Time-Recursive Maximum Likelihood Based Sequence Estimation for Unknown ISI Channels*

Hai Chen Kevin Buckley Richard Perry

Department of Electrical and Computer Engineering
Villanova University
Villanova, PA 19085
E-mail: hchen(buckley,perry)@ece.villanova.edu

Abstract

In this paper MLSE (maximum likelihood sequence estimation) is applied to unknown ISI (inter-symbol interference) fast time-varying channels. Assuming a probabilistic model for an FIR channel, a Maximum A-Posterior (MAP) or ML cost function for the communication symbols is obtained by integration over the channel FIR parameters. A time-recursive (i.e. on-line) sequence estimation algorithm is then identified.

1. Introduction

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 identify the channel blindly. Some assume the channel is time-invariant and are not applicable for fast time-varying channels (e.g. [1, 7]), while others are applicable for channels with fast temporal variation (e.g. Expectation-Maximization (EM) based [4, 10]) but are block processing (off-line) and/or sensitive to initialization. Many adaptive MLSEs have also been developed. Some of these algorithms have difficulty in tracking fast time-varying ISI channels due to: tentative decision delay[3]; or error propagation introduced by a decision-feedback equalizer (DFE) embedded in the MLSE structure [5].

Raheli et al. proposed PSP (Per-Survivor Processing) -based MLSE [6] for tracking fast time-varying channels. PSP has been applied for sequence estimation and channel identification successfully. Adaptive algorithms such as LMS and RLS can be combined with PSP to identify the channel and further estimate the data sequence. A suboptimum PSP adaptive MLSE algorithm was proposed in [2], in which the adaptive MLSE consists of a channel estimator and an MLSE implemented by the Viterbi algorithm, and channel estimation is accomplished for each state in the Viterbi algorithm (VA) along the survivor path associated with each state. LMS and RLS can be applied for channel estimation. This adaptive MLSE has a capability of good tracking performance for fast time-varying ISI channels. However, the choice of the parameters of LMS and RLS in channel identification may still affect the performance of this method.

In order to reduce the performance degradation caused by the channel estimation, this paper introduces a channel statistics-aided MLSE method to estimate the transmitted data sequence on-line. This method marginalizes over the unknown channel coefficient distribution and thus obtains the probability of the received signal given the transmitted signal, which is a function of the given transmitted signal. Then the transmitted signal that maximizes the probability is obtained from this function. The Viterbi and PSP Viterbi algorithms are used to obtain optimum and suboptimum results, respectively. For a temporally independent Gaussian channel, an optimum Viterbi algorithm is derived. For a Gauss-Markov channel, we will use PSP and the generalized Viterbi algorithm (GVA), which retains a fixed number of survivors (L) per state instead of retaining one survivor trellis path per state as in the Viterbi algorithm. It is shown that the proposed method approaches that of ML exhaustive search for reasonably small values of L. This new algorithm is suitable for fast time-varying channels. Simulations prove the advantage of this algorithm over PSP LMS and RLS algorithms.

This paper is organized as follows. In section 2, the system model is introduced and the problem is formulated. Our proposed algorithm is presented in section 3 and section 4 for time-independent Gaussian channel and Gauss-Markov channel respectively. Section 4 will also present an adaptive PSP GVA to reduce the computational complexity. The

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

simulation results are shown in section 5. This section also compares our algorithm with PSP LMS/RLS. Finally, we present our conclusions in section 6.

2. System 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 complex 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 at time k, and n_k is the white Gaussian complex noise at time k with variance σ^2. Let $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$ represent the matrix of channel coefficient vectors over time arranged by columns, and let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]^T$ represent the matrix of transmitted data arranged by rows. Also let $\mathbf{r} = [r_1, \ldots, r_n]^T$ represent the column vector of received data, and $\mathbf{n} = [n_1, \ldots, n_n]^T$ represent the column vector of noise over time. With this notation, the probability density function of the received data, given $\mathbf{H}$ and $\mathbf{A}$, is

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

To minimize BER, we must 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 $f(\mathbf{r}|\mathbf{A})$ can be used instead. 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

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

Evaluating this multidimensional integral and then maximizing the result with respect to $\mathbf{A}$ seems to be an intractable problem in general. However, if the statistics of the channel coefficients can be assumed, it is shown below that in some cases this problem is solvable and that the Viterbi algorithm or GVA can be applied directly to obtain the optimum or suboptimum results. In the following two sections, we will show how the problem is solved over time-independent Gaussian channels and Gauss-Markov channels.

