Space translation properties and the minimum-BER linear-combiner DFE

S. Chen
B. Mulgrew
E.S. Chng
G.J. Gibson

Abstract: Decision feedback in a decision feedback equaliser (DFE) performs a space translation that maps the DFE onto a transversal equaliser in the translated observation space. Properties of DFEs can therefore be analysed more easily by exploiting this geometric translation property. This approach is used to analyse the conventional DFE that employs a linear combination of the channel observations and the past decisions (the linear-combiner DFE). It is demonstrated that the usual minimum mean square error (MMSE) solution does not achieve the full performance potential of the linear-combiner DFE structure. A bit error rate (BER) expression for the linear-combiner DFE with binary signalling is obtained, and a method is proposed to optimally set the coefficients of the linear-combiner DFE. The performance of this minimum-BER (MBER) linear-combiner DFE is much closer to that of the optimal Bayesian DFE, compared with the MMSE linear-combiner DFE.

1 Introduction

Decision feedback is a powerful technique for combating intersymbol interference (ISI) distortion. The conventional DFE [1] is based on a symbol-decision structure that employs a linear combination of the channel observations and the past decisions. We call this DFE the linear-combiner DFE, in contrast to other DFE structures that use nonlinear combinations of the channel observations and the past decisions [2–6]. One advantage of the linear-combiner DFE is its computational simplicity. The optimal solution for the symbol-decision structure with feedback is the Bayesian DFE [4]. The adaptive Bayesian DFE has been shown

to outperform the adaptive maximum likelihood sequence estimator (MLSE) for severely fading mobile channels [7]. This is because the MLSE [8], although offering the best solution for equalisation under stationary conditions, can suffer from the drawback of accumulating channel tracking errors under highly non-stationary environment. The Bayesian DFE can be viewed as a special case of the Bayesian sequence estimation [3, 9].

Previous research has demonstrated that decision feedback in a DFE translates the channel observation space (e.g. [3, 10]). In this paper we further investigate this geometric translation property and derive the explicit formula for performing this space translation. Viewed from the translated observation space, a DFE becomes a simpler transversal equaliser. Many properties of DFEs can therefore be analysed more easily by considering their equivalent forms on the translated space. For example, by adopting this geometric approach, a concise form of the Bayesian DFE has been developed [11] which has certain advantages over the original form of the Bayesian DFE given in [4].

The Wiener or MMSE solution is often said to provide the optimal solution for the coefficients of the linear-combiner DFE. An elegant and rigorous analysis of the MMSE linear-combiner DFE is given in [12]. It is well known however that the MMSE solution does not necessarily correspond to the MBER solution, the BER being the ultimate performance criterion of equalisation. Using the geometric translation approach it becomes obvious that the subsets of the translated channel states corresponding to different decisions are linearly separable. The linear-combiner DFE realises a linear decision boundary in the translated observation space. In the asymptotic case, where the signal-to-noise ratio (SNR) tends to infinity, the hyperplanes of the Wiener decision boundary are orthogonal to the last axis of the translated observation space. We demonstrate that the best or optimal linear decision boundary can be very different from the decision boundary of the Wiener solution.

Since the MMSE solution does not achieve the full performance potential of the linear-combiner DFE structure, a substantial BER reduction over the MMSE solution is possible by searching for a better solution of the linear-combiner DFE. We derive a BER expression for the linear-combiner DFE with binary signalling. Using this BER estimator as the optimisation criterion, a method is proposed to optimally adjust the coefficients of the linear-combiner DFE. The decision

boundary of this MBER linear-combiner DFE is the best linear approximation to the nonlinear Bayesian decision boundary. Adaptive implementation of this MBER linear-combiner DFE is also discussed. A drawback of the MBER linear-combiner DFE is that the computational complexity increases significantly when extending to the multilevel signalling case.

Throughout this study the channel and the symbol constellation are assumed to be real valued. This corresponds to the use of multilevel pulse amplitude modulation (M-PAM) scheme. For the complex-valued channel and modulation schemes, the results of the current study are still valid [5]. Specifically, the channel is modelled as a finite impulse response filter with the transfer function

$$A(z) = \sum_{i=0}^{n_a-1} a_i z^{-i} \qquad (1)$$

where n_a is the length of the channel impulse response and a_i are the channel tap weights. The symbol sequence $\{s(k)\}$ is independently identically distributed (IID) and has an M-PAM constellation defined by the set

$$s_i = 2i - M - 1, \quad 1 \le i \le M \qquad (2)$$

The received signal is given by

$$r(k) = \bar{r}(k) + e(k) = \sum_{i=0}^{n_a-1} a_i s(k-i) + e(k) \qquad (3)$$

where $\bar{r}(k)$ is the noiseless channel observation, $e(k)$ is an IID gaussian noise source with zero mean and variance $E[e^2(k)] = \sigma_e^2$ and is uncorrelated with $s(k)$, and $E[\cdot]$ denotes the expectation operator. The SNR of the system is defined as

$$SNR = E\left[\bar{r}^2(k)\right] / E\left[e^2(k)\right] = \sigma_s^2 \left(\sum_{i=0}^{n_a-1} a_i^2\right) / \sigma_e^2 \qquad (4)$$

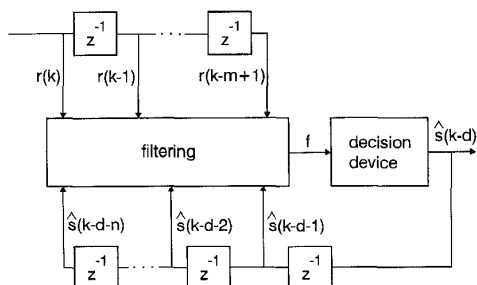where $\sigma_s^2 = E[s^2(k)]$ is the symbol variance.

