Asymptotic Efficiency of a Blind Maximum Likelihood Sequence Detector

Jill K. Nelson and Andrew C. Singer†
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
1308 W. Main Street, MC-228
Urbana, IL 61801
Email: {jknelso1,acsinger}@uiuc.edu

Abstract— **We consider the performance of blind maximum likelihood sequence detection (MLSD) when the recursive least-squares (RLS) algorithm is used to update channel estimates. We employ asymptotic efficiency analysis to characterize the performance of the detector as the signal-to-noise ratio (SNR) approaches infinity. Asymptotic efficiency analysis allows us to quantify the loss in performance due to the presence of inter-symbol interference (ISI) and the lack of channel knowledge. We show that, under certain conditions, the asymptotic efficiency of the detector depends only on a single most-likely noise realization. Our results indicate that the performance of the RLS-based detector is strongly dependent on both the magnitude of the ISI and the number of data samples available.**

I. Introduction

A variety of communication systems are presented with the challenge of equalizing an unknown inter-symbol interference (ISI) channel. As a result, much work has been devoted to developing effective and computationally efficient methods for blind estimation of transmitted data [1]. The blind equalization schemes of interest in this paper are those that perform maximum likelihood sequence detection (MLSD) of the transmitted bits along a set of surviving paths through a trellis, analogous to the Viterbi algorithm [2, 3]. Several blind MLSD schemes employ the recursive least-squares (RLS) or the computationally simpler least mean-square (LMS) algorithm to update channel estimates along paths through the trellis [4–6]. Because of the prevalence of RLS-type estimation methods in blind equalization, as well as in other estimation applications, many researchers have studied the convergence of the RLS algorithm and its asymptotic performance in the limit as the number of available samples approaches infinity [7, 8]. Rather than analyzing the performance of RLS-based blind detectors as time approaches infinity, we study the performance of such detectors in the limit of high signal-to-noise ratio (SNR), i.e. as the noise variance tends to zero.

To perform such high-SNR analysis, we employ asymptotic efficiency, a concept applied widely in multi-user detection and more recently in single-user detection [9, 10]. Computing the asymptotic efficiency of a detector enables us to characterize its performance relative to the performance achievable when no interference is present. Following the definition presented in [10], we define the asymptotic efficiency of a detector as the exponent of decay of its error probability divided by the exponent of decay of the minimum error probability when no ISI is present, asymptotically as SNR tends to infinity. Through our analysis, we are able to quantify the asymptotic loss in performance that results due to the presence of ISI and the lack of channel information.

†This work was supported by the National Science Foundation under CCR-0092598 (CAREER) and by the Motorola Center for Communications.

II. Channel and Equalizer Model

We consider a discrete-time, symbol-spaced, time-invariant channel corrupted by additive white Gaussian noise (AWGN). For simplicity, we restrict our analysis to binary phase-shift keyed (BPSK) encoding over a two-tap channel with system function given by

$$C(z) = 1 + az^{-1}, \ -1 < a < 1. \tag{1}$$

The kth received sample can be written as

$$r_k = b_k + ab_{k-1} + w_k = \mathbf{c}^T \mathbf{b}_k + w_k, \tag{2}$$

where $\{b_k\}_{k=1}^N$ are the transmitted bits in a block of length N, w_k is AWGN independent of b_k with variance $\sigma^2 = N_0/2$, $\mathbf{c}^T = [1\ a]$, and $\mathbf{b}_k^T = [b_k\ b_{k-1}]$.

As a receiver, we consider a detector employing MLSD that has knowledge of the channel length but not of the channel tap values. The detector operates according to the standard Viterbi algorithm (i.e. retains the best path to each state at every time index) [3] and uses a least-squares approximation to estimate the channel along each surviving path. At each time k, the channel estimate along each surviving path is updated via the well-known RLS update [11]

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \frac{(r_{k+1} - \mathbf{b}_{k+1}^T \mathbf{h}_k)(\mathbf{R}_k^{-1} \mathbf{b}_{k+1})}{1 + \mathbf{b}_{k+1}^T \mathbf{R}_k^{-1} \mathbf{b}_{k+1}}, \tag{3}$$

where $\mathbf{h}_k$ denotes the channel estimate at time k, and $\mathbf{R}_k = \sum_{i=1}^k \mathbf{b}_i \mathbf{b}_i^T$ is assumed to be invertible.