3. Time-Independent Gaussian channel

For channel coefficients independent over time, $f(\mathbf{H})$ can be expressed as

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

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

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

where $*$ denotes conjugate, M is the FIR channel length and 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:

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

where

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

($\doteq$ denotes equivalent for optimization purposes).

The time-recursive form for (6) can be expressed as

$$\Lambda_n = \Lambda_{n-1} + \lambda_n \;, \tag{8}$$

where Λ_n is the cumulative cost up to time n. From (6), λ_n is the incremental cost from time $n-1$ to n:

$$\lambda_n = \frac{|r_n - \mathbf{a}_n^T \mathbf{d}_n|^2}{\sigma_n^2} + \log(\sigma_n^2) \;. \tag{9}$$

Since the incremental cost only depends on the state transition at time n, the Viterbi algorithm can be used directly to minimize (8). Therefore, (8) represents a computationally efficient, time-recursive, exact MAP (optimum) algorithm.

4. Gauss-Markov channel

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 [9]. Let the time-varying channel coefficients follow a Gauss-Markov distribution such that, at time k, $\mathbf{h}_k = \alpha \mathbf{h}_{k-1} + \mathbf{v}_k$, 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 $\mathbf{C}$. $f(\mathbf{H})$ can be written as

$$f(\mathbf{H}) = f(\mathbf{h}_1) \prod_{k=2}^{n} f(\mathbf{h}_k|\mathbf{h}_{k-1}). \tag{10}$$

We assume the mean of $\mathbf{v}_k$ is zero and the initial channel coefficient is known and it is represented as $\mathbf{h}_0$. So

$$\mathbf{h}_1 = \alpha\mathbf{h}_0 + \mathbf{v}_1. \tag{11}$$

Substituting (2), (10) and (11) into (3), we obtain

$$\begin{aligned} f(\mathbf{r}|\mathbf{A}) &= E[f(\mathbf{r}|\mathbf{H},\mathbf{A})] \\ &= \int_{\mathbf{H}} f(\mathbf{r}|\mathbf{H},\mathbf{A}) f(\mathbf{H}) d\mathbf{H} \\ &\doteq \int_{\mathbf{H}} \left(\prod_{k=1}^{n} e^{-\frac{|r_k - \mathbf{a}_k^T\mathbf{h}_k|^2}{\sigma^2}}\right) f(\mathbf{h}_1) \cdot \\ & \quad \prod_{k=2}^{n} f(\mathbf{h}_k|\mathbf{h}_{k-1}) d\mathbf{H} \end{aligned} \tag{12}$$

We can marginalize over the channel coefficient distribution step by step. Integrating over $\mathbf{h}_1$, we obtain

$$I_1 = \int_{\mathbf{h}_1} e^{-E_1} d\mathbf{h}_1, \tag{13}$$

with

$$E_1 = (\mathbf{h}_1 - \alpha\mathbf{h}_0)^{*T}\mathbf{C}^{-1}(\mathbf{h}_1 - \alpha\mathbf{h}_0) + \frac{|r_1 - \mathbf{a}_1^T\mathbf{h}_1|^2}{\sigma^2}. \tag{14}$$

After integration over $\mathbf{h}_1$,

$$I_1 \doteq e^{\mathbf{q}_1{}^{*T}\mathbf{G}_1\mathbf{q}_1}|\mathbf{G}_1|e^{-\phi_1(\mathbf{h}_2)} \tag{15}$$

with:

$$\begin{aligned} \mathbf{G}_1 &= \left[\frac{\mathbf{a}_1^*\mathbf{a}_1^T}{\sigma^2} + (1+|\alpha|^2)\mathbf{C}^{-1}\right]^{-1} \\ \mathbf{q}_1 &= \frac{r_1\mathbf{a}_1^*}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{h}_0 \end{aligned} \tag{16}$$

The term $e^{-\phi_1(\mathbf{h}_2)}$ in (15) is a function of $\mathbf{h}_2$ and will be involved in the integral over $\mathbf{h}_2$. The rest of the terms in (15) are the results after the integration over $\mathbf{h}_1$ and are only decided by the input sequence and will be used to compute the maximum likelihood probability.