Fig. 1 Schematic of generic decision feedback equaliser

2 Decision feedback as space translation

The structure of a generic DFE is depicted in Fig. 1. The equalisation process defined in Fig. 1 uses the information present in the observed channel output vector

$$\mathbf{r}(k) = [r(k) \ldots r(k-m+1)]^T \qquad (5)$$

and the past detected symbol vector

$$\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1) \ldots \hat{s}(k-d-n)]^T \qquad (6)$$

to produce an estimate $\hat{s}(k-d)$ of $s(k-d)$. The integers d, m and n are known as decision delay, feedforward

order and feedback order, respectively. Without the loss of generality, $d = n_a - 1$ is chosen to cover the entire channel dispersion, m is related to d by $m = d + 1 = n_a$, and n is given by $n = n_a + m - d - 2 = n_a - 1$. We show that this choice of the DFE structure parameters is sufficient to guarantee the linear separability of the subsets of the channel states related to the different decisions.

Applying the channel model (eqn. 3) to each element of the observation vector (eqn. 5) gives rise to

$$\mathbf{r}(k) = F\mathbf{s}(k) + \mathbf{e}(k) \qquad (7)$$

where $\mathbf{e}(k) = [e(k) \ldots e(k-m+1)]^T$, $\mathbf{s}(k) = [\mathbf{s}_f^T(k) \ \mathbf{s}_b^T(k)]^T$ with

$$\left. \begin{array}{l} \mathbf{s}_f(k) = [s(k) \ldots s(k-d)]^T \\ \mathbf{s}_b(k) = [s(k-d-1) \ldots s(k-d-n)]^T \end{array} \right\} \qquad (8)$$

and the $m \times (d+1+n)$ matrix F has the form

$$F = [F_1 \ \ F_2] \qquad (9)$$

with the $m \times (d+1)$ matrix F_1 and $m \times n$ matrix F_2 defined by

$$F_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix} \qquad (10)$$

and

$$F_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a-1} & 0 & \ddots & \vdots \\ a_{n_a-2} & a_{n_a-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a-2} & a_{n_a-1} \end{bmatrix} \qquad (11)$$

respectively. Under the assumption that the feedback vector is correct, that is, $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$, eqn. 7 can be rewritten as

$$\mathbf{r}(k) = F_1 \mathbf{s}_f(k) + F_2 \hat{\mathbf{s}}_b(k) + \mathbf{e}(k) \qquad (12)$$

Thus the original observation space $\mathbf{r}(k)$ is transformed into a new space $\mathbf{r}'(k)$ owing to decision feedback

$$\mathbf{r}'(k) = \mathbf{r}(k) - F_2 \hat{\mathbf{s}}_b(k) \qquad (13)$$

Furthermore, the elements of $\mathbf{r}'(k)$ can be computed recursively according to the formula

$$\left. \begin{array}{l} r'(k-i) = z^{-1} r'(k-i+1) - a_{n_a-i} \hat{s}(k-d-1), \\ \qquad\qquad\qquad\qquad i = m-1, \ldots, 2, 1 \\ r'(k) = r(k) \end{array} \right\} \qquad (14)$$

where z^{-1} should be interpreted as the unit delay operator. Based on this interpretation of decision feedback, an alternative DFE structure is depicted in Fig. 2. A DFE is reduced to a transversal equaliser in the translated space. Some researchers have realised this space translation nature of decision feedback [3, 10] but they did not go as far as to derive eqn. 14 and Fig. 2. Since the structure of Fig. 2 is equivalent to that of Fig. 1, certain properties of a DFE can be studied by considering its corresponding transversal equaliser, which is an easier task. This is the approach adopted in [11] to derive a concise version of the Bayesian DFE. Iltis [13] has developed an importance sampling simulation technique for evaluating the performance of the Bayesian

equaliser. This technique can readily be applied to evaluate the lower-bound performance (with correct feedback) of the Bayesian DFE based on space translation.
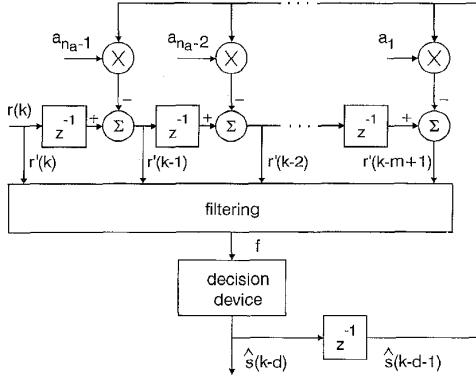

Fig.2 *Schematic of translated decision feedback equaliser.*

We have the following result for the general DFE. Let the $N_f = M^{d+1}$ sequences or states of $s_f(k)$ be $s_{f,j}$ for $1 \le j \le N_f$. The set of the noiseless channel states in the translated space is

$$R' \triangleq \left\{ \mathbf{r}'_j = F_1 \mathbf{s}_{f,j}, \ 1 \le j \le N_f \right\} \qquad (15)$$

This set can be partitioned into M subsets conditioned on $s(k - d) = s_i$, $1 \le i \le M$,

$$R^{(i)} \triangleq \left\{ \mathbf{r}'_j \in R' | s(k - d) = s_i \right\}, \quad 1 \le i \le M \qquad (16)$$

Result 1: $R^{(i)}$, $1 \le i \le M$, are linearly separable.

