Importance Sampling Simulation and Multiple-Hyperplane Realization of the Bayesian Decision Feedback Equaliser

S. Chen and L. Hanzo

Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.
sqc@ecs.soton.ac.uk   lh@ecs.soton.ac.uk

Abstract

For the class of equalisers that employs a symbol-decision finite-memory structure with decision feedback, the optimal solution is known to be the Bayesian decision feedback equaliser (DFE). The complexity of the optimal Bayesian DFE however increases exponentially with the length of the channel impulse response (CIR). It has been noted that, when the signal to noise ratio (SNR) tends to infinity, the decision boundary of the Bayesian DFE is asymptotically piecewise linear and consists of several hyperplanes. This asymptotic property can be exploited for efficient simulation and implementation of the Bayesian DFE. An importance sampling (IS) simulation technique is presented based on this asymptotic property for evaluating the lower-bound bit error rate (BER) of the Bayesian DFE under the assumption of correct decisions being fed back. A design procedure is developed, which chooses appropriate bias vectors for the simulation density to ensure asymptotic efficiency of the IS simulation. As the set of hyperplanes that form the asymptotic Bayesian decision boundary can easily be found, they can be used to partition the observation space. The resulting multiple-hyperplane detector can closely approximate the optimal Bayesian detector, at an advantage of considerably reduced decision complexity.

1   Introduction

Equalisation technique plays an ever-increasing role in combating distortion and interference in communication links [1, 2] and high-density data storage systems [3, 4]. For the class of equalisers based on a symbol-by-symbol decision with

decision feedback, the Bayesian DFE [5, 6, 7] is known to provide the best performance. The complexity of this optimal Bayesian solution, however, increases exponentially with the CIR length, and this limits its practical usefulness. For example, due to its complicated structure, performance analysis of the Bayesian DFE is usually based on conventional Monte Carlo simulation, which is computationally costly even for modest SNR conditions. To obtain a reliable BER estimate, at least 100 errors should occur during a simulation. Thus, for a BER level of 10^{-6}, at least 10^8 data samples are needed. Investigating the Bayesian DFE under BER performance better than 10^{-6} is very difficult if not impossible, using a conventional Monte Carlo simulation. In order for the Bayesian DFE to be more widely adopted in practice, it is also necessary and desired to reduce its implementation complexity without sacrificing performance too much.

Geometrically, the complexity of the Bayesian DFE is a consequence of the need to form the optimal decision boundary that is a hypersurface in the observation space [6]. It can be shown that asymptotically, as the SNR tends to infinity, the Bayesian hypersurface becomes piecewise linear and is made up of a set of hyperplanes [8]. In practice, at large rather than infinite SNR, the performance difference between Bayesian decision boundary and a piecewise linear approximation is negligible. Each of these component hyperplanes is determined by a pair of so-called dominant opposite-class channel states. This asymptotic property can be utilized for various purposes. For instance, in a previous work [9], the Bayesian equalisation solution is approximated by only using the set of the dominant signal state pairs in computation. In this paper, we exploit this asymptotic property to develop an IS simulation technique for performance evaluation of the Bayesian DFE and to implement the Bayesian DFE in a computationally very efficient multiple-hyperplane form.

Iltis [8] developed a randomized bias technique for the IS simulation of Bayesian equalisers without decision feedback. Although it can only guarantee asymptotic efficiency, as defined in [10], for certain channels, this IS simulation technique provides a valuable method in assessing the performance of the Bayesian equaliser. We extend this IS simulation technique to evaluate the lower-bound BER of the Bayesian DFE. By viewing decision feedback as a geometric translation, the Bayesian DFE is "converted" to the Bayesian equalizer in the translated space [11], with a desired property that opposite-class channel states are always linearly separable. A design procedure is developed, which determines the set of hyperplanes that form the asymptotic Bayesian decision boundary and constructs the convex regions associated with individual states by intersecting hyperplanes that are reachable from the states concerned. This provides the appropriate bias vectors for the simulation density to ensure asymptotic efficiency.

A multiple-hyperplane partition technique for equalisation was developed by Kim and Moon [12, 13]. Their design method determines a set of hyperplanes which separate clusters of channel states. A combinatorial search and optimization process is carried out to find these hyperplanes, which is computationally

2

very expensive. The convex regions associated with individual channel states are constructed by appropriately intersecting hyperplanes. The overall decision region is then formed from these convex regions. The decision complexity and performance of the multiple-hyperplane detector are controlled during design by a specified minimum separating distance. Although it is possible to achieve the asymptotic Bayesian solution by an appropriate choice of the minimum separating distance, this is by no means guaranteed as the combinatorial search and optimization process does not necessarily produce the set of hyperplanes which form the asymptotic Bayesian decision boundary. We propose a much simpler alternative design to explicitly realize the asymptotic Bayesian DFE.

