum of the log likelihood function. For discrete parameter estimation, we will address this issue below.

Using (2) in (9) and dropping constants:

$$Q(\mathbf{A}|\mathbf{B}) \doteq \int_{\mathbf{H}} -\sum_{k=1}^{N}|r_k-\mathbf{a}_k^T\mathbf{h}_k|^2\ f(\mathbf{H}|\mathbf{r},\mathbf{B})d\mathbf{H}, \quad (10)$$

$$f(\mathbf{H}|\mathbf{r},\mathbf{B}) \doteq\ f(\mathbf{H})\ \prod_{k=1}^{N} e^{-\frac{|r_k-\mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (11)$$

If we assume that the elements of $\mathbf{H}$ are independent over time, then $f(\mathbf{H})$ can be written as in (4), and $f(\mathbf{H}|\mathbf{r},\mathbf{B})$ becomes

$$f(\mathbf{H}|\mathbf{r},\mathbf{B}) \doteq\ \prod_{k=1}^{n} f(\mathbf{h}_k)\ e^{-\frac{|r_k-\mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (12)$$

Using (12) in (10) we obtain

$$Q(\mathbf{A}|\mathbf{B}) \doteq -\sum_{k=1}^{n} V_k, \quad (13)$$

$$\begin{aligned} V_k &\triangleq |r_k-\mathbf{a}_k^T\mathbf{g}_k|^2 + \mathbf{a}_k^T\mathbf{G}_k\mathbf{a}_k^* \quad (14) \\ \mathbf{g}_k &\triangleq E[\mathbf{h}_k|\mathbf{r},\mathbf{B}] \quad (15) \\ \mathbf{G}_k &\triangleq E[(\mathbf{h}_k-\mathbf{g}_k)(\mathbf{h}_k-\mathbf{g}_k)^{*T}|\mathbf{r},\mathbf{B}]. \quad (16) \end{aligned}$$

V_k represents the Viterbi algorithm incremental cost. In [4] we provide a detailed derivation for both this independent Gaussian channel, and a more realistic Gauss-Markov channel model.

A Note on EM Convergence: For discrete or continuous parameter estimation, it was pointed out above that Jensen's inequality holds, proving that at each iteration of the EM algorithm the negative log likelihood function is not increased. For continuous parameter estimation, it is well known the the EM algorithm converges to a minimum or saddle point of the negative log likelihood function (for a proof, see [10]). This proof is based on the global convergence theorem [12], which is now stated.

Let $\mathcal{E}$ denote an algorithm on $\mathcal{A}$ which, given an initial estimate $\mathbf{A}_0$ generates a sequence $\{\mathbf{A}_k;\ k = 1, 2, ...\}$ as $\mathbf{A}_{k+1} \in \mathcal{E}(\mathbf{A}_k)$. Let $\Gamma \subset \mathcal{A}$ be the solution set. Under the conditions:

1. all points $\mathbf{A}_k$ are contained in a compact set $\mathcal{S} \subset \mathcal{A}$;
2. there is a continuous function Z on $\mathcal{A}$ such that
   (a) if $\mathbf{A_k} \notin \Gamma$, then $Z(\mathbf{A}_{k+1}) < Z(\mathbf{A}_k)$ for all $\mathbf{A}_{k+1} \in \mathcal{E}(\mathbf{A}_k)$
   (b) if $\mathbf{A_k} \in \Gamma$, then $Z(\mathbf{A}_{k+1}) \leq Z(\mathbf{A}_k)$ for all $\mathbf{A}_{k+1} \in \mathcal{E}(\mathbf{A}_k)$:
3. the mapping $\mathcal{E}$ is closed at points outside Γ,

the limit of any convergent subsequence of $\{\mathbf{A}_k\}$ is a solution in Γ.

Let $\{\mathbf{A}_k\} \in \mathcal{A}$ denote a sequence in the discrete space $\mathcal{A}$ of possible sequence estimates. Let $\mathcal{E}$ represent the EM algorithm, and $Z(\mathbf{A}_k) = -\log f(\mathbf{r}|\mathbf{A}_k)$. For discrete parameter estimation, condition 1 does not hold since there is not a compact subset $\mathcal{S} \subset \mathcal{A}$, since $\mathcal{A}$ is a discrete set. As illustrated below, condition 2(a) also does not hold. This is because $\mathbf{A}_k \in \mathcal{E}(\mathbf{A}_k)$ is possible when $\mathbf{A_k} \notin \Gamma$. Thus, the global convergence theorem is not generally applicable to the EM algorithm for discrete parameter estimation.