III. Computation of Asymptotic Efficiency

The minimum bit error probability for a channel with bit energy $1 + a^2$ and no ISI is given by

$$P_{min} = Q\left(\sqrt{\frac{(1+a^2)}{\sigma^2}}\right) \sim \frac{1}{2} \exp\left(-\frac{1+a^2}{2\sigma^2}\right), \tag{4}$$

where $\sim$ indicates that

$$\lim_{x\to\infty} \frac{\log Q(x)}{\log\left(\frac{1}{2}\exp(-x^2/2)\right)} = 1. \tag{5}$$

The asymptotic efficiency of a detector with error probability $P_e = P_e(\sigma^2)$ is then given by

$$\eta = \lim_{\sigma^2\to 0} \frac{\log(2P_e)}{\log(2P_{min})} = \lim_{\sigma^2\to 0} \frac{-\log(2P_e)}{(1+a^2)/(2\sigma^2)}. \tag{6}$$

Again following [10], for a set of outcomes E whose probability depends on σ^2, we define the asymptotic exponent

$$\lambda_E = \lim_{\sigma^2\to 0} \left(-2\sigma^2 \log\left(2P(E)\right)\right). \tag{7}$$

A key property of λ_E is that it is dominated by the most likely event contained in E. Hence, to compute the asymptotic efficiency of a detector, we need consider only the most likely way in which an error can occur. For the two-tap channel $C(z)$, the two closest paths through the trellis (which determine the most likely error event) differ in only one bit. Thus, we study two paths, denoted by A and B, which differ only in their estimates of b_{n+1}. Without loss of generality, we let $\hat{b}_{n+1}^{(A)} = -1$ and $\hat{b}_{n+1}^{(B)} = +1$ denote the estimates of b_{n+1} along paths A and B, respectively. The relevant sections of the two paths through the trellis are shown in Fig. 1.

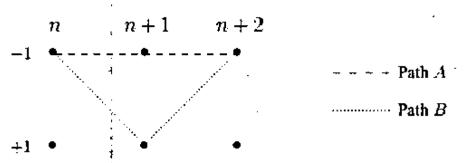

Fig. 1. Section of trellis for which paths A and B differ.

The dominating error event will be a single isolated error, and hence we can assume that paths A and B are correct up to time $n+1$. The RLS channel estimate generated along each path is given by

$$\mathbf{h}_n \triangleq \mathbf{h}_n^{(A)} = \mathbf{h}_n^{(B)} = \mathbf{R}_n^{-1}\mathbf{p}_n, \tag{8}$$

where $\mathbf{p}_n = \sum_{i=1}^{n} r_i \mathbf{b}_i$ [11]. Assuming path A is correct at time $n+1$, the probability that the detector makes an error (i.e. chooses path B) is the probability that $M^{(A)} \geq M^{(B)}$, where the metric of path A is computed as

$$M^{(A)} = \sum_{k=1}^{N} (r_k - \mathbf{h}_{k-1}^{(A)T}\mathbf{b}_k^{(A)})^2 \tag{9}$$

and similarly for $M^{(B)}$ [3]. Since A and B differ only in the $(n+1)$th bit, their metrics differ significantly only in the terms for which $k = n+1$ and $k = n+2$. (In fact, since the channel estimates along the two paths are slightly different, all terms for which $k > n+2$ will also differ slightly. These differences are comparatively small for large block lengths, however, and hence are ignored in our analysis.) Incorporating the RLS updates of the two channel estimates, the metric difference can be written