2   The Bayesian DFE

We will assume that the channel is real-valued and the received signal sample is given by:

$$y(k) = \sum_{i=0}^{n_a - 1} a_i s(k - i) + e(k) \,, \tag{1}$$

where n_a is the CIR length, a_i are the channel taps, the Gaussian white noise $e(k)$ has zero mean and variance σ_e^2, and the transmitted symbol $s(k)$ takes values from the set $\{\pm 1\}$. A DFE uses the observation vector $\mathbf{y}(k) = [y(k) \cdots y(k - m + 1)]^T$ and the past detected symbol vector $\hat{\mathbf{s}}_b(k) = [\hat{s}(k - d - 1) \cdots \hat{s}(k - d - n)]^T$ to produce an estimate $\hat{s}(k - d)$ of $s(k - d)$. Without the loss of generality, the decision delay of $d = n_a - 1$, feedforward order of $m = n_a$ and feedback order of $n = n_a - 1$ are chosen, as this choice is sufficient to guarantee the linear separability [11]. The received signal vector can be expressed as:

$$\mathbf{y}(k) = \mathbf{F}_1 \mathbf{s}_f(k) + \mathbf{F}_2 \mathbf{s}_b(k) + \mathbf{e}(k) \,, \tag{2}$$

where $\mathbf{s}_f(k) = [s(k) \cdots s(k - d)]^T$, $\mathbf{s}_b(k) = [s(k - d - 1) \cdots s(k - d - n)]^T$, $\mathbf{e}(k) = [e(k) \cdots e(k - m + 1)]^T$, and the $m \times (d + 1)$ and $m \times n$ CIR matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ are, respectively,

$$\mathbf{F}_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a - 1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix} \,, \tag{3}$$

$$\mathbf{F}_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a - 1} & 0 & \ddots & \vdots \\ a_{n_a - 2} & a_{n_a - 1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a - 2} & a_{n_a - 1} \end{bmatrix} \,. \tag{4}$$

3

Assuming correct past decisions, we have $\mathbf{y}(k) = \mathbf{F}_1\mathbf{s}_f(k) + \mathbf{F}_2\hat{\mathbf{s}}_b(k) + \mathbf{e}(k)$. Thus the decision feedback translates the original space $\mathbf{y}(k)$ into a new space:

$$\mathbf{r}(k) \triangleq \mathbf{y}(k) - \mathbf{F}_2\hat{\mathbf{s}}_b(k). \tag{5}$$

Let the $N_f = 2^{d+1}$ sequences of $\mathbf{s}_f(k)$ be $\mathbf{s}_{f,j}$, $1 \le j \le N_f$. The set of the noiseless channel states in the translated space is defined as

$$\mathcal{R} \triangleq \{\mathbf{r}_j = \mathbf{F}_1\mathbf{s}_{f,j}, \ \ 1 \le j \le N_f\}, \tag{6}$$

which can be partitioned into the two subsets conditioned on $s(k-d)$:

$$\mathcal{R}^{(\pm)} \triangleq \{\mathbf{r}_j \in \mathcal{R} : s(k-d) = \pm 1\}. \tag{7}$$

We point out that $\mathcal{R}^{(+)}$ and $\mathcal{R}^{(-)}$ are always linearly separable [11]. The optimal equalisation solution, however, is defined by the Bayesian decision function [6, 7]:

$$
\begin{aligned}
f_B\left(\mathbf{r}(k)\right) \ = \ & \sum_{\mathbf{r}_j^{(+)} \in \mathcal{R}^{(+)}} \exp\left(-\left\|\mathbf{r}(k) - \mathbf{r}_j^{(+)}\right\|^2 / 2\sigma_e^2\right) \\
& - \sum_{\mathbf{r}_j^{(-)} \in \mathcal{R}^{(-)}} \exp\left(-\left\|\mathbf{r}(k) - \mathbf{r}_j^{(-)}\right\|^2 / 2\sigma_e^2\right),
\end{aligned}
\tag{8}
$$

assuming equiprobable states. The decision boundary of this Bayesian DFE

$$\mathcal{D}_B \triangleq \{\mathbf{r} : f_B(\mathbf{r}) = 0\} \tag{9}$$

is generally a hypersurface and cannot be realized by one hyperplane. Let we introduce the following definition. A pair of opposite-class states ($\mathbf{r}^{(+)} \in \mathcal{R}^{(+)}, \mathbf{r}^{(-)} \in \mathcal{R}^{(-)}$) is said to be *dominant* if $\forall \mathbf{r}_j \in \mathcal{R}$, $\mathbf{r}_j \neq \mathbf{r}^{(+)}$, $\mathbf{r}_j \neq \mathbf{r}^{(-)}$:

