ELSEVIER

BER analysis of Bayesian equalization using orthogonal hyperplanes

Akrum Elkhazin[*],[1], Konstantinos N. Plataniotis, Subbarayan Pasupathy

The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, 10 King's College Road, Toronto, Canada M5S-3G4

Abstract

Bayesian symbol-by-symbol detection using a finite sequence observation space has been the subject of renewed research interest. The Bayesian transverse equalizer (BTE) and Bayesian decision feedback equalizer (BDFE) are two common Bayesian detectors. It is often difficult to evaluate the bit-error rate (BER) performance of these Bayesian detectors since the BER cannot be analytically evaluated and the high complexity of these detectors makes simulation techniques computationally prohibitive, especially at low BERs. We propose a framework to evaluate the BER for the BTE and a lower bound on the BER for the BDFE. This framework is based on finding an approximation of the conditional error probability for each of the noiseless channel states in the observation space. The optimal Bayesian decision boundary is approximated by a set of hyperplanes, and each hyperplane is rotated some minimal angle to make them mutually orthogonal/parallel. The conditional probability of error can be readily evaluated on the topology of orthogonal/parallel rotated hyperplanes. Our BER evaluation is accurate and does not require simulations. A reduced complexity approach to evaluate the BER is also developed.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Bayesian equalizer; Asymptotic decision boundary; Bit error rate

1. Introduction

Equalization and detection play an integral role in combating distortion and interference in high-speed communication links [1,2] and high-density data storage systems [3,4]. For a class of equalizers based on finite channel observation symbol-by-symbol detection (SBSD), the maximum a posteriori equalizer or Bayesian transverse equalizer (BTE) provides optimal performance [5–7]. For another class of finite channel observation SBSDs with decision feedback, the maximum a posteriori Bayesian decision feedback equalizer (BDFE) provides optimal performance [5,6]. In the case of either the BTE or BDFE, there does not exist a closed form expression for the bit-error rate (BER).

In many communication and data storage systems, it is important to evaluate the BER of the equalization process. In the absence of an analytical formulation, there are several approaches that can be used to evaluate the BER. Using an analytical approximation, an approximate lower bound of the BER for the BTE is presented in [7] and this method

*Corresponding author. Tel.: +1 647 285 7503;
fax: +1 416 978 6845.

E-mail addresses: aelkhazi@comm.utoronto.ca (A. Elkhazin),
kostas@dsp.utoronto.ca (K.N. Plataniotis),
pas@comm.utoronto.ca (S. Pasupathy).

can be easily applied to the BDFE. Simulated results presented in this paper however show that this method does not produce an accurate BER. A direct method to evaluate the BER is through conventional Monte Carlo simulations. Unfortunately, Monte Carlo simulations become computation infeasible at low BERs. This is especially true for channels with long impulse responses since the complexity of either the BTE or BDFE increases exponentially with channel impulse response length. Simulation techniques that employ importance sampling (IS) have been proposed [8,9] to evaluate the BER with fewer simulated bits than conventional Monte Carlo simulations.

In this paper, we develop a method to evaluate the BER for the BTE and BDFE, without simulation. In the case of the BDFE, we make the assumption of correct previous decisions, thus the evaluated BER is a lower bound on the true BER. For both BTE and BDFE, the Bayesian decision boundary is determined by the set of possible noiseless channel outputs or channel states and BER is the average probability of error (P_e) over all channel states. The P_e for a given channel state cannot in general be evaluated analytically. We introduce the orthogonalized decision boundary (ODB) as a means to approximately evaluate the P_e for a given channel state. At asymptotically high SNR, the decision boundary consists of a set planar regions [9,10]. The ODB is formed by rotating the planar regions asymptotic boundary to be orthogonal and/or parallel, forming a hyperprism topology. The P_e can be readily evaluated using Gaussian error functions for the hyperprism topology.

The remainder of this paper is organized as follows. The signal space is defined for both the BTE and BDFE in Section 2 and the asymptotic boundary is described in Section 3. In Section 4, the ODB is developed as a means to evaluate the BER and a reduced complexity ODB (RCODB) is developed in Section 5. The BER performance and computational complexity of the ODB and RCODB methods is compared to other methods in Section 6.

2. Signal space definition

We will assume a real-valued channel with received samples generated by

$$r(k) = \sum_{i=1}^{n_c-1} c_i s(k-i) + n(k), \qquad (1)$$

where $r(k)$ is the observed sample, $s(k)$ is a transmitted symbol, $\{c_i\}$ represent the overall channel response and $n(k)$ is the additive noise taken from a identically independently distributed Gaussian source of variance σ_N^2. The transmitted symbol $s(k)$ can take on values $s_1 = A$, $s_1 = -A$ for some constant A.

The definition of the channel output vector is slightly different for the BTE and BDFE. Accordingly, we use the following two subsections to define parameters relating to the channel output vector.

2.1. BTE definition

For the BTE [5], a channel observation vector $\mathbf{r}(k) = [r(k+d), \ldots, r(k+d-m+1)]^T$ is used to make the kth symbol decision, where d is the decision delay and m is the feedforward order. With no loss of generality, we choose $m = n_c$, $d = n_c - 1$ and express $\mathbf{r}(k)$ in vector form as

$$\mathbf{r}(k) = \mathbf{C}\mathbf{s}(k) + \mathbf{n}(k), \qquad (2)$$

where $\mathbf{s}(k) = [s(k+d), \ldots, s(k-d)]^T$ is the transmitted sequence of symbols that influences $\mathbf{r}(k)$, $\mathbf{n}(k) = [n(k+d), \ldots, n(k)]^T$ is a vector of Gaussian noise samples and $\mathbf{C}$ is a $(m) \times (m - n_c - 1)$