The proof of this result is given in the Appendix (Section 7.1). This result shows that the mapping $F_1 : \mathbf{r}' = F_1 \mathbf{s}_f$ maps linearly separable sets in the s_f-space onto linearly separable sets in the $\mathbf{r}'$-space. This is in contrast to the case of an equaliser without decision feedback where the mapping $F : \mathbf{r} = F\mathbf{s}$ maps a large space $\mathbf{s}$ onto a smaller space $\mathbf{r}$. Hence states which are linearly separable in the $\mathbf{s}$-space will not necessarily be linearly separable in the $\mathbf{r}$-space (see appendix in [14]). Even though $R^{(i)}$, $1 \le i \le M$ are linearly separable, the optimal decision boundary will generally be nonlinear (the Bayesian DFE). However, linear separability of the channel states related to the different decisions is a desired property to have because equalisation performance in this case is generally much better than that of the nonlinear separable case.

We use a simple example to illustrate the space translation property of decision feedback. Consider the channel

$$A_1(z) = 0.5 + 1.0z^{-1} \qquad (17)$$

and the equaliser structure of $d = 1$, $m = 2$ and $n = 1$. Assume that the symbol constellation is 2-PAM, that is, $s(k) \in \{\pm 1\}$. The set of the channel states in the original observation space $\mathbf{r}(k)$ is listed in Table 1 and depicted in Fig. 3. The decision feedback $s(k - 2)$ corresponds to a space translation, the effect of which is also illustrated in Fig. 3. It can be seen that decision feedback effectively merges channel states and this simplifies the decision process. The two subsets of the translated states, the darkened states $\{\mathbf{r}'_1, \mathbf{r}'_2\}$ and $\{\mathbf{r}'_3, \mathbf{r}'_4\}$ in Fig. 3, are obviously linearly separable.
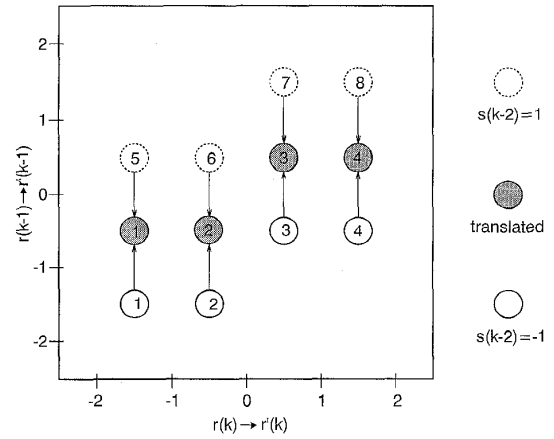

Fig.3 *Illustration of effect of decision feedback $s(k - 2)$ on channel states for channel $A_1(z) = 0.5 + 1.0z^{-1}$ with a 2-PAM constellation*

Table 1: Symbol and channel states for $A_1(z)$ ▢ 0.5 + $1.0z^{-1}$ with 2-PAM constellation

No.	$s(k)$	$s(k-1)$	$s(k-2)$	$r(k)$	$r(k-1)$
1	−1	−1	−1	−1.5	−1.5
2	1	−1	−1	−0.5	−1.5
3	−1	1	−1	0.5	−0.5
4	1	1	−1	1.5	−0.5
5	−1	−1	1	−1.5	0.5
6	1	−1	1	−0.5	0.5
7	−1	1	1	0.5	1.5
8	1	1	1	1.5	1.5

3 Linear-combiner DFE

The linear-combiner DFE is based on a linear filtering of $\mathbf{r}(k)$ and $\hat{\mathbf{s}}_b(k)$, and the decision is made by quantising the filter output

$$f\left(\mathbf{r}(k), \hat{\mathbf{s}}_b(k)\right) = \mathbf{w}^T \mathbf{r}(k) + \mathbf{b}^T \hat{\mathbf{s}}_b(k) \qquad (18)$$

where $\mathbf{w} = [w_0 \ldots w_{m-1}]^T$ and $\mathbf{b} = [b_1 \ldots b_n]^T$ are the coefficients of the feedforward and feedback filters, respectively. Since the linear-combiner DFE is a special case of the generic DFE structure depicted in Fig. 1, by performing the translation of eqn. 13, it is reduced to the equivalent linear equaliser 'without decision feedback':

$$f'\left(\mathbf{r}'(k)\right) = \mathbf{w}^T \mathbf{r}'(k) \qquad (19)$$

The decision boundary of this equivalent linear equaliser consists of $M - 1$ hyperplanes defined by: $\{\mathbf{r}' : \mathbf{w}^T \mathbf{r}' = 2i - M\}$, $1 \le i \le M - 1$. These $M - 1$ parallel hyperplanes can always be designed properly to separate the M subsets of the translated channel states $R^{(i)}$, $1 \le i \le M$. In particular, for $M = 2$, the decision boundary, $\{\mathbf{r}' : \mathbf{w}^T \mathbf{r}' = 0\}$, is a hyperplane passing through the origin of the $\mathbf{r}'(k)$-space.