$$\|\mathbf{r}_j - \mathbf{r}_0\|^2 > \|\mathbf{r}^{(+)} - \mathbf{r}_0\|^2, \tag{10}$$

where $\mathbf{r}_0 = \left(\mathbf{r}^{(+)} + \mathbf{r}^{(-)}\right)/2$. We can now describe the asymptotic Bayesian decision boundary for $\mathrm{SNR} \to \infty$ (or $\sigma_e^2 \to 0$).

Proposition 1 The asymptotic decision boundary $\mathcal{D}_B$ of the Bayesian DFE for large SNR is piecewise linear and made up of a set of L hyperplanes. Each of these hyperplanes is defined by a pair of *dominant* opposite-class states ($\mathbf{r}_l^{(+)} \in \mathcal{R}^{(+)}, \mathbf{r}_l^{(-)} \in \mathcal{R}^{(-)}$), such that the hyperplane is orthogonal to the line connecting the pair of dominant states and passes through the midpoint of the line.

Proof: See [8]. As $\sigma_e^2 \to 0$, a necessary condition for a point $\mathbf{r} \in \mathcal{D}_B$ is

$$\mathbf{r} = \frac{\mathbf{r}_l^{(+)} + \mathbf{r}_l^{(-)}}{2} + \left[\frac{\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}}{2}\right]^{\perp}, \tag{11}$$

4

where $\mathbf{x}^\perp$ denotes an arbitrary vector in the subspace orthogonal to $\mathbf{x}$; and the sufficient conditions for $\mathbf{r} \in \mathcal{D}_B$ are

$$\|\mathbf{r} - \mathbf{r}_l^{(+)}\|^2 < \|\mathbf{r} - \mathbf{r}_i\|^2, \ \forall \mathbf{r}_i \in \mathcal{R}^{(+)}, \ \mathbf{r}_i \neq \mathbf{r}_l^{(+)}, \tag{12}$$

$$\|\mathbf{r} - \mathbf{r}_l^{(-)}\|^2 < \|\mathbf{r} - \mathbf{r}_j\|^2, \ \forall \mathbf{r}_j \in \mathcal{R}^{(-)}, \ \mathbf{r}_j \neq \mathbf{r}_l^{(-)}, \tag{13}$$

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

Proposition 1 follows as a direct consequence. The set of all the dominant state pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}_{l=1}^L$ can easily be determined using a simple algorithm based on the conditions (11)–(14) [8, 9].

3 IS simulation method

An excellent introduction to the IS method can be found in [14]. Since the Bayesian DFE is reduced to the Bayesian equalizer in the translated space, the IS simulation technique of [8] can be extended to evaluate its lower-bound BER under the condition of correct bits being fed back, which is given by:

$$\hat{P}_e = \frac{1}{N_s} \frac{1}{N_k} \sum_{i=1}^{N_s} \sum_{k=1}^{N_k} I_E(\mathbf{r}_i(k)) \frac{p(\mathbf{r}_i(k)|\mathbf{r}_i)}{p^*(\mathbf{r}_i(k)|\mathbf{r}_i)}, \tag{15}$$

where the indicator function $I_E(\mathbf{r}(k)) = 1$ if $\mathbf{r}(k)$ causes an error, and $I_E(\mathbf{r}(k)) = 0$ otherwise; $p(\mathbf{r}_i(k)|\mathbf{r}_i)$ is the true conditional density given $\mathbf{r}_i \in \mathcal{R}^{(+)}$, and $N_s = 2^d$ is the number of states in $\mathcal{R}^{(+)}$; the sample $\mathbf{r}_i(k)$ is generated using the simulation density $p^*(\mathbf{r}_i(k)|\mathbf{r}_i)$ chosen to be

$$p^*(\mathbf{r}_i(k)|\mathbf{r}_i) = \sum_{j=1}^{L_i} p_{j,i} \frac{1}{(2\pi\sigma_e^2)^{\frac{m}{2}}} \exp\left(-\frac{\|\mathbf{r}_i(k) - \mathbf{v}_{j,i}\|^2}{2\sigma_e^2}\right). \tag{16}$$

In (16), L_i is the number of the bias vectors $\mathbf{c}_{j,i} = -\mathbf{r}_i + \mathbf{v}_{j,i}$ for $\mathbf{r}_i \in \mathcal{R}^{(+)}$, $p_{j,i} \geq 0$ for $1 \leq j \leq L_i$, and $\sum_{j=1}^{L_i} p_{j,i} = 1$. An estimate of the IS gain, which is defined as the ratio of the numbers of trials required for the same estimate variance using the Monte Carlo and IS methods, is given in [8]. To achieve asymptotic efficiency, $\{\mathbf{c}_{j,i}\}$ must meet certain conditions [10]. We present the following procedure of constructing $p^*(\mathbf{r}_i(k)|\mathbf{r}_i)$ to meet these conditions.

