ERROR PROBABILITY BOUNDS FOR NONLINEAR CHANNELS WITH INTERSYMBOL INTERFERENCE AND SIGNAL-DEPENDENT CORRELATED NOISE

Aleksandar Kavčić
Division of Engineering and Applied Sciences
Harvard University, Cambridge, MA 02138

Abstract We derive the performance bounds for the signal-dependent correlation-sensitive maximum likelihood sequence detector. Due to the correlation asymmetry of the channel, computations of the bounds are numerically intensive. For this reason, we develop Q-function approximations. An example illustrates the applicability of the bound computing method.

1. Introduction

Analytic methods for evaluating (bounding) the performance of maximum likelihood sequence detectors (MLSDs) in linear channels with intersymbol interference (ISI) and stationary noise have been presented in [1], [2]. For nonlinear ISI channels with signal-dependent and correlated noise (encountered in high density recording), no such method has been developed. In practice, one typically relies on Monte Carlo simulations, but this fails to answer two important questions for nonlinear ISI channels with signal-dependent and correlated noise:

1. Can we determine the error rate of the optimal detector applied to a given channel (even when the detector is too complex to build or simulate)?

2. What error rate can we expect in low error probability regions where Monte Carlo simulations are too computationally expensive to perform?

In this paper, we develop an analytic approach that answers these questions.

It has been demonstrated recently that optimal maximum likelihood sequence detection (MLSD) in nonlinear ISI channels with signal-dependent and correlated noise is performed by the Viterbi algorithm with branch-specific multidimensional branch metrics [3]. The necessary condition for applying this detector is to model the noise as a signal-dependent autoregressive (i.e., Markov) Gaussian process [4]. Using the structure of the signal-dependent Gauss-Markov noise processes, we derive an analytic expression for evaluating the performance bounds of the MLSD. Unfortunately, due to the correlation asymmetry of the signal-dependent channel noise, the computational complexity of the numeric evaluation of this expression is comparable to the Monte Carlo simulation of the detector itself. For this reason, we develop a computationally efficient Q-function approximation for high signal-to-noise ratios (SNRs). We demonstrate the accuracy of this approximation through an example in Section 5.

Throughout the paper, the following notation is used. Underlined characters represent column vectors, while boldface characters represent matrices, where $\mathbf{I}$ denotes the identity matrix. The superscript $^{\mathrm{T}}$ stands for matrix transposition. The notation $\mathrm{P}(A)$ denotes the probability of the event A; $\mathrm{E}[z]$ denotes the expected value of the random variable z; $\mathrm{P}(A|B)$ and $\mathrm{E}[z|B]$ denote the conditional probability and the conditional expectation, conditioned on the event B, respectively.

2. Channel Model and the Maximum Likelihood Sequence Detector

We assume the signal-dependent autoregressive intersymbol interference (ISI) channel model [4]. The sampled channel output of the model is given by

$$z_k = y\left(\underline{a}_k^{k-I}\right) + n_k. \tag{1}$$

Here n_k is the additive Markov signal-dependent correlated noise term which we describe below. The noiseless signal $y\left(\underline{a}_k^{k-I}\right)$ is a function of $I+1$ binary input symbols a_{k-I} through a_k denoted by the vector $\underline{a}_k^{k-I} = [a_{k-I}, \ldots, a_k]^{\mathrm{T}}$. We call I the intersymbol interference length. To allow for general *nonlinear* channels (i.e., channels with nonlinear transition shifts, partial erasure, MR head nonlinearities, etc.), we do *not* model the noiseless response $y\left(\underline{a}_k^{k-I}\right)$ as a convolution between an impulse response and the input symbols.

The noise n_k in Equation (1) is obtained via a signal-dependent autoregressive (AR) L-tap filter, Figure 2

$$n_k = \sigma\left(\underline{a}_k^{k-I}\right) w_k + \underline{b}\left(\underline{a}_k^{k-I}\right)^{\mathrm{T}} \underline{n}_{k-1}^{k-L}. \tag{2}$$

Here w_k is a zero-mean unit-variance white Gaussian process, and $\sigma\left(\underline{a}_k^{k-I}\right)$ is the standard deviation term dependent on $I+1$ binary input symbols $\underline{a}_k^{k-I}$. The symbol $\underline{b}\left(\underline{a}_k^{k-I}\right) = \left[b_L\left(\underline{a}_k^{k-I}\right), \ldots, b_1\left(\underline{a}_k^{k-I}\right)\right]^{\mathrm{T}}$ is a vector

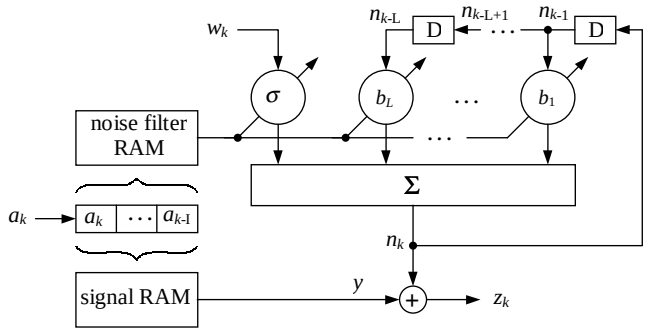

Fig. 1. Channel with intersymbol interference (ISI) of length L and signal-dependent Gauss-Markov noise with memory length L created by a signal dependent L-tap autoregressive filter.

of L autoregressive coefficients whose values depend on the $I+1$ binary input symbols $\underline{a}_k^{k-I}$. The vector $\underline{n}_{k-1}^{k-L}$ collects L previous samples of the noise process n_k, i.e., $\underline{n}_{k-1}^{k-L} = [n_{k-L}, \ldots, n_{k-1}]^{\mathrm{T}}$. We call L the Markov memory length of noise.