$$\begin{aligned} M^{(A)} - M^{(B)} = & \\ & \left(r_{n+1} - \mathbf{p}_n^T\mathbf{R}_n^{-1}\mathbf{b}_{n+1}^{(A)}\right)^2 - \left(r_{n+1} - \mathbf{p}_n^T\mathbf{R}_n^{-1}\mathbf{b}_{n+1}^{(B)}\right)^2 \\ & + \left(r_{n+2} - \left(\mathbf{R}_n^{-1}\mathbf{p}_n + \alpha_n\mathbf{R}_n^{-1}\mathbf{b}_{n+1}^{(A)}\right)^T \mathbf{b}_{n+2}^{(A)}\right)^2 \\ & - \left(r_{n+2} - \left(\mathbf{R}_n^{-1}\mathbf{p}_n + \beta_n\mathbf{R}_n^{-1}\mathbf{b}_{n+1}^{(B)}\right)^T \mathbf{b}_{n+2}^{(B)}\right)^2, \end{aligned} \tag{10}$$

where

$$\alpha_n = \left(\frac{r_{n+1} - \mathbf{b}_{n+1}^{(A)T}\mathbf{R}_n^{-1}\mathbf{p}_n}{1 + \mathbf{b}_{n+1}^{(A)T}\mathbf{R}_n^{-1}\mathbf{b}_{n+1}^{(A)}}\right), \tag{11}$$

and

$$\beta_n = \left(\frac{r_{n+1} - \mathbf{b}_{n+1}^{(B)T}\mathbf{R}_n^{-1}\mathbf{p}_n}{1 + \mathbf{b}_{n+1}^{(B)T}\mathbf{R}_n^{-1}\mathbf{b}_{n+1}^{(B)}}\right). \tag{12}$$

A. Simple approximations to $\mathbf{R}_n$ and $\mathbf{p}_n$

Both $\mathbf{R}_n$ and $\mathbf{p}_n$ are functions of all bits transmitted up to time n. Rather than assume values for $\{b_k\}_{k=1}^{n}$, we approximate $\mathbf{R}_n$ and $\mathbf{p}_n$ by their expected values, $\mathbf{R}_n \approx nI$, where I is the 2 by 2 identity matrix, and $\mathbf{p}_n \approx n\mathbf{c}$. With this approximation, $M^{(A)} - M^{(B)}$ becomes a function of only two variables: w_{n+1} and w_{n+2}.

Claim: The asymptotic exponent λ_{RLS} satisfies

$$\lambda_{RLS} = \min_{D} w_{n+1}^2 + w_{n+2}^2, \tag{13}$$

where $D = \{(w_{n+1}, w_{n+2}) : M^{(A)} - M^{(B)} \geq 0\}$.

Proof: Note that the sum of N independent Gaussian random vectors with mean $\vec{0}$ and covariance matrix K has a Gaussian distribution with mean $\vec{0}$ and covariance matrix NK. Let $\vec{u}$ be a zero-mean Gaussian random vector with covariance matrix $K = \sigma^2 K_0$, and let $\vec{v}_N = \sum_{i=1}^{N} \vec{u}$. Then

$$\lim_{\sigma^2\to 0} -2\sigma^2 \log P(\vec{u} \in F) = \lim_{N\to\infty} -\frac{2}{N\sigma^2} \log P(\vec{v}_N/N \in F). \tag{14}$$

Let the set F be defined as

$$F = \{\vec{v} : R(\vec{v}) \geq 0\} \cap \{\vec{v} : v_i \leq c,\ i = 1, \ldots, L\}, \tag{15}$$

where R is a continuous function, v_i denotes the ith component of $\vec{v}$, L denotes the length of $\vec{v}$, and c is an arbitrarily large scalar. By Ellis's Theorem [12],

$$\limsup_{N\to\infty} \frac{1}{N} \log P(\vec{v}_N/N \in F) \leq -\inf_{\vec{x}\in F} I(\vec{x}), \tag{16}$$

where

$$I(\vec{x}) = \sup_{\theta} [\vec{\theta}^T\vec{x} - \phi(\vec{\theta})], \tag{17}$$

and

$$\phi(\vec{\theta}) = \lim_{N\to\infty} \frac{1}{N} \log E[\exp\{\vec{\theta}^T\vec{v}_N\}]. \tag{18}$$