Each of the L dominant state pairs $\{\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)}\}$ defines a hyperplane $H_l(\mathbf{r}) = \mathbf{w}_l^T \mathbf{r} + b_l = 0$. The weight vector $\mathbf{w}_l$ and bias b_l of the hyperplane are given by:

$$\mathbf{w}_l = \frac{2\left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\right)}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}, \quad b_l = -\frac{(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)})^T(\mathbf{r}_l^{(+)} + \mathbf{r}_l^{(-)})}{\|\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)}\|^2}. \tag{17}$$

Note that the theory of support vector machines [15, 16] has been applied to determine H_l with $(\mathbf{r}_l^{(+)}, \mathbf{r}_l^{(-)})$ as its two support vectors, and H_l is a *canonical* hyperplane having the property $H_l(\mathbf{r}_l^{(+)}) = 1$ and $H_l(\mathbf{r}_l^{(-)}) = -1$.

A state $\mathbf{r}_i \in \mathcal{R}$ is said to be *sufficiently separable* by the hyperplane H_l, if H_l can separate $\mathbf{r}_i$ correctly with $|\mathbf{w}_l^T \mathbf{r}_i + b_l| \geq 1$. Thus, if $\mathbf{w}_l^T \mathbf{r}_i^{(+)} + b_l \geq 1$ for $\mathbf{r}_i^{(+)} \in \mathcal{R}^{(+)}$, $\mathbf{r}_i^{(+)}$ is sufficiently separable by H_l and a separability index $h_{l,i}^{(+)}$ is set to 1; otherwise $h_{l,i}^{(+)} = 0$. Similarly, if $\mathbf{r}_i^{(-)} \in \mathcal{R}^{(-)}$ satisfies $\mathbf{w}_l^T \mathbf{r}_i^{(-)} + b_l \leq -1$, it is sufficiently separable by H_l and $h_{l,i}^{(-)} = 1$; otherwise $h_{l,i}^{(-)} = 0$. The *reachability* of H_l from $\mathbf{r}_i^{(+)} \in \mathcal{R}^{(+)}$ can be tested by computing

$$\mathbf{c}_{l,i} = -0.5 \left(\mathbf{w}_l^T \mathbf{r}_i^{(+)} + b_l \right) \left(\mathbf{r}_l^{(+)} - \mathbf{r}_l^{(-)} \right). \tag{18}$$

If $\mathbf{v}_{l,i} = \mathbf{r}_i^{(+)} + \mathbf{c}_{l,i} \in \mathcal{D}_B$, H_l is said to be reachable from $\mathbf{r}_i^{(+)}$ ($\mathbf{c}_{l,i}$ is then a bias vector), and the reachability index is $\gamma_{l,i} = 1$; otherwise $\gamma_{l,i} = 0$. The process produces the following separability and reachability table:

	$\mathbf{r}_1^{(-)}$	$\cdots$	$\mathbf{r}_{N_s}^{(-)}$	$\mathbf{r}_1^{(+)}$	$\cdots$	$\mathbf{r}_{N_s}^{(+)}$
H_1	$h_{1,1}^{(-)}$	$\cdots$	$h_{1,N_s}^{(-)}$	$h_{1,1}^{(+)} \ (\gamma_{1,1})$	$\cdots$	$h_{1,N_s}^{(+)} \ (\gamma_{1,N_s})$
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
H_L	$h_{L,1}^{(-)}$	$\cdots$	$h_{L,N_s}^{(-)}$	$h_{L,1}^{(+)} \ (\gamma_{L,1})$	$\cdots$	$h_{L,N_s}^{(+)} \ (\gamma_{L,N_s})$

In order to construct a convex region $\mathbf{R}_i^{(+)}$ for $\mathbf{r}_i^{(+)} \in \mathcal{R}^{(+)}$, we select those hyperplanes that can *sufficiently* separate $\mathbf{r}_i^{(+)}$ and that are *reachable* from $\mathbf{r}_i^{(+)}$ with the aid of the above table. This yields the following integer set:

$$G_i^{(+)} \triangleq \{ j : \ h_{j,i}^{(+)} = 1 \text{ and } \gamma_{j,i} = 1 \}. \tag{19}$$