In Figure 2, for the channel model (1) and (2), the maximum likelihood sequence detector (MLSD) is a Viterbi detector [3]. It has 2^{I+L} states, with 2 branches entering and 2 branches leaving each state. Each state x can be named by collecting $I+L$ values of the input symbols. A state x_k at time k can thus be denoted by $x_k = \underline{a}_k^{k-I-L+1}$. A trellis branch is defined by two consecutive states $(x_{k-1}, x_k) = \underline{a}_k^{k-I-L}$. The branch metric is computed as follows. Let $\underline{z}_k^{k-L} = [z_{k-L}, \ldots, z_k]^{\mathrm{T}}$ be the vector of $L+1$ consecutive discrete-time channel outputs. For every branch $(x_{k-1}, x_k) = \underline{a}_k^{k-I-L}$, we form a branch-dependent innovation vector

$$\underline{N}_k^{k-L} = \underline{z}_k^{k-L} - \left[y\left(\underline{a}_{k-L}^{k-I-L}\right), \ldots, y\left(\underline{a}_k^{k-I}\right)\right]^{\mathrm{T}}, \quad (3)$$

where the values $y\left(\underline{a}_k^{k-I}\right)$ are the same as those used in (1). For every branch $(x_{k-1}, x_k) = \underline{a}_k^{k-I-L}$, we next compute the maximum likelihood (ML) branch metric

$$\mathcal{M}\left(\underline{z}_k^{k-L}, \underline{a}_k^{k-I-L}\right) = \ln \sigma^2\left(\underline{a}_k^{k-I}\right) + \frac{\left(\left[-\underline{b}\left(\underline{a}_k^{k-I}\right)^{\mathrm{T}} \;\; 1\right]\underline{N}_k^{k-L}\right)^2}{\sigma^2\left(\underline{a}_k^{k-I}\right)}. \quad (4)$$

The signal-dependent correlation-sensitive MLSD is then implemented by a Viterbi detector with 2^{I+L} states and branch metrics given by (4). An FIR filter implementation is given in [3].

3. Bit Error Probability Bounds

Essentially, the bounds that we will derive here are the same as those presented in [1], [2]. However, since the relative distances between the trellis paths are not symmetric due to the noise correlation asymmetry (signal-dependence), we cannot formulate the bound using the flow-graph transfer function [1]. To formulate the bounds, we first introduce the notation. Then we derive the *binary hypothesis* error probability for asymmetric Gauss-Markov noise channels. Finally, we formulate the upper and lower bounds using the binary hypothesis error probability expressions.

A. Notation

Let $\varepsilon_M = (\Psi'_M, \Psi''_M)$ denote a length-M error event. Here Ψ'_M denotes the correct path through the trellis spanning $M+1$ consecutive states $x'_k, \ldots, x'_{k+M}$. The erroneous path is denoted by Ψ''_M and it spans states $x''_k, \ldots, x''_{k+M}$. Thereby, it is assumed that $x'_k = x''_k$, $x'_{k+M} = x''_{k+M}$, and $x'_{k+m} \neq x''_{k+m}$ for $1 \leq m \leq M-1$. Since a state is defined as $x_k = \underline{a}_k^{k-I-L+1}$, an equivalent notation is $\Psi'_M = \underline{a}'^{k-I-L+1}_{k+M}$ and $\Psi''_M = \underline{a}''^{k-I-L+1}_{k+M}$, where $\underline{a}'^{k-I-L+m+1}_{k+m} \neq \underline{a}''^{k-I-L+m+1}_{k+m}$ for $1 \leq m \leq M-1$ and $\underline{a}'^{k-I-L+m+1}_{k+m} = \underline{a}''^{k-I-L+m+1}_{k+m}$ for all other m.

B. Binary Hypothesis Error Probability

The binary hypothesis error probability $\mathrm{P}_2(\varepsilon_M)$ is the probability of detecting Ψ''_M when Ψ'_M is the true path. This happens when the metric (4) along the path Ψ'_M exceeds the accumulated metric along Ψ''_M, i.e.,

$$\mathrm{P}_2(\varepsilon_M) = \mathrm{P}\left\{\sum_{m=1}^{M} \mathcal{M}\left(\underline{z}_{k+m}^{k+m-L}, \underline{a}'^{k+m-I-L}_{k+m}\right) > \sum_{m=1}^{M} \mathcal{M}\left(\underline{z}_{k+m}^{k+m-L}, \underline{a}''^{k+m-I-L}_{k+m}\right)\right\}. \quad (5)$$

Appendix A shows how to transform (5) into

$$\mathrm{P}_2(\varepsilon_M) = \mathrm{P}\left\{\underline{w}^{\mathrm{T}}\underline{w} > \ln\det \boldsymbol{\Sigma}_{\varepsilon_M} + \left(\underline{w} - \underline{m}_{\varepsilon_M}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\varepsilon_M}^{-1}\left(\underline{w} - \underline{m}_{\varepsilon_M}\right)\right\}, \quad (6)$$

where $\underline{w}$ is an $(L+M)\times 1$ Gaussian random vector with $\mathrm{E}[\underline{w}] = \underline{0}$ and $\mathrm{E}\left[\underline{w}\,\underline{w}^{\mathrm{T}}\right] = \mathbf{I}$; $\boldsymbol{\Sigma}_{\varepsilon_M}$ is a $(L+M)\times(L+M)$ covariance matrix; and $\underline{m}_{\varepsilon_M} = [d, 0, \cdots, 0]^{\mathrm{T}}$, $d \geq 0$. Appendix A relates $\boldsymbol{\Sigma}_{\varepsilon_M}^{-1}$ and d to the parameters of the model equations (1) and (2).