For the Gaussian vector $\vec{v}_N$,

$$\phi(\vec{\theta}) = \frac{1}{2}\vec{\theta}^T K\vec{\theta}. \tag{19}$$

This yields

$$I(\vec{x}) = \sup_{\vec{\theta}}[\vec{\theta}^T\vec{x} - \frac{1}{2}\vec{\theta}^T K\vec{\theta}] = \frac{1}{2}\vec{x}^T K^{-1}\vec{x}, \quad (20)$$

which is achieved for $\vec{\theta} = K^{-1}\vec{x}$. $I(\vec{x})$ is a paraboloid with its minimum at $\vec{x} = \vec{0}$. For a random vector whose elements are i.i.d. with variance σ^2, $K = \sigma^2 I$. It is easy to see for this case that, constrained to the region F, the minimizing value of $\vec{x}$ will be the value closest in Euclidean distance to $\vec{0}$, since

$$\inf_{\vec{x}\in F} I(\vec{x}) = \min_{\vec{x}\in F} \frac{1}{2\sigma^2}\vec{x}^T\vec{x}. \quad (21)$$

Because Ellis's Theorem requires that the set F be compact, we were required to append the restriction that $\{\vec{x} : x_i \leq c,\ i = 1\ldots L\}$ for some finite c. Note, however, that since the minimizing value of $\vec{x}$ is the vector with the minimum Euclidean norm, the restriction of the elements of $\vec{x}$ to finite values is not an active constraint. Hence, minimization over F yields the same result as minimization over the larger set $\{\vec{x} : R(\vec{x}) \geq 0\}$.

Now, define $G = \{\vec{v} : R(\vec{v}) > 0\}$. Ellis's Theorem also implies

$$\liminf_{N\to\infty} \frac{1}{N}\log P\{\vec{v}_N/N \in G\} \geq -\inf_{\vec{x}\in G} I(\vec{x}). \quad (22)$$

Since $I(\vec{x})$ is continuous and F is the closure of G,

$$\inf_{\vec{x}\in G} I(\vec{x}) = \inf_{\vec{x}\in F} I(\vec{x}), \quad (23)$$

and

$$\liminf_{N\to\infty} \frac{1}{N}\log P\{\vec{v}_N/N \in F\} \geq -\inf_{\vec{x}\in F} I(\vec{x}). \quad (24)$$

Thus, the limit exists and is given by

$$\lim_{n\to\infty} -\frac{2}{N\sigma^2}\log P(\vec{v}_N/N \in F) = \min_{\vec{x}\in F}\vec{x}^T\vec{x}.$$

Employing (14), letting $\vec{u} = [w_{n+1}\ w_{n+2}]$, and letting $R = M^{(A)} - M^{(B)}$, we see that

$$\begin{aligned} \lambda_{RLS} &= \lim_{\sigma^2\to 0} -2\sigma^2\log P(\vec{u}\in F) \quad (25)\\ &= \min_{M^{(A)}-M^{(B)}\geq 0} w_{n+1}^2 + w_{n+2}^2, \end{aligned}$$

which proves the claim.

Hence, as SNR tends to infinity, the probability of the most likely error event is a function only of the point of the contour $M^{(A)} - M^{(B)} = 0$ which lies closest in Euclidean distance to the point $(w_{n+1}, w_{n+2}) = (0, 0)$. Using Lagrange multipliers, we solve the constrained minimization problem and compute

$$\eta_{RLS} = \frac{\lambda_{RLS}}{(1+a^2)} \quad (26)$$

for $-1 < a < 1$ and $n = 1, 10, 100, 1000$. We also consider the case in which path B is correct, since the two cases are not symmetric when RLS channel estimation is employed.

Fig. 2 shows the results of the asymptotic efficiency calculations. For comparison, we include the asymptotic efficiency of MLSD when the channel is known, which is equal to 1 for all $a \in (-1, 1)$. As expected, asymptotic efficiency does deteriorate when the channel must be estimated. The reduction in asymptotic efficiency is much more significant for large $|a|$ than for small, especially for smaller values of n. However, as n increases, the asymptotic efficiency of RLS-based MLSD quickly approaches that of MLSD when the channel is known; for $n = 1000$, the asymptotic efficiency of the RLS-based detector is nearly indistinguishable from that of the known-channel detector.