Then $\mathbf{R}_i^{(+)}$ is the intersection of all the half-spaces $\mathcal{H}_j^{(+)} \triangleq \{ \mathbf{r} : H_j(\mathbf{r}) \geq 0 \}$ with $j \in G_i^{(+)}$. In fact, it is not necessary to use every hyperplanes defined in $G_i^{(+)}$ to construct $\mathbf{R}_i^{(+)}$. A subset of these hyperplanes will be sufficient, provided that every opposite-class state in $\mathcal{R}^{(-)}$ can sufficiently be separated by at least one hyperplane in the subset. If such a $G_i^{(+)}$ exists for each $\mathbf{r}_i^{(+)}$, the simulation density constructed with the bias vectors $\{\mathbf{c}_{j,i}\}$, $j \in G_i^{(+)}$, will achieve asymptotic efficiency, since all the hyperplanes defined in $G_i^{(+)}$ are reachable from $\mathbf{r}_i^{(+)}$ and obviously at least one of $\{\mathbf{v}_{j,i}\}$ is the minimum rate point (as defined in [10]), and the error region $\mathcal{E}$ satisfies

$$\mathcal{E} \subset \overline{\mathbf{R}_i^{(+)}} \triangleq \bigcup_{j \in G_i^{(+)}} \mathcal{H}_j^{(-)} \tag{20}$$

Table 1: The separability and reachability table for the CIR of $\mathbf{a} = [-0.8\ 1.0\ -0.5]^T$. The DFE structure is defined by $m = 3$, $d = 2$ and $n = 2$.

	$\mathbf{r}_1^{(-)}$	$\mathbf{r}_2^{(-)}$	$\mathbf{r}_3^{(-)}$	$\mathbf{r}_4^{(-)}$	$\mathbf{r}_1^{(+)}$	$\mathbf{r}_2^{(+)}$	$\mathbf{r}_3^{(+)}$	$\mathbf{r}_4^{(+)}$
H_1	1	1	0	1	0	0	1 (1)	0
H_2	1	0	1	1	1 (1)	1 (1)	0	1 (1)
H_3	1	1	1	1	0	1 (1)	0	0
H_4	0	1	0	0	1 (1)	0	1 (0)	1 (1)
H_5	0	0	1	0	1 (1)	1 (1)	1 (1)	1 (1)

with the half-spaces $\mathcal{H}_j^{(-)} \triangleq \{\mathbf{r} : H_j(\mathbf{r}) < 0\}$.

An example. The IS technique for the Bayesian DFE was simulated using the 3-tap CIR defined by $\mathbf{a} = [-0.8\ 1.0\ -0.5]^T$. The bias vectors were generated using the procedure described above. As in [8], the bias vectors were selected with uniform probability in the simulation. For all the cases, 10^5 iterations were employed at each SNR, averaging over all the possible states in $\mathcal{R}^{(+)}$. Since the channel had a length of $n_a = 3$, the DFE structure was specified by $m = 3$, $d = 2$ and $n = 2$. The asymptotic decision boundary consisted of 5 hyperplanes. Table 1 gives the separability and reachability table for this channel.
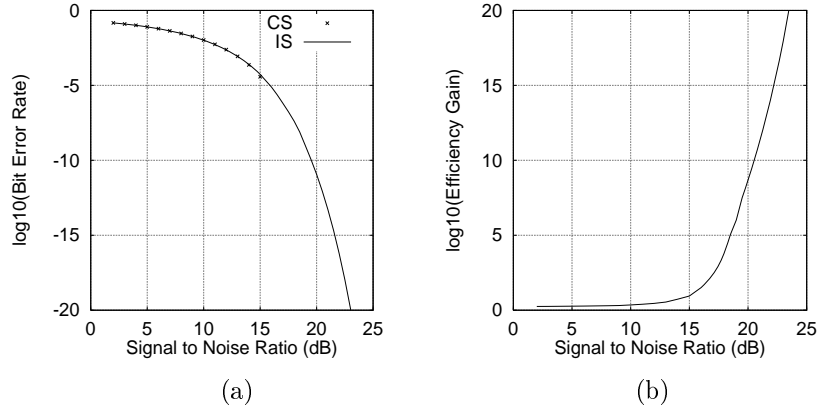

Figure 1: The lower-bound BERs (a) and the IS gain (b) of the Bayesian DFE for the CIR of $\mathbf{a} = [-0.8\ 1.0\ -0.5]^T$ using conventional sampling (CS) and importance sampling (IS) simulation. The DFE structure is defined by $m = 3$, $d = 2$ and $n = 2$.