We now show by example that EM does not necessarily converge to a local discrete minimum of the negative log likelihood function. For the independent Gaussian channel model, let $n = 1$ and $M = 3$, which are small enough to completely analyze by hand to confirm the results. A single BPSK symbol a_1 at time 1 is to be estimated as either +1 or –1. The channel is initialized as $\mathbf{a}_1 = [a_1\ a_0\ a_{-1}]^T = [a_1$ -1 -1$]^T$. For the Gaussian model for the channel coefficients we assume mean $\mathbf{d}_1 = [1\ 1\ 1]^T$ and covariance $\mathbf{C}_1 = v\mathbf{I}$ with $v = 0.05$. The noise variance is $\sigma^2 = 0.3$. The negative log-likelihood cost function of the received data r_1 given $\mathbf{a}_1$ is obtained from (6) and (7). Figure 1(d) illustrates the MAP cost function for $r_1 = -1.8$, which is in the range of r_1 where EM breaks down, as discussed below.

For the EM algorithm, letting $\mathbf{b}_1$ represent an estimate of $\mathbf{a}_1$, the conditional mean and covariance of the channel coefficients given the received data are

$$\mathbf{g}_1 = \mathbf{d}_1 + \frac{\mathbf{C}_1\mathbf{b}_1(r_1 - \mathbf{b}_1^T\mathbf{d}_1)}{\sigma_1^2} \qquad (17)$$

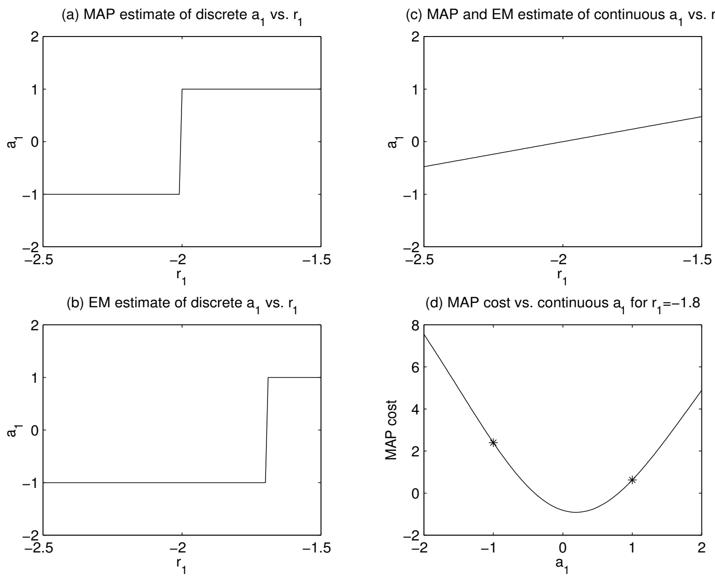

Fig. 1. Estimates of a_1 vs. r_1: (a) Discrete MAP, (b) Discrete EM, (c) Continuous MAP and EM; (d) MAP cost vs. a_1.

$$\mathbf{G}_1 = \mathbf{C}_1 - \frac{\mathbf{C}_1\mathbf{b}_1\mathbf{b}_1^T\mathbf{C}_1}{\sigma_1^2}, \qquad (18)$$

with $\sigma_1^2 = \sigma^2 + \mathbf{b}_1^T\mathbf{C}_1\mathbf{b}_1$. The EM cost function is

$$V_1 = -Q(\mathbf{a}_1|\mathbf{b}_1) \doteq |r_1 - \mathbf{a}_1^T\mathbf{g}_1|^2 + \mathbf{a}_1^T\mathbf{G}_1\mathbf{a}_1. \qquad (19)$$

The EM algorithm was initialized using $\mathbf{b}_1 = [$-1 -1 -1$]^T$. Figures 1(a)-(b) show the MAP and EM discrete estimates of a_1 vs. received data r_1. Note the range of values $-2 \leq r_1 \leq -1.7$ over which the discrete EM estimate is "wrong" in Figure 1(b), i.e. it is not only not equal to the MAP solution, it is not even a local minimum or stationary point of the negative log-likelihood function as can be seen in Figure 1(d). Over the range $-2 \leq r_1 \leq -1.7$, the EM algorithm does not iterate, it is converged after initialization. As an aside, Figure 1(c) shows the EM and MAP estimates (which are identical) for a_1 estimated as a continuous parameter.

This example illustrates that for discrete parameter estimation EM can get stuck at a solution $\mathbf{A}$ which is not a local discrete minimum of the negative log likelihood function. By this we mean that there may be an estimate $\mathbf{A}^{'}$ that differs from the EM solution $\mathbf{A}$ in only one element and that has a lower negative log likelihood cost. We have observed this phenomenon for EM algorithm solutions to several communications and tracking related discrete parameter estimation problems. Nonetheless, researchers have found EM to be useful for discrete parameter estimation. The point here is that care must be taken in employing EM for discrete parameter estimation since convergence to a minimum of the cost can not be guaranteed.