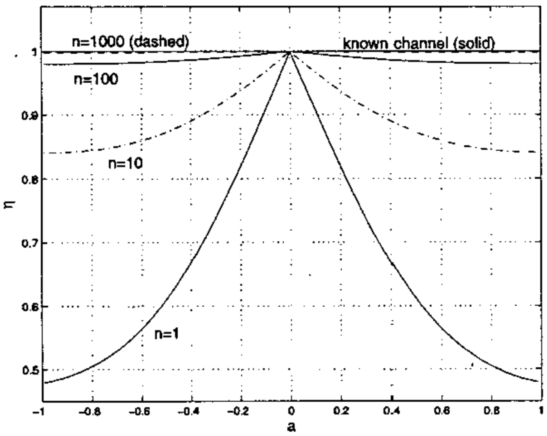

Fig. 2. Asymptotic efficiency of RLS-based MLSD for $a \in (-1, 1)$ and $n = 1, 10, 100, 1000$.

B. More accurate representation of $\mathbf{p}_n$

Rather than setting $\mathbf{p}_n$ equal to its expected value, we now consider a representation of $\mathbf{p}_n$ that includes the noise terms present in the cross-correlation vector. Denoting the sum $\sum_{i=1}^n b_i b_{i-1}$ by γ_n, we can write the autocorrelation matrix and cross-correlation vector as

$$\mathbf{R}_n = \begin{bmatrix} n & \gamma_n \\ \gamma_n & n \end{bmatrix} \quad (27)$$

and

$$\mathbf{p} = \begin{bmatrix} n + a\gamma_n + \sum_{i=1}^n b_i w_i, \\ \gamma_n + an + \sum_{i=1}^n b_{i-1} w_i \end{bmatrix}. \quad (28)$$

Since γ_n is not a function of the channel noise variance, we can consider the conditional asymptotic efficiency for a particular value of γ_n. For the following, we will set

$$\gamma_n = E\{\gamma_n\} = \sum_{i=1}^n E\{b_i b_{i-1}\} = 0. \quad (29)$$

For notational ease, we denote $\sum_{i=1}^n b_i w_i$ by y_n and $\sum_{i=1}^n b_{i-1} w_i$ by z_n. We can now write the autocorrelation matrix and cross-correlation vector as

$$\mathbf{R}_n = nI_2 \quad (30)$$

and

$$\mathbf{p}_n = \begin{bmatrix} n + y_n \\ an + z_n \end{bmatrix}. \quad (31)$$

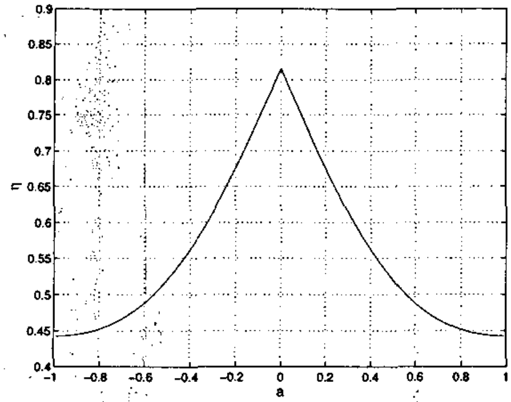

Fig. 3. Asymptotic efficiency of RLS-based MLSD for the first estimated symbol when three training symbols $\{b_k\}_{k=0}^{2} = \{1\ 1\ -1\}$ are first transmitted.