The Wiener or MMSE solution is often said to provide the optimal $\mathbf{w}$ and $\mathbf{b}$. It is however optimal only with respect to the mean square error criterion. Obviously, there must exist a solution $\mathbf{w}_{opt}$ which achieves the best equalisation performance for the structure of eqn. 19. We refer to this $\mathbf{w}_{opt}$ as the MBER solution of the linear-combiner DFE. The MMSE linear-combiner DFE is generally not this MBER solution. A natural question is how different the MMSE and MBER solutions can be. We demonstrate that the performance gap between these two solutions can be large.

3.1 MMSE linear-combiner DFE

The MMSE solution for the linear-combiner DFE is well known [12]. Let $\hat{\mathbf{w}}$ and $\hat{\mathbf{b}}$ be the MMSE solution of $\mathbf{w}$ and $\mathbf{b}$. It can readily be shown that

$$\begin{bmatrix} \hat{\mathbf{w}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} T \\ -F_2^T T \end{bmatrix} \mathbf{c} \qquad (20)$$

where

$$\mathbf{c} = \sigma_s^2 \begin{bmatrix} a_{n_a-1} & a_{n_a-2} & \cdots & a_0 \end{bmatrix}^T \qquad (21)$$

and

$$T = \left(\Gamma - \sigma_s^2 F_2 F_2^T \right)^{-1} \qquad (22)$$

Here

$$\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{m-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \qquad (23)$$

with

$$\gamma_q = \left(\sum_{j=q}^{n_a-1} a_j a_{j-q} \right) \sigma_s^2 + \sigma_e^2 \delta(q), \quad 0 \le q \le m-1 \qquad (24)$$

and $\delta(q)$ is the discrete Dirac delta function. Since $\hat{\mathbf{w}}^T F_2 = -\hat{\mathbf{b}}^T$,

$$\hat{\mathbf{w}}^T \mathbf{r}(k) + \hat{\mathbf{b}}^T \hat{\mathbf{s}}_b(k) = \hat{\mathbf{w}}^T \mathbf{r}'(k) \qquad (25)$$
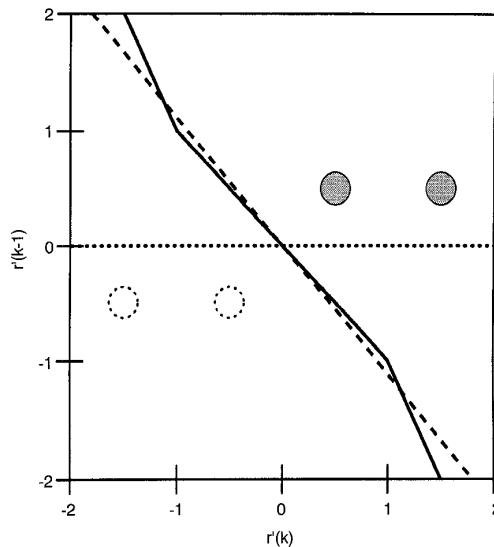
It merely confirms the space translation nature of decision feedback. Thus, when examining the MMSE linear-combiner DFE we can simply study the feedforward part of the solution. In the asymptotic case of $SNR \to \infty$, we have the following result for $\hat{\mathbf{w}}$.

Result 2: In the noise-free case

$$\hat{\mathbf{w}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \frac{1}{a_0} \end{bmatrix}^T \qquad (26)$$

This result can be derived by setting $\sigma_e^2 \to 0$ in eqn. 20. However, an alternative proof is given in the Appendix (Section 7.1). In the limit case of $SNR \to \infty$, the hyperplanes of the MMSE solution are always orthogonal to the last axis of the $\mathbf{r}'(k)$-space regardless of the channel. This cannot be the optimal solution for eqn. 19. Consider the example given in Table 1. The decision boundary of the Wiener solution for $SNR \to \infty$ is depicted in Fig. 4. The best possible linear decision boundary can easily be constructed for this example, and the MMSE solution in this case is far away from the best linear solution.

When the noise is added the hyperplanes of the MMSE linear decision boundary will rotate and are no longer orthogonal to the axis $r'(k-d)$. For the range of meaningful SNRs, however, the gap between the MMSE decision boundary and the best linear boundary can be large. Consider the example of Table 1 again. When $SNR \to 0$, the Wiener decision boundary will rotate towards the line with a slope -2 ($\hat{w}_0/\hat{w}_1 = 2$). For $SNR = 15\,dB$, the Wiener decision boundary is the line with a slope of -0.28 but the best linear decision boundary obtained by minimising the BER has a slope of -1.03. In general the MMSE solution is different from the MBER solution, and searching for the latter is worthwhile since the improvement in the BER performance over the MMSE solution can be substantial, at least for certain channels.