In the general case, the expression on the right-hand side of (6) cannot be simplified by integrating a χ^2-like distribution. These simplifications are possible only in some special cases [5]. A brute force approach is to numerically evaluate (6) using Monte-Carlo simulations, but this is very time consuming for low error rate (high SNR) scenarios. In Section 4, a time-saving Q-function

approximation is developed, leading to accurate evaluations of (6) for high SNR scenarios where Monte Carlo simulations take simply too long.

C. Upper Bound

The bit error probability P_b can be upper bounded by the union bound [1]. Let M denote the length (in branches) of an error event $\varepsilon_M = (\Psi'_M, \Psi''_M)$. For our channel, the minimum possible error event length can be verified to be $M_{min} = I + L + 1$, where I is the ISI length, and L is the Markov memory length of noise. The set of all possible error events of length M is denoted by E_M. Denote by $b_\#(\varepsilon_M)$ the number of erroneous bits corresponding to the error event ε_M. Let $\mathrm{P}(x)$ be the a priori probability of the trellis state x. Similarly, let $\mathrm{P}(x_k|x_{k-1})$ denote the conditional probability of transition from state x_{k-1} to state x_k. The binary hypothesis error probability associated with the error event ε_M is $\mathrm{P}_2(\varepsilon_M)$. The upper bound on bit error probability is given by

$$\mathrm{P}_b \leq \sum_{M=M_{min}}^{\infty} \sum_{\Psi'_M} \left[\mathrm{P}\left(x'_k\right) \prod_{i=1}^{M} \mathrm{P}\left(x'_{k+i}|x'_{k+i-1}\right) \right] \cdot \sum_{\substack{\Psi''_M \text{ such that} \\ \varepsilon_M = (\Psi'_M, \Psi''_M) \in \mathrm{E}_M}} b_\#\left(\varepsilon_M\right) \mathrm{P}_2\left(\varepsilon_M\right), \tag{7}$$

where the corresponding binary hypothesis error probability $\mathrm{P}_2(\varepsilon_M)$ needs to be substituted from Equation (6). If all symbol sequences are equally likely, then $\mathrm{P}\left(x'_k\right) = 2^{-(L+I)}$ and $\mathrm{P}\left(x'_{k+i}|x'_{k+i-1}\right) = 1/2$. A practical method for evaluating the bound in (7) is to truncate the first sum to values of $M < M_{max}$, where M_{max} is a predetermined large enough constant. Notice that in (7), due to the channel asymmetry, the binary hypothesis error probability $\mathrm{P}_2(\varepsilon_M)$ needs to be determined for every qualifying error event ε_M. If $\mathrm{P}_2(\varepsilon_M)$ in (6) is to be computed by Monte Carlo simulation, this would make the bound in (7) unacceptably expensive to compute, especially for low error rates. In Section 4, we provide a computationally efficient approximation.

D. Lower Bound

Obviously, a loose lower bound is obtained by picking any single term of the union bound (7). To get a tighter bound, we modify the genie-assisted bound [2].

Modified genie-assisted bound Denote by $\mathrm{P}(\varepsilon)$ the probability of an error event occurring. Let $\pi_c(\mathrm{P}_c)$ be the probability that the input sequence a' is such that there is a probability $\mathrm{P} \geq \mathrm{P}_c$ of confusion with some allowable sequence a''. Even with a genie-assisted detector that has to choose only between a' and a'', the probability of an error event $\mathrm{P}(\varepsilon)$ will be greater than $\pi_c(\mathrm{P}_c) \cdot \mathrm{P}_c$. Clearly, then

$$\mathrm{P}(\varepsilon) \geq \max_{0 \leq \mathrm{P}_c \leq 1} \left[\pi_c(\mathrm{P}_c) \cdot \mathrm{P}_c\right]. \tag{8}$$

The difference between this formulation and the one in [2] is that we assume $P \geq P_c$ instead of $P = P_c$. In fact, the bound in (8) will never be looser (and in many cases tighter) than the one in [2], even in linear symmetric channels.

We use (8) to find the lower bound for the bit error probability. Consider only error events of $M_{min}= I+L+1$ branches. The number of these error events for binary signaling is $N = 2^{I+L+M_{min}} = 2 \cdot 4^{I+L}$. Label these error events as $\varepsilon_{M_{min}}(1), \varepsilon_{M_{min}}(2), \cdots, \varepsilon_{M_{min}}(N)$, where their order is such that $\mathrm{P}_2(\varepsilon_{M_{min}}(1)) \geq \mathrm{P}_2(\varepsilon_{M_{min}}(2)) \geq \cdots \geq \mathrm{P}_2(\varepsilon_{M_{min}}(N))$, and $\mathrm{P}_2(\varepsilon_{M_{min}}(i))$ is the binary error probability of the error event $\varepsilon_{M_{min}}(i)$. Every error event is a pair $\varepsilon_{M_{min}}(i) = \left(\Psi'_{M_{min}}(i), \Psi''_{M_{min}}(i)\right)$, where $\Psi'_{M_{min}}(i)$ and $\Psi''_{M_{min}}(i)$ are 2 distinct sequences of $M_{min} + 1$ consecutive states; $\Psi'_{M_{min}}(i)$ is the correct path and $\Psi''_{M_{min}}(i)$ is the erroneous path. In binary signaling, for every sequence of $M_{min} + 1$ states $\Psi'_{M_{min}}(i)$ there exists only one erroneous sequence $\Psi''_{M_{min}}(i)$. We therefore have

$$\sum_{i=1}^{N} \mathrm{P}\left(\Psi'_{M_{min}}(i)\right) = \sum_{i=1}^{N} \left[\mathrm{P}\left(x'_k(i)\right) \prod_{l=1}^{M_{min}} \mathrm{P}\left(x'_{k+l}(i)|x'_{k+l-1}(i)\right) \right] = 1, \tag{9}$$

i.e., $\Psi'_{M_{min}}(i)$ are disjunct sequences.