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n_c-1} & 0 & \cdots & 0 \\ 0 & c_0 & \ddots & c_{n_c-2} & c_{n_c-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & c_0 & c_1 & \cdots & c_{n_c-1} \end{bmatrix}. \qquad (3)$$

The vector $\mathbf{s}(k)$ can take on $n_s = 2^{n_c+m-1}$ possible values. Let S denote the set of all possible $\mathbf{s}(k)$ vectors and consider partitioning S into two classes according to the value of $s(k)$

$$S = \bigcup_{1 \leqslant l \leqslant 2} S^{(l)}, \qquad (4)$$

where

$$S^{(l)} \triangleq \{\mathbf{s}(k) | s(k) = s_l\}. \qquad (5)$$

Let $\mathbf{s}_j^{(l)}$, $j \in [1 \cdots 2^{2m-2}]$ be the jth element in $S^{(l)}$ for any ordering of $S^{(l)}$ and let $R^{(l)}$ denote the corresponding set of noiseless channel states given by

$$R^{(l)} = \{\mathbf{r}_j^{(l)} | \mathbf{C}\mathbf{s}_j^{(l)}, \mathbf{s}_j^{(l)} \in S^{(l)}\}. \qquad (6)$$

2.2. BDFE definition

The BDFE [5] uses a channel observation vector $\mathbf{r}(k) = [r(k+d), \ldots, r(k+d-m+1)]^T$ and set of previous decisions $\hat{\mathbf{s}}(k) = [\hat{s}(k-1), \ldots, \hat{s}(k-n_b+2)]$ to make the kth symbol decision, where d, m, and n_b are the decision delay, feedforward and feedback orders, respectively. The choice of $d = n_c - 1$, $m = n_c$, $n_b = n_c - 1$ will be used since this choice is sufficient to guarantee linear separability for different signal classes [5]. With this choice, the channel observation vector can be expressed as

$$\mathbf{r}(k) = \mathbf{C}_1 \mathbf{s}_f(k) + \mathbf{C}_2 \mathbf{s}_b(k) + \mathbf{n}(k), \tag{7}$$

where $\mathbf{s}_f(k) = [s(k+d), \ldots, s(k)]^T$, $\mathbf{s}_b(k) = [s(k-1), \ldots, s(k-d)]^T$, $\mathbf{n}(k) = [n(k+d), \ldots, n(k)]^T$ and

$$\mathbf{C}_1 = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n_c-1} \\ 0 & c_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_1 \\ 0 & \cdots & 0 & c_0 \end{bmatrix},$$

$$\mathbf{C}_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ c_{n_c-1} & 0 & \ddots & \vdots \\ c_{n_c-2} & c_{n_c-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ c_1 & \cdots & c_{n_c-2} & c_{n_c-1} \end{bmatrix} \tag{8}$$

are $(m) \times (d+1)$ and $(m) \times n_b$ channel matrices, respectively. Assuming previous decisions are correct, $\mathbf{s}_b(k)$ can be replaced by an equivalent vector $\hat{\mathbf{s}}_b(k)$ of previous decisions. Thus, the original channel observation space $\mathbf{r}(k)$ can be translated to a new observation space $\mathbf{r}'(k)$ by the following relationship:

$$\mathbf{r}'(k) \triangleq \mathbf{r}(k) - \mathbf{C}_2 \hat{\mathbf{s}}(k) = \mathbf{C}_1 \mathbf{s}_f(k) + \mathbf{n}(k). \tag{9}$$

In the translated channel observation space (9), the BDFE operates like the BTE in (7) with $\mathbf{r}(k)$, $\mathbf{s}(k)$ being replaced by $\mathbf{r}'(k)$, $\mathbf{s}_f(k)$. There are only $n_s = 2^m$ possible noiseless channel states in translated BDFE observation space since $\mathbf{s}_f(k)$ can take on 2^m possible values. Let S denote the set of all possible $\mathbf{s}_f(k)$ vectors and consider partitioning S into two classes according to the value of $s(k)$

$$S = \bigcup_{1 \leqslant l \leqslant 2} S^{(l)}, \tag{10}$$

where

$$S^{(l)} \triangleq \{\mathbf{s}_f(k) | s(k) = s_l\}. \tag{11}$$

Let $\mathbf{s}_j^{(l)}, j \in [1 \cdots 2^m]$ be the jth element in $S^{(l)}$ for any ordering of $S^{(l)}$ and let $R^{(l)}$ denote the corresponding set of noiseless channel states given by

$$R^{(l)} = \{\mathbf{r}_j^{(l)} | \mathbf{C}_1 \mathbf{s}_j^{(l)}, \mathbf{s}_j^{(l)} \in S^{(l)}\}. \tag{12}$$

The signal space definitions for that BDFE in the translated observation space are essentially the same as those for the BTE except that $\mathbf{r}'(k)$, $\mathbf{s}_f(k)$ are used to represent the channel observation vector and transmitted sequence. For simplicity, the notation developed for the BTE will be used in this paper, as the application to the BDFE is straightforward.