Fig. 4 *Asymptotic decision boundaries corresponding to large SNR for channel $A_1(z) = 0.5 + 1.0z^{-1}$ with 2-PAM constellation and decision feedback.*
——— optimal Bayesian
– – – – best linear approximation
·········· Wiener solution

3.2 MBER linear-combiner DFE

For the notational simplicity we restrict the discussion to the 2-PAM constellation. Let R^+ and R^- be the two subsets of the translated channel states R' corresponding to $s(k - d) = \pm1$, respectively. Since R^+ and R^- are linearly separable, the results of Section 7.2 apply. Under the assumption of correct decisions being fed back, the BER of the linear-combiner DFE can be calculated using

$$P_E(\mathbf{w}) = \frac{2}{N_f} \sum_{i=1}^{N_f/2} Q\left(\frac{\rho_i}{\sigma_e} \right) \qquad (27)$$

where

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} \right) dx \qquad (28)$$

$$\rho_i = \frac{|(\mathbf{v} - \mathbf{r}_i')^T \mathbf{w}|}{\|\mathbf{w}\|}, \quad \mathbf{r}_i' \in R^+ \qquad (29)$$

and $\mathbf{v}$ is any point in the decision hyperplane. Since this hyperplane passes through the origin of the $\mathbf{r}'(k)$ space one can always choose $\mathbf{v} = \mathbf{0}$. For the general M-PAM case, similar results can be derived but computation will increase dramatically as M increases.

It is obvious that the MMSE solution does not minimise $P_E(\mathbf{w})$. The optimal linear-combiner DFE should minimise the BER of eqn. 27. The following algorithm can be employed to obtain the optimal weight vector $\mathbf{w}_{opt}$ for the MBER linear-combiner DFE.

Algorithm 1

Step 1. Use a channel estimator to obtain a channel model and an estimate of the noise variance

Step 2. Compute the subset of channel states R^+ and use the low noise Wiener solution (eqn. 26) as the initial value $\mathbf{w}(0)$

Step 3. Use the gradient algorithm

$$\mathbf{w}(l+1) = \mathbf{w}(l) - \frac{\eta}{\left\| \frac{\partial P_E(\mathbf{w}(l))}{\partial \mathbf{w}} \right\|} \frac{\partial P_E(\mathbf{w}(l))}{\partial \mathbf{w}} \qquad (30)$$

to optimise $\mathbf{w}$, where η is an adaptive gain.

The derivative of $P_E(\mathbf{w})$ with respect to $\mathbf{w}$ can be found in the Appendix (Section 7.2). The gradient algorithm (eqn. 30) is an offline optimisation procedure and does not involve any channel observation $\mathbf{r}'(k)$. Thus the algorithm 1 is suitable for application to stationary channels. For nonstationary channels it is desirable to update the weights after each new observation sample is taken, and the following recursive adaptive algorithm can be used to achieve this purpose.
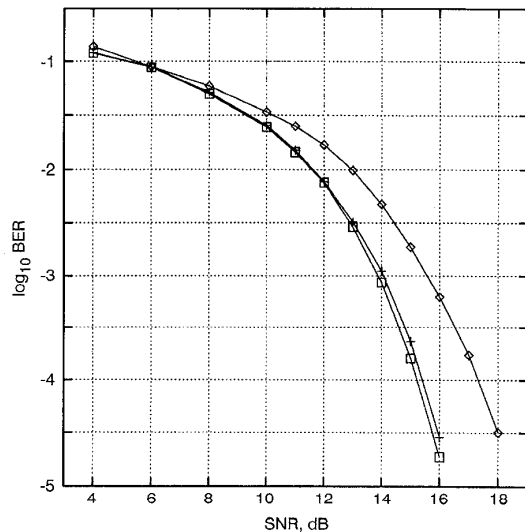
Algorithm 2. At the sample k:

Step 1. Use the least mean square (LMS) algorithm to update the channel estimate $\mathbf{a}(k) = [a_0(k) \ ... \ a_{n_a-1}(k)]^T$ and a noise variance estimator to update $\sigma_e^2(k)$

Step 2. Compute the subset of the channel states $R^+(k)$ and the gradient $\partial P_E(\mathbf{w}(k-1))/\partial \mathbf{w}$

Step 3. Update the equaliser's weights according to

$$\mathbf{w}(k) = \mathbf{w}(k-1) - \frac{\eta}{\left\| \frac{\partial P_E(\mathbf{w}(k-1))}{\partial \mathbf{w}} \right\|} \frac{\partial P_E(\mathbf{w}(k-1))}{\partial \mathbf{w}}$$

(31)

Computational complexity of the adaptive MBER linear-combiner DFE is considerably more than that of the standard adaptive MMSE linear-combiner DFE. However, the performance gain justifies the increase in computation. Some of the channel states $\mathbf{r}'_i \in R^+$ are far away from the decision boundary and contribute little to the performance criterion of eqn. 27. Computational requirements of the MBER linear-combiner DFE can be reduced significantly by neglecting these states from the optimisation procedure with little performance degradation. For example, consider the case of Fig. 4. By just using the single state at (0.5, 0.5) in the optimisation, little performance degradation will occur, compared with using the full set R^+ of the two states.

Fig.5 *Performance comparison for channel $A_1(z) = 0.5 + 1.0z^{-1}$ and 2-PAM constellation with detected symbols being fed back*
—◇— MMSE DFE
—+— MBER DFE
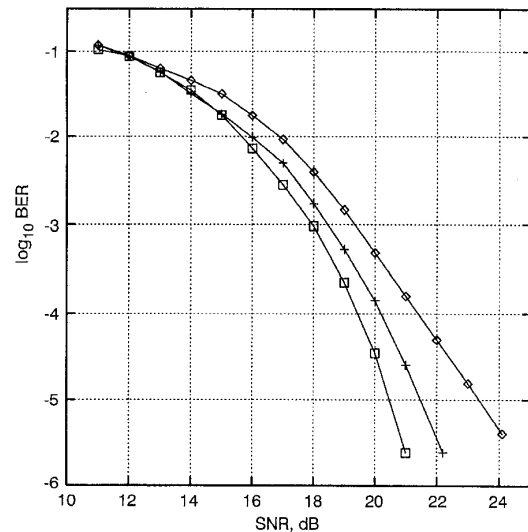—□— Bayesian DFE

3.3 Simulation study

Two examples were used to compare the MBER and MMSE solutions of the linear-combiner DFE. The optimal weight vector $\mathbf{w}_{opt}$ for the linear-combiner DFE was obtained using the gradient algorithm (eqn. 30).