Substituting into (10), the metric difference now becomes a function of w_{n+1}, w_{n+2}, y_n, and z_n. The problem of finding the most likely values of w_{n+1}, w_{n+2}, y_n, and z_n is now a constrained minimization in four variables, all of which have zero-mean Gaussian distributions. While w_{n+1} and w_{n+2} are independent of the other two variables and of each other, the function to be minimized is complicated in that y_n and z_n are not jointly Gaussian. To see this, consider $n = 1$. Then $y_1 = b_1 w_1$, and $z_1 = b_0 w_1$. Clearly, both y_1 and z_1 are zero-mean Gaussians with variance σ^2, but given y_1, the value of z_1 is known up to a sign change. The resulting joint pdf of y_1 and z_1 is given by

$$f(y_1, z_1) = \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y_1^2}{2\sigma^2}} \left(\delta(z_1 - y_1) + \delta(z_1 + y_1)\right). \tag{32}$$

For general n, the joint pdf of y_n and z_n has a complicated form that does not allow for a simple minimization function like that used in the two-variable case. However, we can analyze the estimation of the first data symbol, since this special case is a function of four jointly Gaussian random variables. We consider a scenario in which three training symbols are transmitted, following which the first data symbol is transmitted and estimated. We choose training symbols $\{b_k\}_{k=0}^{2} = \{1\ 1\ -1\}$, which yields $\gamma_2 = 0$ in line with our previous assumption. The corresponding autocorrelation matrix and cross-correlation vector are $\mathbf{R} = 2I_2$ and

$$\mathbf{p} = \begin{bmatrix} 2 + b_1 w_1 + b_2 w_2 \\ 2a + b_0 w_1 + b_1 w_2 \end{bmatrix} = \begin{bmatrix} 2 + w_1 - w_2 \\ 2a + w_1 + w_2 \end{bmatrix}. \tag{33}$$

Following the argument used in the simplifed analysis, we minimize the function

$$f(w_1, w_2, w_3, w_4) = w_1^2 + w_2^2 + w_3^2 + w_4^2 \tag{34}$$

subject to the constraint $M^{(A)} - M^{(B)} = 0$. The resulting asymptotic efficiency as a function of a is shown in Fig. 3. When the noise terms in $\mathbf{p}_n$ are considered, the shape of the asymptotic efficiency curve is similar to that found via the simpler analysis. However, inclusion of the noise terms y_n and z_n does result in a noticeable decrease in asymptotic efficiency, especially for small $|a|$. For example, in the simpler analysis, the asymptotic efficiency reached 1 for $a = 0$, but when the new noise terms are considered, the maximum asymptotic efficiency (still occurring for $a = 0$) is only slightly above 0.8.

We now further consider the joint probability density function of y_n and z_n for arbitrary values of n. Suppose $b_0 = b_n = -1$. (These elements are fixed because b_0 is an assumed "prior bit" that appears only in z_n, and b_n is given by the paths A and B.) We claim that the joint probability density function of y_n and z_n is given by

$$f(y_n, z_n) = \frac{1}{2^{(n-1)}} \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} \frac{\exp\left\{-\frac{y_n^2 + z_n^2 - 2\rho_i y_n z_n}{\sigma^2(1-\rho_i^2)}\right\}}{2\pi\sigma^2\sqrt{1-\rho_i^2}}, \tag{35}$$

where $\rho_i = \frac{n-4i}{n}$, and $\lfloor n/2 \rfloor$ denotes the largest integer less than or equal to $n/2$. We derive this density function by first conditioning on the value of $\{b_k\}_{k=1}^{n-1}$. When $\{b_k\}_{k=1}^{n-1}$ is known, y_n and z_n are jointly Gaussian with covariance given by

$$\begin{aligned} E\left[y_n z_n | \{b_k\}_{k=1}^{n-1}\right] &= \sum_{i=1}^{n}\sum_{j=1}^{n} b_i b_{j-1} E[w_i w_j] \qquad (36) \\ &= \sigma^2 \sum_{i=1}^{n} b_i b_{i-1}. \end{aligned}$$

Hence, the covariance of y_n and z_n is completely determined by the value of $\sum_{i=1}^{n} b_i b_{i-1}$. The value of this summation is determined by the number of transitions (from +1 to -1 or from -1 to +1) in the sequence $\{b_k\}_{k=0}^{n}$.