If we choose $\mathrm{P}_c = \mathrm{P}_2(\varepsilon_{M_{min}}(i))$ for some i where $1 \leq i \leq N$, then $\pi_c\left(\mathrm{P}_c\right) \geq \sum_{j=1}^{i} \mathrm{P}\left(\Psi'_{M_{min}}(i)\right)$ because there may be error events of length $M > M_{min}$ whose binary hypothesis error probability is greater than P_c. In the interest of keeping the bound computation simple, we ignore these error events of length $M > M_{min}$. Since for all error events of length M_{min} the number of erroneous bits is $b_\#\left(\varepsilon_{M_{min}}(i)\right) = 1$, we have

$$\mathrm{P}_b \geq \pi_c\left(\mathrm{P}_c\right) \cdot \mathrm{P}_c \geq \left[\sum_{j=1}^{i} \mathrm{P}\left(\Psi'_{M_{min}}(j)\right) \right] \mathrm{P}_2\left(\varepsilon_{M_{min}}(i)\right) \tag{10}$$

for any $1 \leq i \leq N$. Maximizing the right hand side of (10), we obtain a tighter lower bound

$$\mathrm{P}_b \geq \max_{1 \leq i \leq N} \left\{ \left[\sum_{j=1}^{i} \mathrm{P}\left(\Psi'_{M_{min}}(j)\right) \right] \mathrm{P}_2\left(\varepsilon_{M_{min}}(i)\right) \right\}. \tag{11}$$

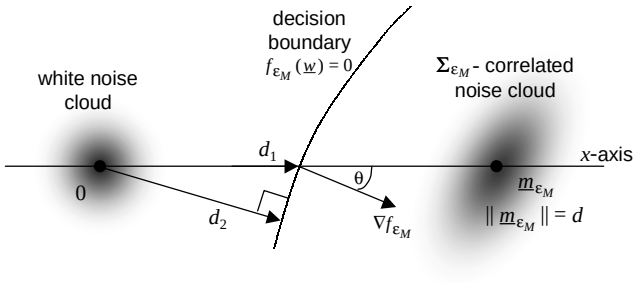

Fig. 2. Binary hypothesis signal and noise constellation encountered when computing the binary hypothesis error probability.

When all sequences $\Psi'_{M_{min}}(i)$ are equiprobable, we have $\mathrm{P}\left(\Psi'_{M_{min}}(j)\right) = \frac{1}{N} = \frac{1}{2}4^{-(I+L)}$. The modified genie-assisted lower bound is then

$$\mathrm{P}_b \geq \frac{1}{2}4^{-(I+L)} \max_{1\leq i\leq N} \left[i \cdot \mathrm{P}_2\left(\varepsilon_{M_{min}}(i)\right)\right]. \qquad (12)$$

Since the terms $\mathrm{P}_2\left(\varepsilon_{M_{min}}(i)\right)$ have already been computed as part of (7), no further computation is needed for the lower bound. Notice also that while (12) is always a lower bound, it may not be the tightest of the modified genie-assisted bounds because we considered only the minimum-length error events. Examples in Section 5 show, however, that this bound is relatively tight.

4. Q-function Approximations

For the bounds (7) and (12) to be of practical use, their computation must be much less expensive than simulating the detector. However, the evaluation of (6) is generally a complex task, which, if done by Monte Carlo simulation, requires the same time as simulating the detector itself. Computing (6) requires integrating a zero-mean $\mathbf{I}$-covariance multivariate Gaussian pdf over a region where

$$f_{\varepsilon_M}(\underline{w}) = \underline{w}^{\mathrm{T}}\underline{w} - \ln\det\boldsymbol{\Sigma}_{\varepsilon_M} - \left(\underline{w}-\underline{m}_{\varepsilon_M}\right)^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon_M}^{-1}\left(\underline{w}-\underline{m}_{\varepsilon_M}\right) > 0 \qquad (13)$$

Since a Gaussian function has a very fast decay, we approximate (6) by

$$\mathrm{P}_2(\varepsilon_M) \approx \mathrm{Q}(d_2), \qquad (14)$$

where d_2 is the point on $f_{\varepsilon_M}(\underline{w}) = 0$ closest to the origin and $\mathrm{Q}(x) = (2\pi)^{-\frac{1}{2}} \int_x^\infty \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$, see Figure 2.

Determining d_2 in (14) is a problem of minimizing $\|\underline{w}\|$ subject to $f_{\varepsilon_M}(\underline{w}) = 0$. A numeric solution can be obtained by iteratively approaching the point on the boundary $f_{\varepsilon_M}(\underline{w}) = 0$, for which the direction of $\underline{w}$ matches the direction of the gradient $\nabla f_{\varepsilon_M}(\underline{w})$ up to a small error. This approach is good for large $d = \left\|\underline{m}_{\varepsilon_M}\right\|$, i.e., for large signal to noise ratios (SNRs).

$\left(\underline{a}_k^{k-2}\right)^{\mathrm{T}}$	$y\left(\underline{a}_k^{k-2}\right)$	$\sigma\left(\underline{a}_k^{k-2}\right)$	$b_1\left(\underline{a}_k^{k-2}\right)$	$b_2\left(\underline{a}_k^{k-2}\right)$
000	-0.2	1.0	-0.1	0
001	0.9	1.4	-0.3	-0.1
010	-0.1	1.6	-0.6	-0.2
011	1.2	1.2	-0.2	-0.1
100	-1.0	1.2	-0.2	-0.1
101	0.1	1.6	-0.6	-0.2
110	-0.8	1.4	-0.3	-0.1
111	0.2	1.0	-0.1	0

TABLE I
Nonlinearly distorted $1 - D^2$ partial response channel (PR4) with ISI length $I = 2$ and Markov memory length $L = 2$.