The first example was the channel given in Table 1. The decision boundary of the MBER linear-combiner DFE is depicted in Fig. 4 under the title 'best linear approximation' to emphasise the fact that it is the best linear approximation to the optimal nonlinear Bayesian boundary. This example clearly demonstrates that the MMSE solution does not achieve the full performance potential of the linear-combiner DFE structure. Fig. 5 compares the BERs as a function of SNR with detected symbols being fed back for the Bayesian, MBER linear-combiner and MMSE linear-combiner DFEs. For this example, the MBER linear-combiner DFE is far superior over the MMSE solution and is very close to the optimal nonlinear Bayesian solution.

The second example was a five-tap channel with the transfer function given by

$$A_2(z) = 0.227 + 0.466z^{-1} + 0.688z^{-2}$$
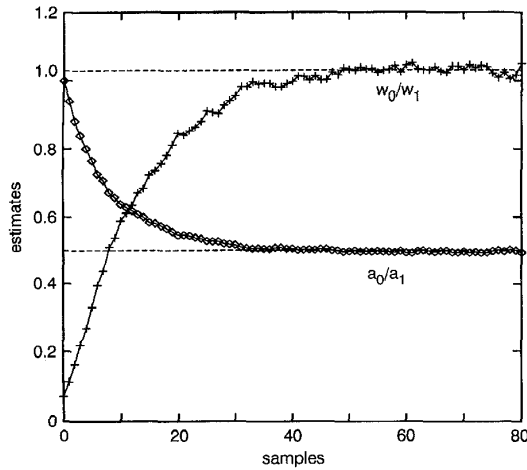$$+ 0.466z^{-3} + 0.227z^{-4}$$
(32)

The structure of the DFE was chosen to be $d = 4$, $m = 5$ and $n = 4$. The BERs of the Bayesian, MBER linear-combiner and MMSE linear-combiner DFEs with detected symbols being fed back are plotted in Fig. 6, where it can be seen that the performance of the MBER linear-combiner DFE is significantly better than that of the MMSE solution. The performance gap between the Bayesian DFE and the MBER linear-combiner DFE confirms the fact that the real optimal solution for the DFE structure of Fig. 1 is generally nonlinear. The best linear solution is suboptimal in nature. However, the usual MMSE solution is inferior to this best linear solution.

Fig.6 *Performance comparison for channel $A_2(z) = 0.227 + 0.466z^{-1} + 0.688z^{-2} + 0.466z^{-3} + 0.227z^{-4}$ and 2-PAM constellation with detected symbols being fed back*
—◇— MMSE DFE
—+— MBER DFE
—□— Bayesian DFE

The convergence behaviour of the algorithm 2 was tested using the following example. Initially, the channel had a transfer function $A_3(z) = 0.8 + 0.8z^{-1}$ with $SNR = 15$dB. At the sample $k = 0$, the channel jumped to the transfer function $A_1(z) = 0.5 + 1.0z^{-1}$. The LMS algorithm was used to estimate the channel taps with an adaptive gain 0.1 and eqn. 31 was used to update the equaliser weights with $\eta = 0.1$. The trajectories of

320

the channel estimates $a_0(k)/a_1(k)$ and the equaliser weights $w_0(k)/w_1(k)$, averaged over 50 different runs, are plotted in Fig. 7. It can be seen that the convergence speed of this adaptive procedure is reasonable.

Fig. 7 *Trajectories of channel estimates and equaliser weights for channel $A_1(z) = 0.5 + 1.0z^{-1}$ with 2-PAM constellation*
Two lines indicate respective optimal values

4 Conclusions

The geometric translation property of the decision feedback in the DFE structure has been investigated in this paper. Basically, the decision feedback performs a space translation that maps the DFE onto an equivalent transversal equaliser in the translated observation space. In particular, viewed from the translated observation space, the linear-combiner DFE is reduced to a simpler linear equaliser. We have shown that, in the translated observation space, the subsets of channel states corresponding to the different decisions are always linearly separable and, under very low noise conditions, the hyperplanes of the Wiener decision boundary are orthogonal to the last axis of the translated space. This demonstrates that the MMSE solution does not achieve the best possible performance of the linear-combiner DFE structure. Based on a BER expression, a novel MBER linear-combiner DFE has been derived for 2-PAM constellation, which achieves the full performance potential of the linear-combiner DFE structure and offers the best linear approximation to the nonlinear Bayesian solution. This MBER linear-combiner DFE can be extended to the general M-PAM case but computational requirements will increase significantly as M increases.