There are n terms in the summation of interest corresponding to the possibility of up to n transitions. Each transition contributes -1 to the sum, and each "non-transition" contributes $+1$ to the sum. The covariance of y_n and z_n when conditioned on $\{b_k\}_{k=1}^{n-1}$ is then $\sigma^2(n - 2N_T)$, where N_T denotes the number of transitions in $\{b_k\}_{k=1}^{n-1}$.

To determine the unconditioned joint density of y_n and z_n, we compute

$$f(y_n, z_n) = \sum_{\{b_k\}_{k=1}^{n-1}} f(y_n, z_n | \{b_k\}_{k=1}^{n-1}) P(\{b_k\}_{k=1}^{n-1}), \tag{37}$$

where $P(\{b_k\}_{k=1}^{n-1})$ denotes the probability of the sequence $\{b_k\}_{k=1}^{n-1}$. Since we assume the transmitted bits are i.i.d. and equally likely to be +1 or -1, $P(\{b_k\}_{k=1}^{n-1}) = 2^{-(n-1)}$ for any $\{b_k\}_{k=1}^{n-1}$, yielding

$$f(y_n, z_n) = 2^{-(n-1)} \sum_{\{b_k\}_{k=1}^{n-1}} f(y_n, z_n | \{b_k\}_{k=1}^{n-1}).$$

It is clear that the joint pdf will be a Gaussian mixture generated by a sum of equally weighted zero-mean Gaussian probability density functions.

We now determine how many realizations of $\{b_k\}_{k=1}^{n-1}$ result in k transitions for k between 0 and n. Note that an unconstrained binary sequence of length $n+1$ could have any number of transitions from 0 to n. However, since we have constrained the first and last element of the sequence to be -1, the sequence $\{b_k\}_{k=0}^{n}$ must have an even number of transitions. The number of ways to have k transitions in a sequence of length $n+1$ with fixed endpoints is equivalent to choosing k elements from a set of size n. Hence, for k even, the number of sequences $\{b_k\}_{k=1}^{n-1}$ that contain k transitions is $\binom{n}{k}$, and the associated covariance between y_n and z_n is $(n-2k)\sigma^2$. Combining these findings, we reach the result given by (35).

Because the joint pdf of w_{n+1}, w_{n+2}, y_n, and z_n is a Gaussian mixture, letting $\sigma^2 \to 0$ can no longer be equivalently represented as taking the normalized sum of N identically distributed random variables and letting $N \to \infty$. Hence, the tools used to determine the general form of the asymptotic efficiency in the simplified case no longer apply. However, under certain conditions, it can be shown that the asymptotic efficiency in the Gaussian-mixture case is still a function of only a single point. Let P_i and P_j denote the probabilities that $R(\vec{x}) \geq 0$ for some continuous function R when $\vec{x}$ is a random variable with with distribution $p_i(\vec{x})$ and $p_j(\vec{x})$, respectively. We say that P_i dominates P_j if

$$\lim_{\sigma^2 \to 0} -2\sigma^2 \frac{P_j}{P_i} = 0. \tag{38}$$

Consider a vector $\vec{x}$ of random variables with joint distribution given by

$$f(\vec{x}) = \sum_{j=1}^{L} \alpha_j p_j(\vec{x}), \tag{39}$$

where $\sum_{j=1}^{L} \alpha_j = 1$, and for $j = 1, \ldots, L$, each $p_j(\vec{x})$ is the probability density function of a vector of zero-mean jointly Gaussian random variables, each element of the vector having marginal variance σ^2.