3. Asymptotic decision boundary

At asymptotically high SNR, it has been shown in [9] that the decision boundary that separates $R^{(1)}$ and $R^{(2)}$ is formed by a piecewise linear set of hyperplanes. Let D_{asym} denote the decision boundary at an asymptotically high SNR. A necessary condition for a point $\mathbf{r}_B \in D_{\text{asym}}$ to be in the asymptotic boundary is

$$\mathbf{r}_B = \frac{\mathbf{r}^{(1)} + \mathbf{r}^{(2)}}{2} + \left[\frac{\mathbf{r}^{(1)} - \mathbf{r}^{(2)}}{2}\right]^{\perp}, \tag{13}$$

where $\mathbf{r}^{(1)} \in R^{(1)}$, $\mathbf{r}^{(2)} \in R^{(2)}$ and $[\mathbf{x}]^{\perp}$ denotes an arbitrary subspace orthogonal to $[\mathbf{x}]$. The sufficient conditions for $\mathbf{r}_B \in D_{\text{asym}}$ are

$$\|\mathbf{r}_B - \mathbf{r}^{(1)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_j^{(1)}\|^2 \quad \forall \mathbf{r}_j^{(1)} \in R^{(1)}, \ \mathbf{r}_j^{(1)} \neq \mathbf{r}^{(1)},$$

$$\|\mathbf{r}_B - \mathbf{r}^{(2)}\|^2 < \|\mathbf{r}_B - \mathbf{r}_j^{(2)}\|^2 \quad \forall \mathbf{r}_j^{(2)} \in R^{(2)}, \ \mathbf{r}_j^{(2)} \neq \mathbf{r}^{(2)},$$

$$\|\mathbf{r}_B - \mathbf{r}^{(1)}\|^2 = \|\mathbf{r}_B - \mathbf{r}^{(2)}\|^2. \tag{14}$$

If there exist an $\mathbf{r}_B$ satisfying (13), (14), then channel state pair $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ forms a *dominant pair*. To test if a pair $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ is dominant, the linear constraints in (14) can be posed as a linear program [10]. Let R_{asym} denote the set of dominant pairs known as the dominant set and let N_{DS} denote the number of pairs in R_{asym}. For the ith dominant pair in any ordering of the dominant set, let G_i denote the region $\mathbf{r}_B \in D_{\text{asym}}$ for which (13), (14) hold. A region G_i is a bounded subspace of a hyperplane that is

orthogonal to the line segment joining the ith dominant pair.

4. Orthogonalized decision boundary

The formation of the ODB for an arbitrary channel state can be summarized as follows. The first step is to represent each planar region in the asymptotic boundary with a point of local minimum distance (LMD) to the given channel state. For each LMD point, we define an LMD vector as a vector from the channel state to the LMD point and an LMD hyperplane as a hyperplane orthogonal to the LMD vector that includes the LMD point. Each planar region in the asymptotic boundary is approximated by an LMD hyperplane. This approximation is typically exact for planar regions that are at a small Euclidean distance to the given channel state. The set of LMD hyperplanes are rotated about the given channel state some minimum angle so that they are mutually orthogonal and/or parallel. After rotation, there may be several parallel hyperplanes in a particular direction from the given channel state. All but the closest parallel hyperplane in particular direction is omitted, and the remaining set of rotated hyperplanes form the ODB.

For a particular channel state $\mathbf{r}$, let $\mathbf{p}_i \in G_i$ be the point in G_i that is of minimum Euclidean distant to $\mathbf{r}$. For notational simplicity, we have omitted the subscripts and superscripts in $\mathbf{r}_j^{(l)}$. The set of points $\{\mathbf{p}_i : i \in [1, \ldots, N_{\text{DS}}]\}$ are the LMD points in D_{asym} to a channel state $\mathbf{r}$. Each channel state has a different set of LMD points. Associated with an LMD point $\mathbf{p}_i$ is an LMD vector $\mathbf{v}_i = \mathbf{p}_i - \mathbf{r}$ that is the displacement from the channel state $\mathbf{r}$ to the LMD point $\mathbf{p}_i$. The LMD points and vectors may be non-unique since it is possible that an LMD point may be at the boundary of two planar regions of the asymptotic decision boundary. The LMD point for a region G_i that is determined by a dominant pair $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ can be found by solving the following quadratic program (QP):

$$\underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{r} - \mathbf{x}\|^2 \tag{15}$$

constrained by

$$(\mathbf{r}^{(1)} - \mathbf{r}^{(2)})^{\mathrm{T}}\mathbf{x} = (\mathbf{r}^{(1)} - \mathbf{r}^{(2)})^{\mathrm{T}}\left(\frac{\mathbf{r}^{(1)} + \mathbf{r}^{(2)}}{2}\right)^{\mathrm{T}},$$

$$(\mathbf{r}_j^{(1)} - \mathbf{r}^{(1)})^{\mathrm{T}}\mathbf{x} \leqslant \frac{\|\mathbf{r}_j^{(1)}\|^2 - \|\mathbf{r}^{(1)}\|^2}{2},$$

$$(\mathbf{r}_j^{(2)} - \mathbf{r}^{(2)})^{\mathrm{T}}\mathbf{x} \leqslant \frac{\|\mathbf{r}_j^{(2)}\|^2 - \|\mathbf{r}^{(2)}\|^2}{2}. \tag{16}$$

The constraints in (16) come directly from the necessary (13) and sufficient (14) conditions for a pair $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ to be dominant. For each planar region G_i, let H_i be an LMD hyperplane that includes the point $\mathbf{p}_i$ and is orthogonal to the LMD vector $\mathbf{v}_i$. The hyperplane H_i approximates the region G_i about the corresponding LMD point. If one member of the ith dominant pair is the channel state $\mathbf{r}$, then the hyperplane exactly represents the planar region G_i in the locality of $\mathbf{p}_i$ since G_i will be a subset of H_i. This can be seen for the dominant channel state $\mathbf{r}$ in Fig. 1 for the channel $c(z) = 1 + 0.4z^{-1}$. The LMD hyperplanes H_1, H_2 exactly represent G_1, G_2 in the region over which G_1, G_2 exist. The hyperplane H_3 approximately represents the region G_3 about the LMD point $\mathbf{p}_3$, although this hyperplane is superfluous since the other two LMD hyperplanes H_1, H_2 already capture the part of the asymptotic boundary that is close to $\mathbf{r}$. After the rotation process, superfluous hyperplanes will be omitted, as they will become parallel to some other rotated hyperplane that is closer to $\mathbf{r}$.