The general form of the results after integral over $\mathbf{h}_k$ ($k = 2, ..., n-1$) will be

$$I_k \doteq e^{\mathbf{q}_k{}^{*T}\mathbf{G}_k\mathbf{q}_k}|\mathbf{G}_k|e^{-\phi_k(\mathbf{h}_{k+1})}, \tag{17}$$

with

$$\begin{aligned} \mathbf{G}_k &= \left[\frac{\mathbf{a}_k^*\mathbf{a}_k^T}{\sigma^2} + \mathbf{C}^{-1} + |\alpha|^2\left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{G}_{k-1}\mathbf{C}^{-1}\right)\right]^{-1} \\ \mathbf{q}_k &= \frac{r_k\mathbf{a}_k^*}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_{k-1}\mathbf{q}_{k-1}. \end{aligned} \tag{18}$$

For the integral over $\mathbf{h}_n$,

$$I_n \doteq e^{\mathbf{q}_n{}^{*T}\mathbf{G}_n\mathbf{q}_n}|\mathbf{G}_n| \tag{19}$$

with

$$\begin{aligned} \mathbf{G}_n &= \left[\frac{\mathbf{a}_n^*\mathbf{a}_n^T}{\sigma^2} + \mathbf{C}^{-1} - |\alpha|^2\mathbf{C}^{-1}\mathbf{G}_{n-1}\mathbf{C}^{-1}\right]^{-1} \\ \mathbf{q}_n &= \frac{r_n\mathbf{a}_n^*}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_{n-1}\mathbf{q}_{n-1} \end{aligned} \tag{20}$$

After the integration over all the channel coefficients, we have the likelihood function

$$f(r|\mathbf{A}) \doteq \prod_{k=1}^{n} e^{\mathbf{q}_k{}^{*T}\mathbf{G}_k\mathbf{q}_k}|\mathbf{G}_k|. \tag{21}$$

If we take the negative natural logarithm of the above equation, we will obtain the time-recursive form of the cost function. However, from the expressions of $\mathbf{G}_k$ and $\mathbf{q}_k$ we know the transition costs depend on all the previous states, so the standard Viterbi algorithm can not be applied here. The optimum results can only be obtained by exhaustive search. However, a suboptimum algorithm which we call GM may be derived and we introduce this algorithm as follows. Observe equations (16), (18) and (21). Note that the parameters in the computation of the cost such as g_k and $\mathbf{G}_k$ are actually a function of all the previous transmitted data symbols. This observation suggests that we use per-survivor processing (PSP) [6] to decrease the computational complexity. It is well known that if the quantities corresponding to the computation of MLSE are unknown, we may use data-aided estimation to estimate them. In PSP, these quantities are estimated by data-aided estimation, but the sequence used in data-aided estimation of the transition cost corresponding to a particular state transition is the sequence associated with the survivor path before the current transition. The path obtained this way should approach the optimum results due to reduced error propagation. Furthermore, in our algorithm, GVA is applied [8], we compute the g_k and $\mathbf{G}_k$ associated with the survivors of the trellis, and these g_k and $\mathbf{G}_k$ are used to compute the current transition cost and to find the survivors at the next stage. At each state of each stage, we keep more than one survivor and we denote this fixed number of survivors as L. As L increases, the algorithm will approach the optimum results. This algorithm is a suboptimum but effective method for adaptive MLSE.