Claim: When P_i dominates P_j for all $j \neq i$, then

$$\lim_{\sigma^2 \to 0} -2\sigma^2 \log P(R(\vec{x}) \geq 0) = \lim_{\sigma^2 \to 0} -2\sigma^2 \log P_i. \tag{40}$$

Proof:

$$\begin{aligned}
&\lim_{\sigma^2 \to 0} -2\sigma^2 \log \int_{R(\vec{x})\geq 0} f(\vec{x}) d\vec{x} \qquad (41)\\
&= \lim_{\sigma^2 \to 0} -2\sigma^2 \log \sum_{j=1}^{L} \int_{R(\vec{x})\geq 0} \alpha_j p_j(\vec{x}) d\vec{x}\\
&= \lim_{\sigma^2 \to 0} -2\sigma^2 \log \sum_{j=1}^{L} \alpha_j P_j\\
&= \lim_{\sigma^2 \to 0} -2\sigma^2 \log \alpha_i P_i + \lim_{\sigma^2 \to 0} -2\sigma^2 \log \left(1 + \sum_{j\neq i} \frac{\alpha_j P_j}{\alpha_i P_i}\right)\\
&= \lim_{\sigma^2 \to 0} -2\sigma^2 \log \alpha_i P_i\\
&\quad + \lim_{\sigma^2 \to 0} -2\sigma^2 \left(\sum_{j\neq i} \frac{\alpha_j P_j}{\alpha_i P_i} + O\left(\left(\sum_{j\neq i} \frac{\alpha_j P_j}{\alpha_i P_i}\right)^2\right)\right)\\
&= \lim_{\sigma^2 \to 0} -2\sigma^2 \log \alpha_i P_i\\
&= \lim_{\sigma^2 \to 0} -2\sigma^2 \log P_i,
\end{aligned}$$

where the fourth equality follows from a Taylor Series expansion of the log term, the fifth equality follows from the supposition that P_i dominates P_j for all $j \neq i$, and the last equality holds because α_j is a constant independent of σ^2. Hence, the asymptotic efficiency in the Gaussian mixture case is equivalent to the asymptotic efficiency under the assumption that $\vec{x}$ is distributed according to the dominating Gaussian distribution in the sum.

IV. Conclusion

Using asymptotic efficiency as a metric, we have characterized the high-SNR performance of a blind maximum-likelihood sequence detector using RLS to update the channel estimate along each path. When the autocorrelation matrix and crosscorrelation vector are approximated by their expected values, the asymptotic efficiency can be computed as the constrained minimization of a simple quadratic function. Under certain conditions on the joint distribution of the noise elements, the same computation can be applied when a more accurate representation of the crosscorrelation vector is used. The asymptotic efficiency results reveal that the high-SNR performance of the detector studied varies significantly with channel knowledge and with the magnitude of the ISI present.

References

[1] Z. Ding and Y. Li, *Blind equalization and identification.* Marcel Dekker, 2001.

[2] R. Raheli, A. Polydoros, and T. Ching-Kae, "Per-survivor processing: a general approach to MLSE in uncertain environments," *IEEE Transactions on Communications*, vol. 43, pp. 354–364, February-April 1995.

[3] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, pp. 363–378, May 1972.

[4] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Transactions on Communications*, vol. 42, pp. 1000–1011, February-April 1994.

[5] J. Lin, F. Ling, and J. Proakis, "Joint data and channel estimation for TDMA mobile channels," *International Journal of Wireless Information Networks*, vol. 1, pp. 229–238, October 1994.

[6] R. Iltis, J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filtering," *IEEE Transactions on Communications*, vol. 42, pp. 1017–1032, February-April 1994.

[7] E. Eweda, "Convergence analysis of adaptive filtering algorithms with singular data covariance matrix," *IEEE Transactions on Signal Processing*, vol. 49, pp. 334–343, February 2001.

[8] W. Leon and D. Taylor, "Steady-state tracking analysis of the RLS algorithm for time-varying channels: a general state-space approach," *IEEE Communications Letters*, vol. 7, pp. 236–238, May 2003.

[9] S. Verdu, *Multiuser Detection.* Cambridge University Press, 1998.

[10] J. K. Nelson, A. C. Singer, U. Madhow, and C. S. McGahey, "BAD: Bidirectional arbitrated decision feedback equalization," February 2003. Submitted to *IEEE Trans. on Comm.*

[11] S. Haykin, *Adaptive Filter Theory.* Prentice-Hall, 1996.

[12] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation.* John Wiley and Sons, 1990.

[13] K. A. Ross, *Elementary Analysis: The Theory of Calculus.* Springer-Verlag, 1998.