To form the ODB for a channel state $\mathbf{r}$, the LMD hyperplanes must be rotated about $\mathbf{r}$ so that they are mutually orthogonal/parallel. This is equivalent to rotating the LMD vectors to make them orthogonal
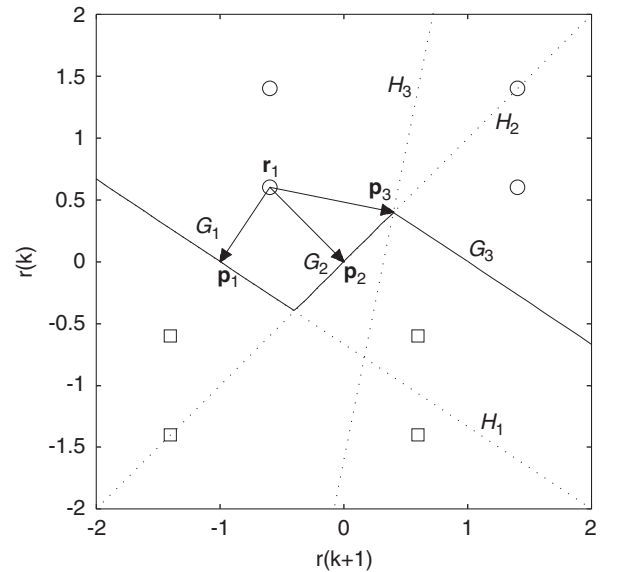

Fig. 1. Asymptotic decision boundary for $c(z) = 1 + 0.4z^{-1}$. Decision boundary near $\mathbf{r}$ exactly represented by hyperplanes H_1, H_2.

or collinear. Let $\hat{N}$ be a set of normal vector given by

$$\hat{N} = \left\{ \{\hat{\mathbf{n}}_1, \ldots, \hat{\mathbf{n}}_n\} : n \leqslant m, \hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_k = \begin{cases} 1, i = k \\ 0, i \neq k \end{cases} \right\}$$
(17)

that form a basis for the rotated LMD vectors. Let $\mathbf{v}'_i$ be a rotation of $\mathbf{v}_i$ by some minimal angle so that $\mathbf{v}'_i$ is collinear with one of the normal vector. A rotated LMD vector $\mathbf{v}'_i$ is given by

$$\mathbf{v}'_i = \left\{ \delta \|\mathbf{v}_i\| \hat{\mathbf{n}}_l : \delta \mathbf{v}_i \cdot \hat{\mathbf{n}}_l = \underset{a,k}{\mathrm{argmax}}\{a\mathbf{v}_i \cdot \hat{\mathbf{n}}\} \right\},$$
(18)

where $\{a, \delta\} \in \{+1, -1\}$ accounts for the fact that each normal vector can define a hyperplane with positive and negative displacements from $\mathbf{r}$. For a rotated LMD vector $\mathbf{v}'_i$, let H'_i be the corresponding rotated LMD hyperplane that is orthogonal to $\mathbf{v}'_i$ and includes the point $(\mathbf{r} + \mathbf{v}'_i)$. In a direction $\delta \hat{\mathbf{n}}_n$ from the channel state $\mathbf{r}$, there may be one or more rotated hyperplanes, but only the closest of these hyperplane is used in the orthogonalized boundary. If there exist a rotated LMD vector such $\delta \mathbf{v}'_i \cdot \hat{\mathbf{n}}_n > 0$, then the ODB will include a hyperplane given by

$$\mathbf{h}_{k,\delta} = \mathbf{r} + d_{k,\delta} \hat{\mathbf{n}}_k + [\hat{\mathbf{n}}_k]^\perp, \quad \delta d_{k,\delta} > 0,$$
(19)

where $d_{k,\delta}$ is the magnitude of the smallest rotated vector in the direction $\delta \hat{\mathbf{n}}_k$ that is given by

$$d_{k,\delta} = \underset{\mathbf{v}'_i}{\mathrm{argmax}} \|\mathbf{v}'_i\|, \quad \delta \mathbf{v}'_i \cdot \hat{\mathbf{n}}_k = \|\mathbf{v}'_i\|.$$
(20)

If there is no $\mathbf{v}'_i$ such that $\delta \mathbf{v}'_i \cdot \hat{\mathbf{n}}_k > 0$, then the ODB will not have a boundary in the direction $\delta \hat{\mathbf{n}}_k$ and we set $d_{k,\delta} = \infty$. Since there will usually be one or more directions that have no boundary, the ODB is in general an unbounded hyperprism. With the ODB, the probability of error for a channel state $\mathbf{r}$ can be approximately evaluated as

$P(\mathrm{error}|\mathbf{r})$

$$\approx 1 - \prod_{i=1}^{n} \left(1 - Q\left(\frac{d_{i,+1}}{\sigma_N}\right) - Q\left(\frac{-d_{i,-1}}{\sigma_N}\right) \right), \quad (21)$$

where $Q(x)$ is defined by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(\frac{-x^2}{2}\right).$$
(22)

The BER is the average P_e over all the channel states in a class $R^{(l)}$ that is given by

$$\mathrm{BER} = \frac{1}{n_s/2} \sum_{j=1}^{n_s/2} P(\mathrm{error}|\mathbf{r}_j^{(l)}),$$
(23)

where $n_s/2$ is the number of channel states in a class $R^{(l)}$. There is no difference between using $R^{(1)}$ or $R^{(2)}$ to evaluate (23) since the channel states in $R^{(1)}$ and $R^{(2)}$ are symmetric about the decision boundary [11]. The choice of the normal vectors $\hat{N}$ can affect the accuracy of the evaluated probability of error in (21). Since the probability of error is largely determined by the G_i regions close to $\mathbf{r}$, it is advantageous to minimize the rotation angle for those LMD vectors with a small magnitude. This can be done by ordering the LMD vectors in increasing magnitude of $\|\mathbf{v}_i\|$ and applying the following greedy algorithm to define $\hat{N}$:

---

1 $\hat{\mathbf{n}}_1 = \mathbf{v}_1/\|\mathbf{v}_1\|$;
2 $n = 1$;
3 **for** $i = 1$ to N_{DS};
4   **if** $\mathbf{v}_i - \sum_{k=1}^{n}(\mathbf{v}_i \cdot \hat{\mathbf{n}}_k)\hat{\mathbf{n}}_k > \|\mathbf{v}_i\|^2/\sqrt{2}$
5     $\hat{\mathbf{n}}_{n+1} = (\mathbf{v}_i - \sum_{k=1}^{n}(\mathbf{v}_i \cdot \hat{\mathbf{n}}_k)\hat{\mathbf{n}}_k)/\|\mathbf{v}_i - \sum_{k=1}^{n}(\mathbf{v}_i \cdot \hat{\mathbf{n}}_k)\hat{\mathbf{n}}_k\|$;
6     $n = n + 1$;
7     $\hat{N} \longleftarrow \hat{\mathbf{n}}_n$;
8   **end if**
9 **end for**

The preceding algorithm incrementally defines the set of normal vectors by considering LMD vectors in increasing order of $\|\mathbf{v}_i\|$. For the ith LMD vector, if the angle between $\mathbf{v}_i$ and the subspace defined by $\hat{N}$ is less than $\pi/2$, then (13) maps $\mathbf{v}_i$ to a normal vector already defined in $\hat{N}$. If this angle is greater than $\pi/2$ then the condition expression on line 4 will be valid and a new normal vector is defined on line 6. This new normal vector makes a smaller angle with $\mathbf{v}_i$ than any existing normal vectors.

The ODB can be thought of as a geometric tool whose purpose is to choose the $d_{k,\delta}$ offsets that are needed for the P_e calculation in (21). Accordingly, the exact orientation of the normal vector is not of critical importance as long as LMD vectors that are separated by a large angular distance are aligned with different normal vectors. Provided that each LMD vector is rotated to the angular direction of the closest normal vector, the magnitude of the angular rotation has a small effect on the accuracy of the probability of error calculation. This can be seen in Fig. 2 for the LMD hyperplanes H_1, H_2.
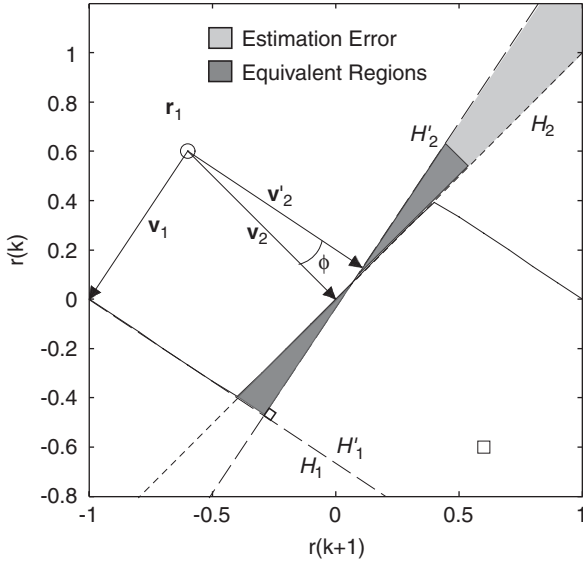
Fig. 2. Depiction of error incurred by rotating hyperplanes.

Since the vectors $\mathbf{v}_1, \mathbf{v}_2$ have a large angular separation, after rotation, they are aligned with different normal vectors. For the purposes of evaluating the probability of error, the decision boundary formed by the rotated hyperplanes H'_1, H'_2 approximately represents the boundary formed by H_1, H_2. The volume under a circular Gaussian pdf centered at $\mathbf{r}$ is the same in the two dark shaded regions. The estimation error incurred in using the rotated hyperplane boundary is shown in the light shaded region. Since the light shaded region is far from $\mathbf{r}$, the volume of the Gaussian pdf in this region is small. As a result, the angle ϕ that the hyperplane H_2 is rotated does not produce a significant error in estimating the probability of error.

5. Reduced complexity orthogonal decision boundary

The ODB method described in the previous section is rather computationally intensive. The bulk of the computational load lies in finding the asymptotic decision boundary and in finding the set of LMD points for each channel state. A channel observation space has n_s noiseless channel state, where $n_s = 2^{2m-1}$ for the BTE and $n_s = 2^m$ for the BDFE. To find the asymptotic boundary, one has to solve a total of $(n_s/2)^2$ linear programs, one for each possible pair of opposite class channel states [10]. Finding the set of LMD points for a particular channel state requires one to solve the QP defined by (15), (16) once for each of the $n_s/2$ LMD points. Since the BER is evaluated as the average prob-

ability of error over $n_s/2$ channel states, a total of $(n_s/2)^2$ quadratic programs must be solved. To avoid solving $(n_s/2)^2$ linear and quadratic programs, we develop a reduce complexity orthogonalized decision boundary (RCODB) method to evaluate the BER. This RCODB method forms an approximate asymptotic decision boundary without solving any linear programs. In addition, the RCODB uses a set of LMD points that are common to all the channel states and these points are found without solving quadratic programs.