Substituting d_2 into the expression for the binary hypothesis error probability (14), and then substituting $\mathrm{P}_2(\varepsilon_M)$ from (14) into (7) and (12), we obtain computationally efficient approximate bit error probability bounds for the MLSD, whose accuracy is demonstrated through an example in the next section.

5. Numeric Evaluations

We consider a non-ideal PR4 channel (ISI length $I = 2$) with signal-dependent and correlated noise whose Markov memory length is $L = 2$. The channel target values are given in Table I, and are not symmetric to model nonlinear effects (such as MR head nonlinearity, nonlinear transition shift and partial erasure). The Markov model parameters (Table I) are non-uniform to model the signal-dependent character of the correlated noise. The numeric evaluations of the parameters involved in computing the bounds are presented in Appendices A and B, where Appendix B is entirely devoted to the channel described in Table I.

Figure 3 shows the performance of the MLSD for the nonlinearly distorted PR4 channel with signal-dependent noise described in Table I. The MLSD has $2^{I+L} = 16$ states. The different values of the signal-to-noise ratio (SNR) shown in Figure 3 are obtained by scaling the noiseless channel response (column denoted by $y\left(\underline{a}_k^{k-2}\right)$ in Table I) while keeping the relative distance between the signal points the same. This corresponds to scaling the signal power while the noise constellation is kept the same. In Figure 3, the solid line represents the simulated performance of the ML sequence detector (MLSD). We do not show the MLSD performance curve beyond SNR=14dB since the Monte Carlo simulations take too long beyond that point. In Figure 3, the simulated performance of the Euclidean Viterbi detector (that assumes white channel noise) ap-

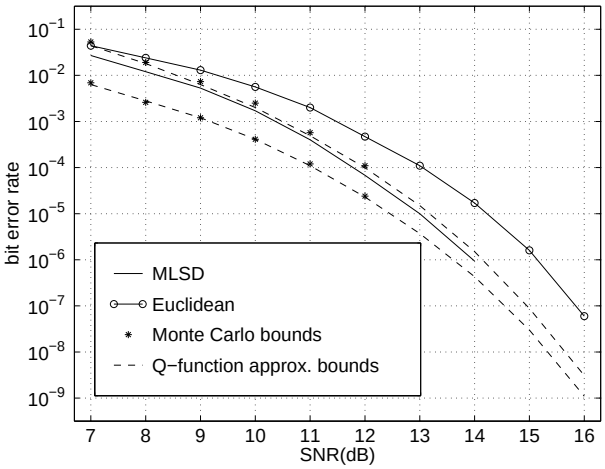

Fig. 3. Performance evaluations of detectors applied to the channel in Table I.

$\left(\underline{a}_k^{k-1}\right)^{\mathrm{T}}$	$y\left(\underline{a}_k^{k-1}\right)$	$\sigma\left(\underline{a}_k^{k-1}\right)$	$b_1\left(\underline{a}_k^{k-1}\right)$
00	-0.1	1.0	-0.2
01	1.2	1.8	-0.6
10	-0.8	1.2	-0.6
11	0.1	1.0	-0.2

TABLE II
Nonlinearly distorted $1-D$ partial response channel with ISI length $I=1$ and Markov memory length $L=1$.

plied to the channel in Table I has a 1 dB loss at the error rate 10^{-6}.

The stars (*) in Figure 3 show the error probability bounds computed by evaluating the binary hypothesis error probability (6) with Monte Carlo simulations for SNR$\leq$12dB where the Monte Carlo simulations are still computationally feasible. On the other hand, the approximate bounds (dashed lines in Figure 3), computed by approximating the binary hypothesis error probability with a Q-function (14), can be computed at a fractional computational cost for any SNR and are extremely close to the bounds computed by Monte Carlo simulations. Note also that with the Q-function approximate bounds we are able to roughly estimate the MLSD performance above SNR$\geq$14dB where simulations of the MLSD last too long. As expected, the bounds are tightening as the SNR increases which helps us evaluate the MLSD performance at high SNRs where Monte Carlo simulations of the MLSD are impractical.

The same general conclusions hold for the nonlinearly distorted $1-D$ partial response channel with ISI length $I=1$ and Markov memory length $L=1$ (Table II). Performance and bound curves are shown in Figure 4.

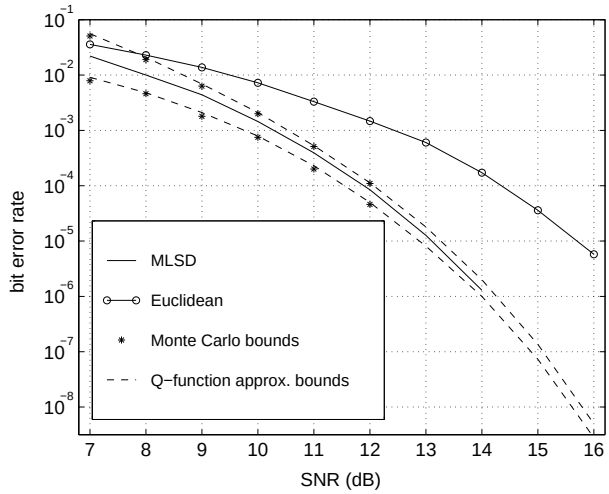

Fig. 4. Performance evaluations of detectors applied to the channel in Table II.