5 Acknowledgment

6 References

1 QURESHI, S.U.H.: 'Adaptive equalization', *Proc. IEEE*, 1985, **73**, (9), pp. 1349–1387
2 SIU, S., GIBSON, G.J., and COWAN, C.F.N.: 'Decision feedback equalisation using neural network structures and performance comparison with the standard architecture', *IEE Proc. 1, Commun. Speech Vis.*, 1990, **137**, (4), pp. 221–225
3 WILLIAMSON, D., KENNEDY, R.A., and PULFORD, G.W.: 'Block decision feedback equalization', *IEEE Trans. Commun.*, 1992, **40**, (2), pp. 255–264

4 CHEN, S., MULGREW, B., and McLAUGHLIN, S.: 'Adaptive Bayesian equaliser with decision feedback', *IEEE Trans. Signal Process.*, 1993, **41**, (9), pp. 2918–2927
5 CHEN, S., McLAUGHLIN, S., and MULGREW, B.: 'Complex-valued radial basis function network, Part II: application to digital communications channel equalisation', *EURASIP Signal Process. J.*, 1994, **36**, pp. 175–188
6 CHA, I., and KASSAM, S.A.: 'Channel equalization using adaptive complex radial basis function networks', *IEEE J. Sel. Areas Commun.*, 1995, **13**, (1), pp. 122–131
7 CHEN, S., McLAUGHLIN, S., MULGREW, B., and GRANT, P.M.: 'Adaptive Bayesian decision feedback equaliser for dispersive mobile radio channels', *IEEE Trans. Commun.*, 1995, **43**, (5), pp. 1937–1946
8 FORNEY, G.D.: 'Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference', *IEEE Trans.*, 1972, **IT-18**, (3), pp. 363–378
9 ABEND, K., and FRITCHMAN, B.D.: 'Statistical detection for communication channels with intersymbol interference', *Proc. IEEE*, 1970, **58**, (5), pp. 779–785
10 CLARK, A.P., LEE, L.H., and MARSHALL, R.S.: 'Developments of the conventional non-linear equaliser', *IEE Proc. F, Commun. Radar Signal Process.*, 1982, **129**, (2), pp. 85–94
11 CHEN, S., McLAUGHLIN, S., MULGREW, B., and GRANT, P.M.: 'Bayesian decision feedback equaliser for overcoming co-channel interference', *IEE Proc. Commun.*, 1996, **143**, (4) pp. 219–225
12 CIOFFI, J.M., DUDEVOIR, G.P., EYUBOGLU, M.V., and FORNEY, G.D.: 'MMSE decision-feedback equalizers and coding - part I: equalization results', *IEEE Trans. Commun.*, 1995, **43**, (10), pp. 2582–2594
13 ILTIS, R.A.: 'A randomized bias technique for the importance sampling simulation of Bayesian equalizers', *IEEE Trans. Commun.*, 1995, **43**, (2/3/4), pp. 1107–1115
14 GIBSON, G.J., SIU, S., and COWAN, C.F.N.: 'The application of nonlinear structures to the reconstruction of binary signals', *IEEE Trans. Signal Process.*, 1991, **39**, (8), pp. 1877–1884

7 Appendix

7.1 Proof of results 1 and 2

Consider the matrix F_1 which maps the m-dimensional space $\mathbf{s}_f$ onto the m-dimensional space $\mathbf{r}'$

$$\mathbf{r}' = F_1 \mathbf{s}_f \tag{33}$$

The hyperplane in the $\mathbf{s}_f$-space

$$\mathbf{w}_s^T \mathbf{s}_f = 2i - M \tag{34}$$

divides the set of all $\mathbf{s}_{f,j}$ into two disjoint sets, where $1 \le i \le M - 1$. The decision metric for an $\mathbf{s}_{f,j}$ in the set $\{\mathbf{s}_{f,j} : 0 < \mathbf{w}_s^T \mathbf{s}_{f,j} - (2i - M)\}$ is

$$d_j = \mathbf{w}_s^T \mathbf{s}_{f,j} - (2i - M) \tag{35}$$

Since F_1 is a square upper triangular Toeplitz matrix, its eigenvalues are its diagonal elements which are all a_0. Thus F_1 is full rank and its inverse always exists. Let $\mathbf{r}'_j = F_1 \mathbf{s}_{f,j}$. Then $\mathbf{s}_{f,j} = F_1^{-1} \mathbf{r}'_j$ and we write the decision metric in terms of $\mathbf{r}'_j$:

$$d_j = \mathbf{w}_s^T F_1^{-1} \mathbf{r}'_j - (2i - M) = \mathbf{w}^T \mathbf{r}'_j - (2i - M) \tag{36}$$

where

$$\mathbf{w} = F_1^{-T} \mathbf{w}_s \tag{37}$$

Therefore we operate on the $\mathbf{r}'$-space and produce the same decision metric. Hence if we start with linearly separable states in the $\mathbf{s}_f$-space, the states after the mapping are also linearly separable in the $\mathbf{r}'$-space.

The set $\{\mathbf{s}_{f,j} : s(k - d) > s_i\}$ is linearly separable from the set $\{\mathbf{s}_{f,j} : s(k - d) \le s_i\}$, where s_i is defined in eqn. 2, since a suitable set of m weights that define the separating hyperplane can be chosen as

$$\mathbf{w}_s = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^T \tag{38}$$

Thus the two sets $\{R^{(l)}, 1 \le l \le i\}$ and $\{R^{(l)}, i + 1 \le l \le M\}$ in the $\mathbf{r}'$-space are also linearly separable, and the weight vector of the separating hyperplane is

$$\mathbf{w} = F_1^{-T} \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^T \tag{39}$$

Since F_1 is upper triangular with all the diagonal elements being a_0, F_1^{-1} is lower triangular with all the diagonal elements being $1/a_0$. Hence

$$\mathbf{w} = [\,0 \quad 0 \quad \cdots \quad 0 \quad \tfrac{1}{a_0}\,]^T \qquad (40)$$

Notice that the weight vector (eqn. 40) is in fact the Wiener solution for the linear-combiner DFE in the case of zero noise. Let $i = 1, ..., M - 1$, we conclude that $R^{(l)}$, $1 \le l \le M$, are linearly separable.