The states $\mathbf{r}_1^{(+)}$ and $\mathbf{r}_4^{(+)}$ require the two hyperplanes H_2 and H_4 to separate them from all the opposite-class states, and H_2 and H_4 are reachable from the both states. Thus, there are two bias vectors for $\mathbf{r}_1^{(+)}$ and $\mathbf{r}_4^{(+)}$, respectively, and $\mathcal{E} \subset \mathcal{H}_2^{(-)} \bigcup \mathcal{H}_4^{(-)}$. The state $\mathbf{r}_2^{(+)}$ is separated from $\mathcal{R}^{(-)}$ by the single reachable

hyperplane H_3. Thus, there exists one bias vector for $\mathbf{r}_2^{(+)}$ and $\mathcal{E} \subset \mathcal{H}_3^{(-)}$. The state $\mathbf{r}_3^{(+)}$ is separated from $\mathcal{R}^{(-)}$ by the two reachable hyperplanes H_1 and H_5, there are two bias vectors for $\mathbf{r}_3^{(+)}$ and $\mathcal{E} \subset \mathcal{H}_1^{(-)} \bigcup \mathcal{H}_5^{(-)}$. Asymptotic efficiency of the IS simulation is therefore guaranteed for this example.

Fig. 1 (a) shows the lower-bound BERs obtained using the IS and conventional sampling (CS) simulation methods, respectively. It can be seen that the conventional Monte Carlo results for low SNR conditions agreed with those of the IS simulation. The estimated IS gains, depicted in Fig. 1 (b), indicate that exponential IS gains were obtained with increasing SNRs. For SNR=20 dB, the BER of the Bayesian DFE with correct bits being fed back calculated by the IS technique is 1.2×10^{-11}. The CS method could not work under the same SNR condition and, to achieve the same BER estimation accuracy, it would require approximately 4.8×10^8 times of the samples needed by the IS simulation method. As the IS method used 4×10^5 data samples, the CS method would require approximately 2×10^{14} samples to achieve a similar estimation variance.
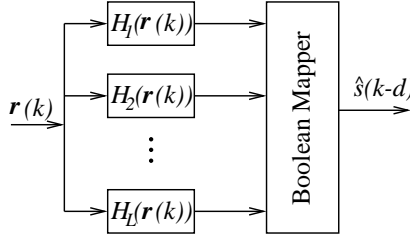

Figure 2: Multiple-hyperplane detector for realizing the asymptotic Bayesian DFE.

4   Multiple-hyperplane detector

Since the set of the L hyperplanes that form the asymptotic Bayesian decision boundary can easily be obtained, they can be used for partitioning the observation space to form a multiple-hyperplane detector which has a structure as depicted in Fig. 2. To construct such a detector, there is no need to test whether a hyperplane H_l is reachable from each state in $\mathcal{R}^{(+)}$ and only a separability table is required. To construct a convex region $\mathbf{R}_i^{(+)}$ covering $\mathbf{r}_i^{(+)} \in \mathcal{R}^{(+)}$, select hyperplanes which can sufficiently separate $\mathbf{r}_i^{(+)}$ from the separability table and denote

$$\tilde{G}_i^{(+)} \stackrel{\triangle}{=} \{l : \ h_{l,i}^{(+)} = 1\}. \tag{21}$$

Then $\mathbf{R}_i^{(+)}$ is obtained by the intersection of all the $\mathcal{H}_j^{(+)}$ with $j \in \tilde{G}_i^{(+)}$

$$\mathbf{R}_i^{(+)} = \bigcap_{j \in \bar{G}_i^{(+)}} \mathcal{H}_j^{(+)}. \tag{22}$$

Table 2: Comparison of decision complexity for the full Bayesian and multiple-hyperplane detectors. L (usually $\ll 2^{n_a}$) is the number of hyperplanes, and n_a is the CIR length. The DFE structure is chosen to be $m = n_a$, $d = n_a - 1$ and $n = n_a - 1$.