5. Simulation results

5.1. Time-Independent Gaussian channel

To illustrate the effectiveness of the computationally efficient, time-recursive algorithm described above for the time-independent, Gaussian model, 10,000 trials were performed for each of several values of SNR from 5 to 17 dB. For each trial, random BPSK transmitted symbol values were generated using a uniform distribution, so the symbol values -1 and 1 were equally likely. A channel FIR length of $M = 3$ was used. The channel coefficients were generated as independent (over time) Gaussian random variables, with mean $\mathbf{d}_k$ changing linearly from 1 to 2 over the block of time. The following constant channel coefficient covariance

$$\mathbf{C} = v \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \quad (22)$$

was used, and the initial channel symbols were fixed as $a_i = -1, i = -M+2, \cdots, 0$. Figure 1 shows the BER vs. SNR for several values of the channel coefficient variance parameter v. When v is smaller, the algorithm performs better, approaching closely the performance of the known channel MLSE.

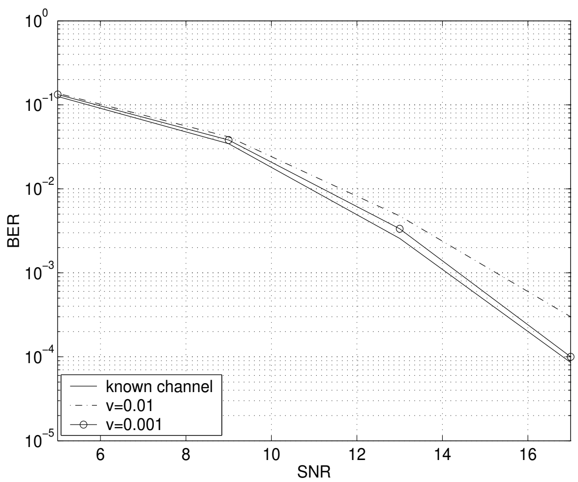

Figure 1. BER vs. SNR for time-varying independent Gaussian channel

5.2. Gauss-Markov channel

For the Gauss-Markov channel simulations, we used covariance

$$\mathbf{C} = v \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (23)$$

The initial channel coefficient vector $\mathbf{h}_0$ is constant. The following figures show results of the GM algorithm. PSP LMS and RLS algorithms are used for comparison. The simulations were run for 1000 trials with a block length of 100. The known channel estimation is presented as a reference bound. Simulations show that when v is small, GM results approach the known channel. Figures 2 and 3 show the results when $v = 0.001$ and $v = 0.01$.

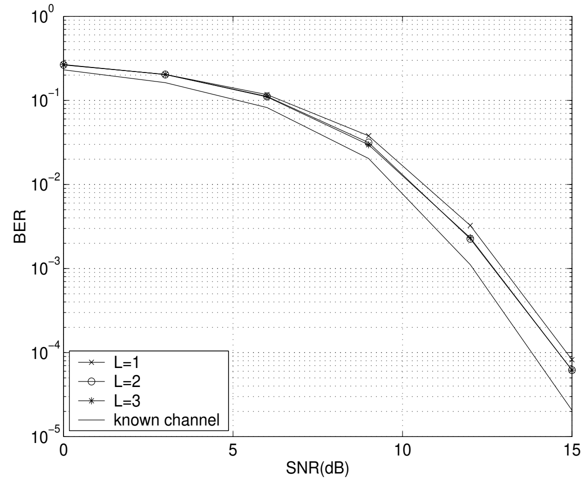

Figure 2. BER vs. SNR for Gaussian-Markov channel with v=0.001

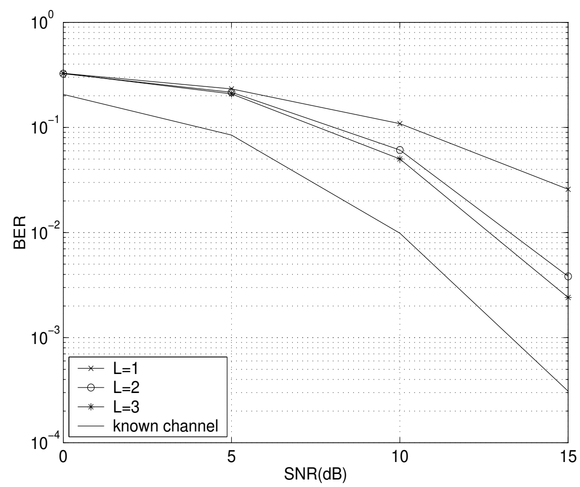

Figure 3. BER vs. SNR for Gaussian-Markov channel with v=0.01

It can be observed that as L increases, the performance gets better for the same v, and the improvement is greater for larger v (i.e. for faster changing channels). Simulations also show that PSP GM performance approaches that of exhaustive search (although it is not shown in the figures). Figures 4 and 5 show the comparison of PSP LMS and RLS with GM when v is 0.01 and 0.001 respectively. The step size in LMS and forgetting factor in RLS were chosen by