7.2 BER of linear equaliser $f(\mathbf{r}(k)) = \mathbf{w}^T\mathbf{r}(k)$ for 2-PAM constellation

There are two subsets of the channel states R^+ and R^- related to $s(k - d) = \pm 1$, respectively. Let Z^+ and Z^- be the regions of $\mathbf{r}(k)$ related to the decisions $\hat{s}(k - d) = \pm 1$, respectively. The BER is given by

$$P_E = \sum_{\mathbf{r}_i \in R^+} \beta_i \int_{\mathbf{r} \in Z^-} p_\mathbf{r}(\mathbf{r}|\mathbf{r}_i)d\mathbf{r} + \sum_{\mathbf{r}_j \in R^-} \beta_j \int_{\mathbf{r} \in Z^+} p_\mathbf{r}(\mathbf{r}|\mathbf{r}_j)d\mathbf{r} \qquad (41)$$

where $p_r(\mathbf{r}(k)|\mathbf{r}_i)$ is the probability density function (PDF) of $\mathbf{r}(k)$ conditioned on the received channel state being $\mathbf{r}_i$ and β_i is the *a priori* probability of $\mathbf{r}_i$. Let the number of the channel states be N_s. For the symmetric and IID symbol constellation, $\beta_i = 1/N_s$ and eqn. 41 is reduced to

$$P_E = \frac{2}{N_s} \sum_{i=1}^{N_s/2} P_e(\mathbf{r}_i), \quad \mathbf{r}_i \in R^+ \qquad (42)$$

where

$$P_e(\mathbf{r}_i) = \int_{\mathbf{r} \in Z^-} p_\mathbf{r}(\mathbf{r}|\mathbf{r}_i)d\mathbf{r}, \quad \mathbf{r}_i \in R^+ \qquad (43)$$

is the conditional error probability when the received channel state is $\mathbf{r}_i \in R^+$.

When the two subsets R^+ and R^- are linearly separable, that is R^+ and R^- can be separated by the decision hyperplane $\mathbf{w}^T\mathbf{r} = 0$, the BER expression can further be simplified. An orthogonal transformation $\mathbf{x} = L\mathbf{r}$ can be constructed which rotates the bases so that one of the transformed bases, say x_1, is parallel to $\mathbf{w}$, the

normal of the decision hyperplane. Since $LL^T = I$ and the noise $e(k)$ has an IID gaussian PDF, the conditional error probability (eqn. 43) is reduced to

$$
\begin{aligned}
P_e(\mathbf{r}_i) &= \int_{\rho_i}^{\infty} p_x(x_1)dx_1 \int_{-\infty}^{\infty} p_x(x_2)dx_2 \cdots \int_{-\infty}^{\infty} p_x(x_m)dx_m \\
&= \int_{\rho_i}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{x^2}{2\sigma_e^2}\right) dx \triangleq Q\left(\frac{\rho_i}{\delta_e}\right)
\end{aligned}
$$

$$(44)$$

where

$$\rho_i = \frac{|(\mathbf{v} - \mathbf{r}_i)^T\mathbf{w}|}{\|\mathbf{w}\|} \qquad (45)$$

is the euclidean distance between $\mathbf{r}_i$ and the decision hyperplane, and $\mathbf{v}$ is any point in this hyperplane. The BER of the linear equaliser in this case can be expressed as

$$P_E(\mathbf{w}) = \frac{2}{N_s} \sum_{i=1}^{N_s/2} Q\left(\frac{\rho_i}{\sigma_e}\right), \quad \mathbf{r}_i \in R^+ \qquad (46)$$

Here we have included $\mathbf{w}$ in the expression to emphasise that for a given channel the BER depends on the equaliser weights. The derivative of $P_e(\mathbf{r}_i)$ with respect to w_j is

$$
\begin{aligned}
\frac{\partial P_e(\mathbf{r}_i)}{\partial w_j} &= \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{\rho_i^2}{2\sigma_e^2}\right) \mathrm{sgn}\big((\mathbf{v} - \mathbf{r}_i)^T\mathbf{w}\big) \\
&\quad \times \big[(\mathbf{v} - \mathbf{r}_i)^T\mathbf{w}\|\mathbf{w}\|^{-3}w_j - \|\mathbf{w}\|^{-1}(v_j - r_{ij})\big] \\
&\qquad 0 \le j \le m - 1
\end{aligned}
$$

$$(47)$$

where $\mathrm{sgn}(\cdot)$ is the signum function, v_j and r_{ij} are the jth elements of $\mathbf{v}$ and $\mathbf{r}_i$, respectively. The derivative of $P_E(\mathbf{w})$ with respect to w_j is then given by

$$\frac{\partial P_E(\mathbf{w})}{\partial w_j} = \frac{2}{N_s} \sum_{i=1}^{N_s/2} \frac{\partial P_e(\mathbf{r}_i)}{\partial w_j}, \quad \mathbf{r}_i \in R^+ \qquad (48)$$