6. Conclusion

We have presented an analytic method for evaluating (bounding) the performance of the maximum likelihood sequence detector (MLSD) in channels with nonlinear ISI and signal-dependent and correlated noise. The computational complexities of numerically evaluating these bounds are comparable to the Monte Carlo simulation of the detector. However, computationally efficient Q-function approximation for high SNRs are shown to be good approximations to the true bounds computed by Monte Carlo simulations. They are particularly useful for evaluating the performance of the optimal detector applied to any nonlinear ISI channel with signal-dependent and correlated noise at error rates unachievable through Monte Carlo simulation, or when the MLSD is too complex to build (or simulate).

A Decision Boundary Transformations for Gauss-Markov Noise

In (5), the choice of k is arbitrary. For convenience, set $k=L$. Using the structural properties of Gauss-Markov covariance matrices [6], we transform (5) to

$$\mathrm{P}_2(\varepsilon_M)=\mathrm{P}\left\{\ln\det\mathbf{C}_{\Psi'_M}+\left(\underline{N'}^1_{M+L}\right)^{\mathrm{T}}\mathbf{C}^{-1}_{\Psi'_M}\underline{N'}^1_{M+L}\right.$$
$$\left.>\ln\det\mathbf{C}_{\Psi''_M}+\left(\underline{N''}^1_{M+L}\right)^{\mathrm{T}}\mathbf{C}^{-1}_{\Psi''_M}\underline{N''}^1_{M+L}\right\}. \quad (15)$$

Here $\mathbf{C}_{\Psi'_M}$ and $\mathbf{C}_{\Psi''_M}$ are the covariance matrices of $\underline{z}^1_{L+M}$ when $\underline{a'}^{1-I}_{L+M}$ and $\underline{a''}^{1-I}_{L+M}$ are the transmitted

(written) binary sequences, respectively. The innovation vector $\underline{N'}_{L+M}^{1}$ is defined as

$$\underline{N'}_{L+M}^{1} = \underline{z}_{L+M}^{1} - \underline{y'} = \underline{z}_{L+M}^{1} - \left[y\left(\underline{a'}_{1}^{1-I}\right), ..., y\left(\underline{a'}_{L+M}^{L+M-I}\right) \right]^{\mathrm{T}}, \tag{16}$$

where the values $y\left(\underline{a'}_{k}^{k-I}\right)$ are the same as those used in (1). The vectors $\underline{N''}_{M+L}^{1}$ and $\underline{y''}$ are defined similarly to (16), where the superscript $'$ is replaced by $''$.

$\mathbf{C}_{\Psi'_M}$ and $\mathbf{C}_{\Psi''_M}$ are the conditional covariance matrices of the noise process n_k for their respective transmitted (written) sequences. We next use the results on Gauss-Markov covariance matrices given in [6],[7]) to express $\mathbf{C}_{\Psi'_M}^{-1}$ (or $\mathbf{C}_{\Psi''_M}^{-1}$) in terms of the AR parameters $\underline{b}\left(\underline{a}_{k}^{k-I}\right)$ and $\sigma\left(\underline{a}_{k}^{k-I}\right)$ in (2).

Since the process n_k is conditionally Gauss-Markov with memory length L, $\mathbf{C}_{\Psi'_M}^{-1}$ is an L-banded matrix [6],[7]. Its upper Cholesky decomposition is

$$\mathbf{C}_{\Psi'_M}^{-1} = \mathbf{U}_{\Psi'_M} \mathbf{D}_{\Psi'_M} \mathbf{U}_{\Psi'_M}^{\mathrm{T}}. \tag{17}$$

Here, according to [6], [7], the upper triangular L-banded matrix $\mathbf{U}_{\Psi'_M}$ is

$$\mathbf{U}_{\Psi'_M} = \begin{bmatrix} \mathbf{U}_{1'} & -\underline{b}\left(\underline{a'}_{L+1}^{L+1-I}\right) & & \mathbf{0} \\ & 1 & \ddots & \\ & & \ddots & -\underline{b}\left(\underline{a'}_{L+M}^{L+M-I}\right) \\ \mathbf{0} & & & 1 \end{bmatrix}, \tag{18}$$

and the diagonal matrix $\mathbf{D}_{\Psi'_M}$ is

$$\mathbf{D}_{\Psi'_M} = \mathrm{diag}\Big[\mathbf{D}_{1'} \;\; 1/\sigma^2\left(\underline{a'}_{L+1}^{L+1-I}\right) \;\; \cdots \;\; 1/\sigma^2\left(\underline{a'}_{L+M}^{L+M-I}\right) \Big]. \tag{19}$$

The values $\underline{b}\left(\underline{a'}_{i}^{i-I}\right)$ and $\sigma\left(\underline{a'}_{i}^{i-I}\right)$ in (18) and (19) are the same as those in the model (2). The upper triangular matrix $\mathbf{U}_{1'}$ and the diagonal matrix $\mathbf{D}_{1'}$ are the upper Cholesky factors of

$$\mathbf{C}_{1'}^{-1} = \mathbf{U}_{1'} \mathbf{D}_{1'} \mathbf{U}_{1'}^{\mathrm{T}}, \tag{20}$$

where $\mathbf{C}_{1'}$ is the conditional covariance matrix of $\underline{z}_{L}^{1}$ when the sequence of transmitted (or written) symbols is $\underline{a'}_{L}^{1-I}$. The elements of $\mathbf{C}_{1'}$ can be obtained by solving a linear system of rearranged signal-dependent Yule-Walker equations [6], [8], a tedious but straightforward task, best illustrated through an example (Appendix B).