	Bayesian DFE	Multiple-hyperplane detector
Multiplications	$(n_a + 1) \times 2^{n_a}$	$n_a \times L$
Additions	$n_a \times 2^{n_a+1} - 1$	$n_a \times L$
Others	$2^{n_a} \exp(\cdot)$ function evaluations	logic ANDs $\leq 2^{n_a-1}$ a logic OR

Again, a subset of the hyperplanes defined by (21) is enough in the construction of $\mathbf{R}_i^{(+)}$, provided that every state in $\mathcal{R}^{(-)}$ can sufficiently be separated by at least one hyperplane in the subset. The overall decision region $\mathbf{R}^{(+)}$ associated with the decision $\hat{s}(k - d) = 1$ is simply formed as the union of all the $\mathbf{R}_i^{(+)}$

$$\mathbf{R}^{(+)} = \bigcup_{i=1}^{N_s} \mathbf{R}_i^{(+)} . \tag{23}$$

The resulting multiple-hyperplane detector is now completely defined. Let a threshold detector output $\beta_j(\mathbf{r}(k))$ for a linear discriminant function $H_j(\mathbf{r}(k))$ have Boolean logic value 1 or 0 depending on $\mathbf{r}(k) \in \mathcal{H}_j^{(+)}$ or not. A Boolean logic value $\theta_i^{(+)}(\mathbf{r}(k))$ indicating whether $\mathbf{r}(k) \in \mathbf{R}_i^{(+)}$ or not is obtained via a logic AND operation of $\{\beta_j(\mathbf{r}(k)) : j \in \tilde{G}_i^{(+)}\}$. A Boolean logic value indicating whether $\mathbf{r}(k) \in \mathbf{R}^{(+)}$ (that is, $\hat{s}(k - d) = 1$) or not is obtained via a logic OR operation of $\{\theta_i^{(+)}(\mathbf{r}(k))\}$ for all i. This detector achieves asymptotically the optimal Bayesian performance since it realizes exactly the asymptotic Bayesian decision boundary. Table 2 compares decision complexity for the full Bayesian DFE and the multiple-hyperplane detector. The multiple-hyperplane detector generally has much simpler decision complexity than the full Bayesian detector, since usually $L \ll N_f$.

An example. The CIR was given by $\mathbf{a} = [0.4\ 0.7\ 0.4]^T$. The structure parameters of the DFE were set to $m = 3$, $d = 2$ and $n = 2$. The asymptotic decision boundary consisted of 5 hyperplanes. Table 3 gives the separability table for this channel. The state $\mathbf{r}_1^{(+)}$ requires the two hyperplanes H_1 and H_2 to be separated from all the opposite-class states $\mathcal{R}^{(-)}$ and, therefore, the convex region $\mathbf{R}_1^{(+)}$ for $\mathbf{r}_1^{(+)}$ is the intersection of the two half-spaces $\mathcal{H}_1^{(+)}$ and $\mathcal{H}_2^{(+)}$. The states $\mathbf{r}_2^{(+)}$ and $\mathbf{r}_3^{(+)}$ are separated from $\mathcal{R}^{(-)}$ by the two hyperplanes H_3 and H_4. Thus $\mathbf{R}_2^{(+)} = \mathbf{R}_3^{(+)}$ is the intersection of the half-spaces $\mathcal{H}_3^{(+)}$ and $\mathcal{H}_4^{(+)}$. The state $\mathbf{r}_4^{(+)}$ is separated by the single hyperplane H_5 from all the opposite-class states, and the convex region $\mathbf{R}_4^{(+)}$ for $\mathbf{r}_4^{(+)}$ is the half-space $\mathcal{H}_5^{(+)}$ defined by H_5. The overall decision region $\mathbf{R}^{(+)}$ is the union of $\mathbf{R}_1^{(+)}$, $\mathbf{R}_2^{(+)}$ and $\mathbf{R}_4^{(+)}$.

9

Table 3: The separability table for the CIR of $\mathbf{a} = [0.4\ 0.7\ 0.4]^T$. The DFE structure is defined by $m = 3$, $d = 2$ and $n = 2$.

	$\mathbf{r}_1^{(-)}$	$\mathbf{r}_2^{(-)}$	$\mathbf{r}_3^{(-)}$	$\mathbf{r}_4^{(-)}$	$\mathbf{r}_1^{(+)}$	$\mathbf{r}_2^{(+)}$	$\mathbf{r}_3^{(+)}$	$\mathbf{r}_4^{(+)}$
H_1	1	0	0	0	1	1	1	1
H_2	0	1	1	1	1	0	0	0
H_3	1	1	1	0	0	1	1	1
H_4	0	0	0	1	1	1	1	0
H_5	1	1	1	1	0	0	0	1

The resulting 5-hyperplane detector requires 15 multiplications and 15 additions to detect a symbol, compared with 32 multiplications, 47 additions and 8 $\exp(\cdot)$ evaluations required by the full Bayesian DFE. The BERs of this multiple-hyperplane detector are compared with those of the full Bayesian DFE in Fig. 3, under different SNR conditions. The BER results were obtained with detected symbols being fed back. It can be seen from Fig. 3 that there exists hardly any BER performance difference between the two equalisers for this channel.
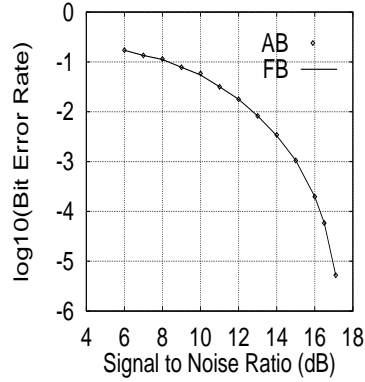