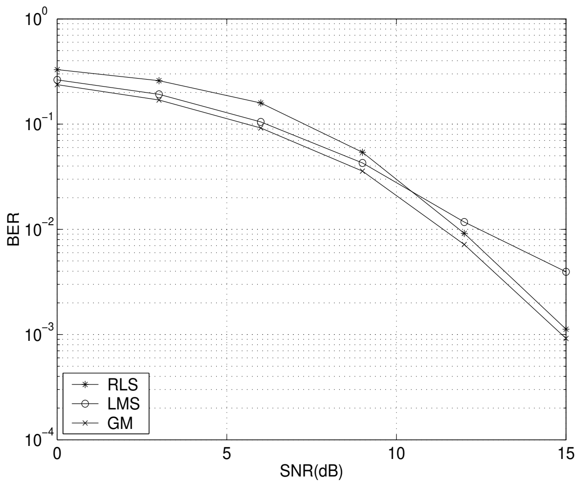

Figure 4. BER performance comparison for GM and PSP with LMS and RLS channel estimator with v=0.01

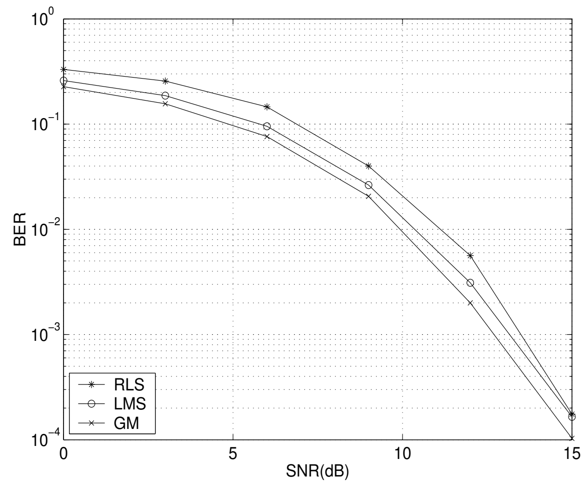

Figure 5. BER performance comparison for GM and PSP with LMS and RLS channel estimator with v=0.001

trial and error to obtain best possible results. The simulations show that GM performs better than PSP LMS/RLS.

6. Conclusion

In this paper we described the time-independent Gaussian and Gauss-Markov algorithms . We developed the optimum time-recursive algorithm for time-independent Gaussian channels. We developed an efficient time-recursive approximation to the optimal Gauss-Markov algorithm. We also compared performance of these algorithms to established suboptimal time-recursive algorithms.

References

[1] M. Ghosh and C. L. Weber. Maximum-likelihood blind equalization. *Optical Engineering*, 31(6):1224–1228, June 1992.

[2] K. M. Hiroshi Kubo and T. Fujino. An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels. *IEEE Transactions on Communications*, pages 1872 – 1880, February 1994.

[3] 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*, pages 120–124, Jan 1973.

[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] S. U. H. Qureshi. An adaptive decision-feedback receiver using maximum-likelihood sequence estimation. *Proc. 1973 IEEE Int. Conf. Commun.*, pages 14.10–14.16, 1973.

[6] 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*, 43(2/3/4):354–364, Feb/Mar/Apr 1995.

[7] N. Seshadri. Joint data and channel estimation using blind trellis search techniques. *IEEE Trans. on Communications*, 42(2/3/4):1000–1011, April 1994.

[8] N. Seshadri and C. Sundberg. List viterbi decoding algorithms with applications. *IEEE Transactions on Communications*, pages 313–323, February/March/April 1994.

[9] M. Stojanovic and Z. Zvonar. Performance of multiuser diversity reception in Rayleigh fading CDMA channels. *IEEE Trans. on Communications*, 47(3):356–359, Mar 1999.

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