Once the Cholesky decompositions (17) of $\mathbf{C}_{\Psi'_M}^{-1}$ and $\mathbf{C}_{\Psi''_M}^{-1}$ are computed, we write the binary hypothesis error probability (6). Thereby, the parameters in (6) have the following properties: $\underline{w}$ is an $(L+M)\times 1$ Gaussian random vector with $\mathrm{E}[\underline{w}] = \underline{0}$ and $\mathrm{E}\left[\underline{w}\,\underline{w}^{\mathrm{T}}\right] = \mathbf{I}$; $\underline{m}_{\varepsilon_M}$ is

$$\underline{m}_{\varepsilon_M} = \mathbf{Q}\mathbf{D}_{\Psi'_M}^{1/2} \mathbf{U}_{\Psi'_M}^{\mathrm{T}} \left(\underline{y''} - \underline{y'}\right) = \left[\begin{array}{cccc} d & 0 & \cdots & 0 \end{array} \right]^{\mathrm{T}}, \tag{21}$$

where $d = \left\| \mathbf{D}_{\Psi'_M}^{1/2} \mathbf{U}_{\Psi'_M}^{\mathrm{T}} \left(\underline{y''} - \underline{y'}\right) \right\|$, $\underline{y'}$ and $\underline{y''}$ are defined in (16), $\|\cdot\|$ is the L_2 norm, $\mathbf{Q}$ is a unitary matrix (e.g., the Householder reflector $\mathbf{Q} = 2\underline{v}\,\underline{v}^{\mathrm{T}}/\|\underline{v}\|^2 - \mathbf{I}$, where $\underline{v} = \underline{q}/\|\underline{q}\| + \underline{e}_1$, $\underline{q} = \mathbf{D}_{\Psi'_M}^{1/2} \mathbf{U}_{\Psi'_M}^{\mathrm{T}} \left(\underline{y''} - \underline{y'}\right)$, and $\underline{e}_1 = [1, 0, \cdots, 0]^{\mathrm{T}}$, [9]); finally $\mathbf{\Sigma}_{\varepsilon_M}$ satisfies

$$\begin{aligned} \mathbf{\Sigma}_{\varepsilon_M}^{-1} &= \mathbf{Q}\mathbf{D}_{\Psi'_M}^{-1/2} \mathbf{U}_{\Psi'_M}^{-1} \mathbf{C}_{\Psi''_M}^{-1} \mathbf{U}_{\Psi'_M}^{-\mathrm{T}} \mathbf{D}_{\Psi'_M}^{-1/2} \mathbf{Q}^{\mathrm{T}} \\ \text{and} \quad \det \mathbf{\Sigma}_{\varepsilon_M} &= \frac{\det \mathbf{D}_{\Psi'_M}}{\det \mathbf{D}_{\Psi''_M}}. \end{aligned} \tag{22}$$

B Covariance Matrix Computation Example

We describe here the computation of the $L \times L$ conditional covariance matrices $\mathbf{C}_1$ (Equation (20)) for the example in Table I. There, $I = 2$ and $L = 2$. The channel model (1) and (2) is rewritten for convenience

$$n_k = \sigma\left(\underline{a}_{k}^{k-2}\right) w_k + b_1\left(\underline{a}_{k}^{k-2}\right) n_{k-1} + b_2\left(\underline{a}_{k}^{k-2}\right) n_{k-2} \tag{23}$$

$$z_k = y\left(\underline{a}_{k}^{k-2}\right) + n_k \tag{24}$$

where w_k is a zero-mean unit-variance Gaussian noise process while $y\left(\underline{a}_{k}^{k-2}\right)$, $\sigma\left(\underline{a}_{k}^{k-2}\right)$, $b_1\left(\underline{a}_{k}^{k-2}\right)$ and $b_2\left(\underline{a}_{k}^{k-2}\right)$ are in Table I. Since $L = 2$ and $I = 2$, $\mathbf{C}_1$ is

$$\begin{aligned} \mathbf{C}_1\left(\underline{a}_{k}^{k-3}\right) &= \begin{bmatrix} \mathrm{E}\left[n_{k-1}^2 | \underline{a}_{k}^{k-3}\right] & \mathrm{E}\left[n_{k-1} n_k | \underline{a}_{k}^{k-3}\right] \\ \mathrm{E}\left[n_{k-1} n_k | \underline{a}_{k}^{k-3}\right] & \mathrm{E}\left[n_k^2 | \underline{a}_{k}^{k-3}\right] \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{E}\left[n_{k-1}^2 | \underline{a}_{k}^{k-3}\right] & c_1\left(\underline{a}_{k}^{k-3}\right) \\ c_1\left(\underline{a}_{k}^{k-3}\right) & c_0\left(\underline{a}_{k}^{k-3}\right) \end{bmatrix}. \end{aligned} \tag{25}$$

Since n_{k-1} is independent of a_k

$$\begin{aligned} \mathrm{E}\left[n_{k-1}^2 | \underline{a}_{k}^{k-3}\right] &= \mathrm{E}\left[n_{k-1}^2 | \underline{a}_{k-1}^{k-3}\right] = \frac{1}{2}\mathrm{E}\left[n_{k-1}^2 \,\middle|\, \begin{bmatrix} 0 \\ \underline{a}_{k-1}^{k-3} \end{bmatrix} \right] \\ &\quad + \frac{1}{2}\mathrm{E}\left[n_{k-1}^2 \,\middle|\, \begin{bmatrix} 1 \\ \underline{a}_{k-1}^{k-3} \end{bmatrix} \right] = \frac{1}{2} \sum_{a_{k-4}=0}^{1} c_0\left(\underline{a}_{k-1}^{k-4}\right). \end{aligned} \tag{26}$$