Figure 3: Performance comparison of the multiple-hyperplane detector (AB: points) and the full Bayesian DFE (FB: solid line) with detected symbols being fed back for the CIR of $\mathbf{a} = [0.4\ 0.7\ 0.4]^T$. The DFE structure is defined by $m = 3$, $d = 2$ and $n = 2$.

5   Conclusions

An asymptotic property of the optimal Bayesian decision boundary has been utilized for efficient IS simulation and implementation of the Bayesian DFE. In the first application, we have extended the randomized bias technique for IS

simulation of [8] to evaluate the lower-bound BER of the Bayesian DFE. A design procedure has been presented for constructing the simulation density that meets the asymptotic efficiency conditions. Although asymptotic efficiency of the IS simulation for the general channel has not rigorously been proven, we are unable to find a counter example suggesting that the asymptotic efficiency conditions are not met. The more difficult problem of how to derive an upper-bound BER of the Bayesian DFE, taking into account error propagation, remains an open question and is still under investigation. In the second application, the set of hyperplanes that form the asymptotic Bayesian decision boundary is used to partition the observation space. The resulting multiple-hyperplane detector is guaranteed to achieve asymptotically the optimal Bayesian performance and has a much lower decision complexity compared with the full Bayesian DFE.

References

[1] Qureshi, S.U.H. (1985). Adaptive equalization. *Proc. IEEE*, **73** (9), pp.1349–1387.

[2] Proakis, J.G. (1995). *Digital Communications*, 3rd edition, McGraw-Hill: New York.

[3] Moon, J. (1998). The role of SP in data-storage systems. *IEEE Signal Processing Magazine*, **15** (4), pp.54–72.

[4] Proakis, J.G. (1998). Equalization techniques for high-density magnetic recording. *IEEE Signal Processing Magazine*, **15** (4), pp.73–82.

[5] Abend, K. and Fritchman, B.D. (1970). Statistic detection for communication channels with intersymbol interference. *Proc. IEEE*, **58** (5), pp.779–785.

[6] Chen, S., Mulgrew, B. and McLaughlin, S. (1993). Adaptive Bayesian equaliser with decision feedback. *IEEE Trans. Signal Processing*, **41** (9), pp.2918–2927.

[7] Chen, S., McLaughlin, S., Mulgrew, B. and Grant, P.M. (1995). Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels. *IEEE Trans. Communications*, **43** (5), pp.1937–1946.

[8] Iltis, R.A. (1995). A randomized bias technique for the importance sampling simulation of Bayesian equalizers. *IEEE Trans. Communications*, **43** (2/3/4), pp.1107–1115.

[9] Chng, E.S., Mulgrew, B., Chen, S. and Gibson, G. (1995). Optimum lag and subset selection for radial basis function equaliser. *Proc. 5th IEEE Workshop Neural Networks for Signal Processing* (Cambridge, USA), Aug.31-Sept.2, pp.593–602.

[10] Sadowsky, J.S. and Bucklew, J.A. (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Information Theory*, **36** (3), pp.579–588.

[11] Chen, S., Mulgrew, B., Chng, E.S. and Gibson, G. (1998). Space translation properties and the minimum-BER linear-combiner DFE. *IEE Proc. Communications*, **145** (5), pp.316–322.

[12] Kim, Y. and Moon, J. (1998). Delay-constrained asymptotically optimal detection using signal-space partitioning. *Proc. IEEE Int. Conf. Communications* (Atlanta, USA).

[13] Kim, Y. and Moon, J. (2000). Multi-dimensional signal space partitioning using a minimal set of hyperplanes for detecting ISI-corrupted symbols. *IEEE Trans. Communications*, to appear.

[14] Smith, P.J., Shafi, M. and Gao, H. (1997). Quick simulation: a review of importance sampling techniques in communications systems. *IEEE J. Selected Areas in Communications*, **15** (4), pp.597–613.

[15] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*, Springer-Verlag: New York.

[16] Chen, S., Gunn, S. and Harris, C.J. (2000). Decision feedback equalizer design using support vector machines. *IEE Proc. Vision, Image and Signal Processing*, **147** (3), pp.213–219.