To reduce the complexity of the finding of the orthogonal decision boundary, we first define an approximate asymptotic decision boundary. The asymptotic boundary is formed by the set of dominant channels states. For a pair $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ to be dominant, there must exist a point $\mathbf{r}_B = (\mathbf{r}^{(1)} + \mathbf{r}^{(2)})/2 + [(\mathbf{r}^{(1)} - \mathbf{r}^{(2)})/2]^\perp$ that is closer to $\mathbf{r}^{(1)}$ or $\mathbf{r}^{(2)}$ than to any other channel states. This does not imply that the midpoint $\mathbf{m} = (\mathbf{r}^{(1)} + \mathbf{r}^{(2)})/2$ of the line segment between $\mathbf{r}^{(1)}$ and $\mathbf{r}^{(2)}$ is part of asymptotic decision boundary. If $\mathbf{m}$ is not part of the asymptotic boundary then there is one or more other channel states that determine the asymptotic decision in the region that is local to the line segment between $\mathbf{r}^{(1)}$ and $\mathbf{r}^{(2)}$. Conversely, the region of the asymptotic boundary determined by $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ must a relatively large distance from $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ and any other channel states. The regions of the asymptotic boundary that have a large Euclidean distance to every channel state will not significantly affect the accuracy of the evaluated BER since it is unlikely that a channel observation will fall into one of these regions. Accordingly, a reduced dominant set (RDS), denoted $\hat{R}_{\mathrm{asym}}$ is defined as all pairs $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ where $\mathbf{m}$ is the point is closer to $\mathbf{r}^{(1)}$ or $\mathbf{r}^{(2)}$ than any other channel states. The RDS can be found using the following algorithm [11]:

```
1   N_RDS = 0;
2   for r_i^(1) ∈ R^(1);
3       for r_j^(2) ∈ R^(2);
4           x = (r_i^(1) + r_j^(2))/2;  α = ‖r_i^(1) − x‖²;
5           if (‖r_k^(1) − x‖² > α, ∀r_k^(1) ∈ R^(1), k ≠ i) and
                (‖r_k^(2) − x‖² > α, ∀r_k^(2) ∈ R^(2), k ≠ j);
6               N_RDS = N_RDS + 1;
7               R̂_asm ⟵ (r_i^(1), r_j^(2));
8           end if
9       end for
10  end for
```

A set of approximate LMD points can be found directly from the RDS. For a planar region $G \in D_{asym}$ that is determined by a dominant pair $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$, let $\mathbf{p}$ be the LMD point in G with respect to a channel state $\mathbf{r}$ and consider substituting $\mathbf{p}$ with an approximate LMD point

$$\hat{\mathbf{p}} = (\mathbf{r}^{(1)} + \mathbf{r}^{(2)})/2. \qquad (24)$$

The set of approximate LMD points defined in (24) are the same as the set of asymptotically efficient bias vectors used in [8,9]. It is not surprising that the approximate LMD points in (24) can be used to adequately represent asymptotic boundary. The implications of using $\hat{\mathbf{p}}$ to approximate $\mathbf{p}$ can be understood by considering the following four cases:

(1) $\|\mathbf{r} - \mathbf{p}\| < \|\mathbf{r}^{(1)} - \hat{\mathbf{p}}\|$,
(2) $\|\mathbf{r} - \mathbf{p}\| = \|\mathbf{r}^{(1)} - \hat{\mathbf{p}}\|$,
(3) $\|\mathbf{r} - \mathbf{p}\| > \|\mathbf{r}^{(1)} - \hat{\mathbf{p}}\|$,
(4) $\|\mathbf{r} - \mathbf{p}\| \gg \|\mathbf{r}^{(1)} - \hat{\mathbf{p}}\|$.

With no loss of generality, assume $\mathbf{r} \in R^{(1)}$ and that the magnitude difference in case 3 is small. In case 1, $(\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$ does not belong to the RDS and so $\mathbf{p}$ must be a large distance from $\mathbf{r}$ and all other channel states. In this case, $\mathbf{p}$ can be omitted since it will not significantly affect the probability of error calculation. In case 2, $\mathbf{p} = \hat{\mathbf{p}}$ since $\hat{\mathbf{p}}$ is the closest point in G to $\mathbf{r}^{(1)}$ and $\|\mathbf{r}^{(1)} - \mathbf{x}\| \leqslant \|\mathbf{r} - \mathbf{x}\|, \mathbf{x} \in G$. In case 3, we note that the vector $\mathbf{p} - \hat{\mathbf{p}}$ is orthogonal to $\mathbf{r}^{(1)} - \hat{\mathbf{p}}$ which leads to

$$\|\mathbf{p} - \hat{\mathbf{p}}\|^2 = \|\mathbf{r}^{(1)} - \mathbf{p}\|^2 + \|\mathbf{r}^{(1)} - \hat{\mathbf{p}}\|^2 \qquad (25)$$

$$\leqslant \|\mathbf{r} - \mathbf{p}\|^2 + \|\mathbf{r}^{(1)} - \hat{\mathbf{p}}\|^2. \qquad (26)$$

As long as $\|\mathbf{r}^{(1)} - \hat{\mathbf{p}}\|$ is only slightly large than $\|\mathbf{r}^{(1)} - \mathbf{p}\|$, we have $\mathbf{p} \approx \hat{\mathbf{p}}$. In case 4, the error $\|\mathbf{p} - \hat{\mathbf{p}}\|$ in estimating the LMD point $\mathbf{p}$ with $\hat{\mathbf{p}}$ may be large, but this error does not significantly affect the BER since the distance between G and $\mathbf{r}$ is large.