It suffices to determine $c_0\left(\underline{a}_{k}^{k-3}\right)$ and $c_1\left(\underline{a}_{k}^{k-3}\right)$ to find $\mathbf{C}_1$ in (25). The conditional expectation of (23) squared is

$$\begin{aligned} \mathrm{E}\left[n_k^2 | \underline{a}_{k}^{k-3}\right] &= \sigma^2\left(\underline{a}_{k}^{k-2}\right) + b_1^2\left(\underline{a}_{k}^{k-2}\right) \mathrm{E}\left[n_{k-1}^2 | \underline{a}_{k}^{k-3}\right] \\ &\quad + b_2^2\left(\underline{a}_{k}^{k-2}\right) \mathrm{E}\left[n_{k-2}^2 | \underline{a}_{k}^{k-3}\right] \\ &\quad + 2 b_1\left(\underline{a}_{k}^{k-2}\right) b_2\left(\underline{a}_{k}^{k-2}\right) \mathrm{E}\left[n_{k-2} n_{k-1} | \underline{a}_{k}^{k-3}\right]. \end{aligned} \tag{27}$$

The conditional expectation of (23) times n_{k-1} is

$$\begin{aligned} \mathrm{E}\left[n_{k-1}n_k|\underline{a}_k^{k-3}\right] &= b_1\left(\underline{a}_k^{k-2}\right)\mathrm{E}\left[n_{k-1}^2|\underline{a}_k^{k-3}\right] \\ &\quad + b_2\left(\underline{a}_k^{k-2}\right)\mathrm{E}\left[n_{k-2}n_{k-1}|\underline{a}_k^{k-3}\right]. \end{aligned} \tag{28}$$

Essentially, (27) and (28) are rearranged signal-dependent Yule-Walker equations [3]. We further manipulate them to solve for $c_0\left(\underline{a}_k^{k-3}\right)$ and $c_1\left(\underline{a}_k^{k-3}\right)$. Similar to (26), we find

$$\mathrm{E}\left[n_{k-2}^2|\underline{a}_k^{k-3}\right] = \frac{1}{4}\sum_{a_{k-5}=0}^{1}\sum_{a_{k-4}=0}^{1} c_0\left(\underline{a}_{k-2}^{k-5}\right) \tag{29}$$

$$\mathrm{E}\left[n_{k-2}n_{k-1}|\underline{a}_k^{k-3}\right] = \frac{1}{2}\sum_{a_{k-4}=0}^{1} c_1\left(\underline{a}_{k-1}^{k-4}\right). \tag{30}$$

Substituting (26), (29) and (30) into (27) and (28), and using the definitions of $c_0\left(\underline{a}_k^{k-3}\right)$ and $c_1\left(\underline{a}_k^{k-3}\right)$ in (25)

$$\begin{aligned} c_0\left(\underline{a}_k^{k-3}\right) &= \sigma^2\left(\underline{a}_k^{k-2}\right) + \frac{1}{2}b_1^2\left(\underline{a}_k^{k-2}\right)\sum_{a_{k-4}=0}^{1} c_0\left(\underline{a}_{k-1}^{k-4}\right) \\ &\quad + \frac{1}{4}b_2^2\left(\underline{a}_k^{k-2}\right)\sum_{a_{k-5}=0}^{1}\sum_{a_{k-4}=0}^{1} c_0\left(\underline{a}_{k-2}^{k-5}\right) \\ &\quad + b_1\left(\underline{a}_k^{k-2}\right)b_2\left(\underline{a}_k^{k-2}\right)\sum_{a_{k-4}=0}^{1} c_1\left(\underline{a}_{k-1}^{k-4}\right) \end{aligned} \tag{31}$$

$$\begin{aligned} c_1\left(\underline{a}_k^{k-3}\right) &= \frac{1}{2}b_1\left(\underline{a}_k^{k-2}\right)\sum_{a_{k-4}=0}^{1} c_0\left(\underline{a}_{k-1}^{k-4}\right) \\ &\quad + \frac{1}{2}b_2\left(\underline{a}_k^{k-2}\right)\sum_{a_{k-4}=0}^{1} c_1\left(\underline{a}_{k-1}^{k-4}\right). \end{aligned} \tag{32}$$

Although tedious to derive, (31) and (32), when written for all 16 combinations of four symbols a_{k-3}, a_{k-2}, a_{k-1} and a_k, represent a *linear* system of equations whose solutions $c_0\left(\underline{a}_k^{k-3}\right)$ and $c_1\left(\underline{a}_k^{k-3}\right)$ are straight-forwardly obtained. These solutions give the matrices $\mathbf{C}_1$ needed in (20). The error probability bounds computation follows as in Appendix A and Sections 3 and 4.

References

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

[2] G. D. Forney Jr., "Lower bounds on error probability in the presence of large intersymbol interference," *IEEE Trans. Commun.*, vol. 20, pp. 76–77, Feb. 1972.

[3] A. Kavčić and J. M. F. Moura, "The Viterbi algorithm and Markov noise memory," *IEEE Trans. Inform. Theory*, submitted Nov. 1997. 28 pages, accepted for publication in Jan. 2000 issue.

[4] A. Kavčić and A. Patapoutian, "A signal-dependent autoregressive channel model," *IEEE Trans. Magn.*, vol. 35, September 1999.

[5] K. S. Miller, *Multidimensional Gaussian Distributions*. New York: John Wiley, 1964.

[6] A. Kavčić and J. M. F. Moura, "Gauss-Markov approximations and information loss," tech. rep., Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, March 1998. 30 pages, submitted for publication.

[7] A. Kavčić and J. M. F. Moura, "Correlation structures for optimizing information criteria," in *Proc. IEEE Information Theory Workshop on Detection, Estimation, Classification and Imaging*, (Santa Fe, NM), p. 25, Feb. 1999.

[8] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs: Prentice Hall, 2nd ed., 1991.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore: The Johns Hopkins University Press, 1983.