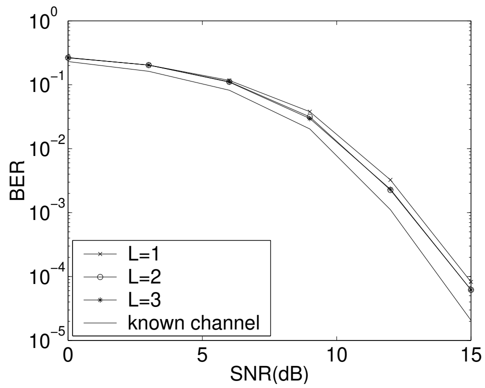

Fig. 2. BER for Gauss-Markov PSP

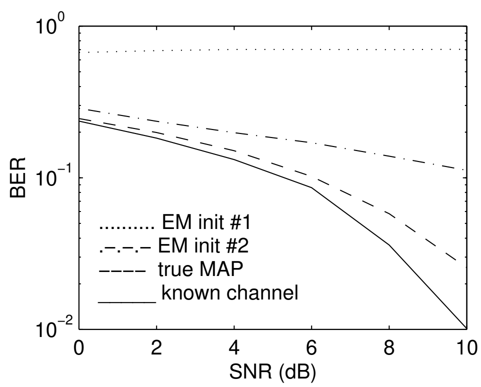

Fig. 3. BER for different Gauss-Markov EM initializations

5. NUMERICAL RESULTS

Figures 2 and 3 show typical results from Monte-Carlo simulations using Gauss-Markov channel coefficients. Figure 2 shows the direct MAP PSP GVA algorithm for a Gauss-Markov channel with zero mean and covariance $\mathbf{C} = v\mathbf{I}$ with $v = 0.001$. This illustrates that even for small L near optimum results can be obtained. Figure 3 is a comparison of EM and direct MAP, using parameters $n = 10$, $M = 3$, $\alpha = 0.995$, $v = 0.01$, $\mathbf{d} = \mathbf{0}$, and $\mathbf{C} = v\ \mathbf{I}$. The direct MAP BER results were obtained by an exhaustive search of the ML cost function. Two different initializations for the EM algorithm were used, both of which produce unreasonably large BER. Initialization 1 (init #1) used known $\mathbf{h}_1$ and $\mathbf{h}_k = \alpha\mathbf{h}_{k-1},\ k = 2\ldots n$. Initialization 2 used known $\mathbf{h}_1$ and 0's for the initial symbol estimates. Other researchers have found that EM initialization can be improved using pilot bits. We have performed many simulations which confirm that EM with proper initialization can provide MAP results.

6. SUMMARY

We describe MAP, approximate (PSP based) MAP and EM algorithms for sequence estimation for fast time-varying channels. Comparing EM with direct MAP, EM offers computational advantages. However, with EM, since the parameters to be estimated are discrete valued, extra care must be taken in initializing the algorithm.

7. REFERENCES

[1] 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, no. 2/3/4, pp. 354–364, Feb/Mar/Apr 1995.

[2] H. Chen, K. Buckley, and R. Perry, "Time-recursive maximum likelihood based sequence estimation for unknown ISI channels," *Proc. 33-rd Asilomar Conference on Circuits, Systems and Computers*, 2000.

[3] N. Seshadri and C. Sundberg, "List Viterbi decoding algorithms with applications," *IEEE Transactions on Communications*, pp. 313–323, February/March/April 1994.

[4] R. Perry, W. Berger, and K. Buckley, "EM algorithm for sequence estimation over Gauss–Markov ISI channels," *ICC 2000 Conference, New Orleans, Louisiana*, June 2000.

[5] C. N. Geroghiades and D. L. Snyder, "The expectation-maximization algorithm for symbol unsynchronized sequence detection," *IEEE Trans. on Communications*, vol. 39, no. 1, pp. 54–61, Jan 1991.

[6] Laurie B. Nelson and H. Vincent Poor, "Iterative multiuser receivers for CDMA channels: An EM-based approach," *IEEE Trans. on Communications*, vol. 44, no. 12, pp. 1700–1710, Dec 1996.

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

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

[9] 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.

[10] C. F. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983.

[11] A. Logothetis and V. Krishnamurthy, "Expectation maximization algorithms for map estimation of jump markov linear systems," *IEEE Transactions on Signal Processing*, August 1999.

[12] D. G. Luenberger, *Linear and Nonlinear Programming, 2nd ed.*, Addision-Wesley, 1984.