6. Performance comparison

Monte Carlo simulations were performed with three sample channels to evaluate the BER estimation accuracy of the ODB and RCODB methods. The ODB and RCODB methods were also compared to the BER estimation method in [7] and the importance sampling method in [8] both in terms of accuracy and computational complexity. Monte Carlo simulations were performed until at least
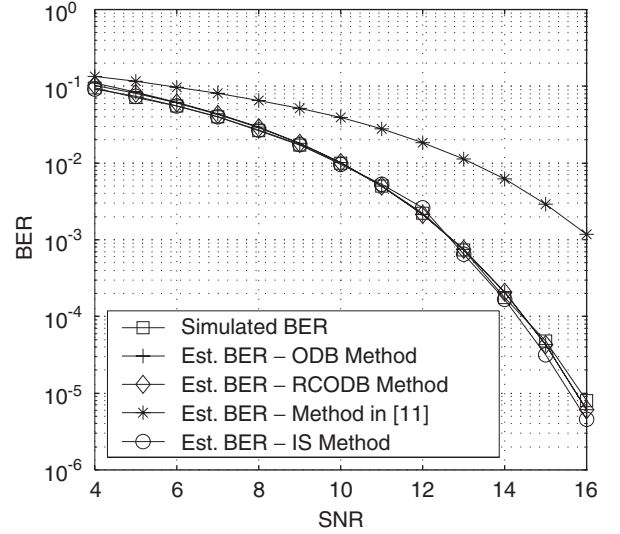

Fig. 3. Comparison of BER evaluation methods for BTE applied to channel $\mathbf{c}(z) = 1 + 0.4z^{-1}$.

200 errors occurred and 10^5 bits were used in the IS simulations.

In the first example, the BTE is applied to a channel with an impulse response specified by

$$\mathbf{c}(z) = 1 + 0.4z^{-1}. \qquad (27)$$

The BTE decision delay and channel observation length were set to $d = 1$ and $m = 2$. The Estimated BERs for the ODB method, RCODB method, IS methods and method in [7] are compared to the BER obtained through simulations in Fig. 3. The method in [7] produced a BER curve that deviated from the simulated BER with increasing SNR. The ODB, RCODB, and IS methods all produced relatively accurate BER curves for moderate-to-high SNR values.

For the second example, the BDFE is applied to channel taken from [5] with channel impulse response given by

$$\mathbf{c}(z) = -0.2052 - 0.5131z^{-1} + 0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4}. \qquad (28)$$

The decision delay and channel observation length were set to $d = 4$ and $m = 5$. The BER curve for the ODB method, RCODB method, IS and the method in [7] is compared to the simulated BER in Fig. 4. The ODB, RCODB and IS methods produced a tight lower bound on the BER. The method in [7] produced a close BER estimate at lower SNR values, but diverged for larger SNR values.
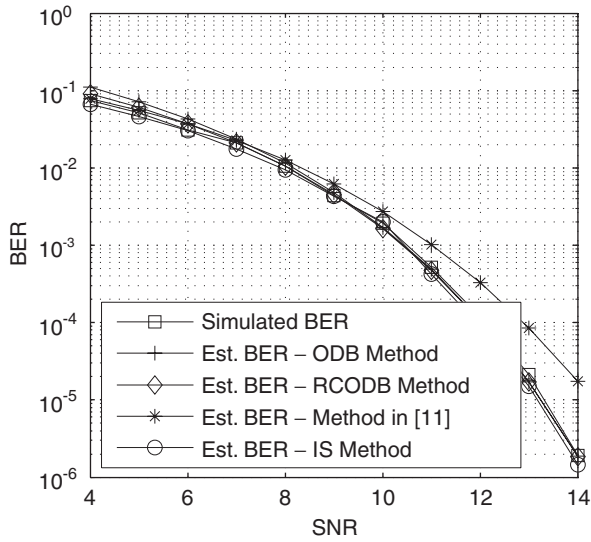
Fig. 4. Comparison of BER evaluation methods for BDFE applied to channel $\mathbf{c}(z) = -0.2052 - 0.5131z^{-1} + 0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4}$.

Fig. 5. Comparison of BER evaluation methods for BDFE applied to channel $\mathbf{c}(z) = 0.35 + 0.8z^{-1} + z^{-2} + 0.8z^{-3}$.

The channel used in the second example was moderately frequency selective. In the third example, we apply the BDFE to a highly frequency selective channel taken from [10,11] with an impulse response specified by

$$\mathbf{c}(z) = 0.35 + 0.8z^{-1} + z^{-2} + 0.8z^{-3}. \tag{29}$$

The decision delay and channel observation length were set to $d = 3$ and $m = 4$. The BER curve for the ODB method, RCODB method, IS and the method in [7] is compared to the simulated BER in Fig. 5. Also shown in Fig. 5 is the simulated BER curve for the BDFE with correct decision feedback. The BER curves for the ODB, RCODB and IS methods closely follow the BER curve for the BDFE with correct decision feedback, while the curve for the method in [7] diverged. There is a significant performance gap between the BER curve for the BDFE with and without correct decision at low SNRs. This gap is the result of error propagate that tends to have a lesser effect at higher SNRs. In comparison to the second example, error propagation has a larger effect on BER performance for highly frequency selective channels.

The computational load of the ODB and RCODB methods is summarized in Table 1, along with the complexity of the IS method and the method in [7]. The number of channel states is $n_s = 2^{2m-1}$ for the BTE and $n_s = 2^m$ for the BDFE. Since the majority of elementary operations in all four
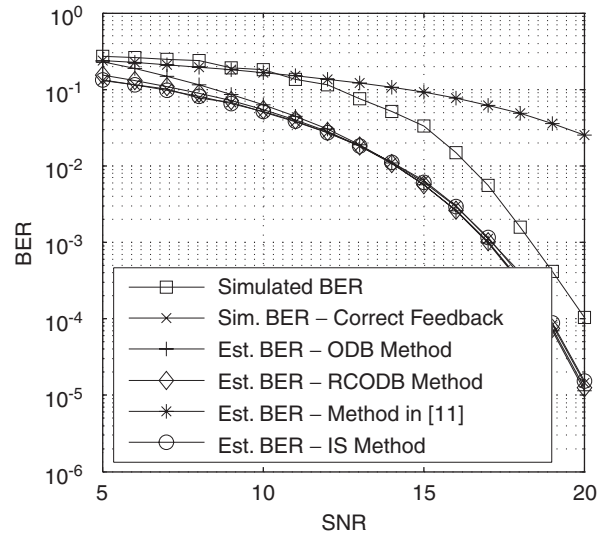
methods are performed on vectors of length m, the number of additions, subtractions and multiplications is specified in vector quantities. The overall complexity of the IS method is comprised of a fixed complexity involved in finding the bias vectors and a variable complexity associated with simulating each transmitted bits. The ODB method has the highest complexity since $(n_s/2)^2$ linear and quadratic programs that must be solved, where each LP and QP has m variables and has $n_s - 1$ linear constraints. The RCODB method has a significantly lower computational load that is dominated by the $(n_s/2) \times n_s$ vector multiplications required to find the LMD points. The fixed complexity of the IS method is slightly less than the complexity of the RCODB. Finally, the method in [7] has the least computational load since the number of elementary vector operations increases with n_s^2.

7. Conclusions

An orthogonalized decision boundary has been proposed as a tool to evaluate the BER for the BTE and a lower bound on the BER for the BDFE. Simulated results have shown that the ODB method produces an accurate BER curve, as accurate as the importance sampling method. A reduced complexity ODB method has also been developed that has complexity that is comparable to the fixed complexity of the IS method. The BER produced by

Table 1
Computational complexity comparison

Orthogonalized decision boundary
Form R_{asym}
LPs $= (n_s/2)^2$
Find LMD points
QPs $= (n_s/2)^2$
Find normal vectors ($n_s/2$ chan. states)
Vect.add./sub. $\leqslant (n_s/2) \times (N_{DS} \times m + m(m+1)/2)$
Vect.multiplications $\leqslant (n_s/2) \times (N_{DS} \times (m+1) + m)$
Rotate LMD vector
Vect.multiplications $\leqslant (n_s/2) \times N_{DS} \times m$
Evaluate BER
$Q(\cdot)$ functions $\leqslant (n_{s/2}) \times (2m)$

Reduced complexity orthogonalized decision boundary
Find LMD points
Vect.add./sub. $= (n_s/2)^2 \times 2 + N_{RDS}$
Vect.multiplications $= (n_s/2)^2 \times n_s + N_{RDS}$
Find normal vectors ($N_s/2$ chan. states)
Vect.add./sub. $\leqslant (n_s/2) \times (N_{DS} \times m + m(m+1)/2)$
Vect.multiplications $\leqslant (n_s/2) \times (N_{RDS} \times (m+1) + m)$
Rotate LMD vectors
Vect.multiplications $\leqslant (n_s/2) \times N_{RDS} \times m$
Evaluate BER
$Q(\cdot)$ functions $\leqslant (n_s/2) \times (2m)$

Importance sampling method overhead
Find bias vectors
Vect.add./sub. $= (n_s/2)^2 \times 2 + N_{RDS}$
Vect.multiplications $= (n_s/2)^2 \times n_s + N_{RDS}$

Method in [7]
Evaluate BER
Vect.add./sub. $= (n_s/2) \times (n_s + 1)$
Vect.multiplications $= (n_s/2) \times (2n_s)$
$\exp(\cdot)'s = (n_s/2) \times n_s$
$\ln(\cdot)'s = n_s/2$
$Q(\cdot)$ functions $= n_s/2$

RCODB method is essentially the same as that of the ODB method for moderate-to-high SNRs.

References

[1] E. Lee, D. Messerschmitt, Digital Communications, second ed., Kluwer, Boston, 1994.

[2] S. Qureshi, Adaptive equalization, Proc. IEEE 73 (September 1985) 1349–1387.

[3] J. Moon, The role of SP in data-storage systems, IEEE Signal Process. Mag. 15 (4) (April 1998) 54–72.

[4] J. Proakis, Equalization techniques for high-density magnetic recording, IEEE Signal Process. Mag. 15 (4) (April 1998) 73–82.

[5] S. Chen, B. Mulgrew, S. McLaughlin, Adaptive Bayesian equalizer with decision feedback, IEEE Trans. Signal Process. 41 (9) (September 1993) 2918–2926.

[6] J. Montalvao, B. Dorizzi, J. Mota, Why use Bayesian equalization based on finite data blocks?, Signal Process. 81 (1) (January 2001) 137–147.

[7] K. Yao, L. Milstein, On ML bit detection of binary signals with intersymbol interference in Gaussian Noise, IEEE Trans. Comm. (September 1975) 971–976.

[8] S. Chen, Importance sampling simulation simulations for evaluating the lower-bound BER of the Bayesian DFE, IEEE Trans. Comm. 50 (2) (February 2000) 179–182.

[9] R.A. Iltis, A randomized bias technique for the importance sampling simulation of Bayesian equalizers, IEEE Trans. Comm. 43 (2) (February–March–April 1995) 1107–1115.

[10] J. Moon, T. Jeon, Sequence detection for binary ISI channels using signal-space partitioning, IEEE Trans. Comm. 46 (7) (July 1998) 891–901.

[11] S. Chen, B. Mulgrew, L. Hanzo, Asymptotic Bayesian decision feedback equalizer using a set of hyperplanes, IEEE Trans. Signal Process. 48 (12) (December 2000) 3493–3500.