Minimum-SER linear-combiner decision feedback equaliser

S.Chen and B.Mulgrew

Abstract: The paper considers the conventional decision feedback equaliser (DFE) that employs a linear combination of the channel observations and past decisions. An expression of the symbol error rate (SER) is derived for the linear-combiner DFE with the general M-PAM constellation by utilising a geometric translation property of decision feedback. A method is developed to optimise the coefficients of the linear-combiner DFE to achieve the minimum-SER (MSER) solution. The performance of this MSER linear-combiner DFE is superior to the usual minimum mean square error (MMSE) solution.

1 Introduction

Equalisation is a powerful technique for combating distortion and interference in communication links [1, 2] and high-density data storage systems [3, 4]. The conventional DFE, in particular, is widely used in practice as it provides a good balance between performance and complexity. The conventional DFE [1] is based on a symbol-decision structure that employs a linear combination of the channel observations and past decisions. We will refer to this DFE as the linear-combiner DFE to distinguish it from other DFE structures that use nonlinear combinations of the channel observations and past decisions [5–10]. The Wiener or MMSE solution [11] is often said to provide the optimal solution for the linear-combiner DFE. However, the MMSE solution is not the MSER solution, the SER being the ultimate performance criterion of equalisation.

It is known that decision feedback in a DFE performs a space translation [6, 12]. Previous study [13, 14] has further developed this geometric translation property and derived the explicit recursive formula for performing the space translation. In the translated observation space, a DFE is reduced to a transversal equaliser and, furthermore, the subsets of the translated channel states related to different decisions are always linearly separable. In the asymptotic case of large signal to noise ratio (SNR), the hyperplanes of the Wiener decision boundary are orthogonal to the last axis of the translated observation space [14], which clearly illustrates why the MMSE solution does not achieve the full performance potential of the linear-combiner DFE structure.

A new contribution of this paper is the derivation of an SER expression of the linear-combiner DFE for the general M-PAM constellation by using the geometric translation approach. This allows an algorithm to be developed to obtain the MSER solution by minimising this SER criterion. Simulation results show that the MSER solution can offer a substantial SER reduction over the MMSE solution. A drawback of the MSER linear-combiner DFE is that the computational complexity increases significantly for high order signalling, compared with the MMSE solution.

In a recent work [15], an approximate MSER solution of the linear equaliser was derived for the special case of equalisable channels. Equalisability corresponds to the linear separability of channel states related to the different decisions. It is well known that linear separability is not guaranteed when a linear equaliser is used [16]. In contrast, our MSER solution is exact and is not restricted to equalisable channels, as the decision feedback always makes channel states linearly separable. For the linear equaliser with equalisable channels, our solution is also valid. The approach of [15], however, does have an advantage that it can be implemented adaptively.

We will assume that the channel and the symbol constellation are real-valued. For the complex-valued channel and modulation schemes, the results of this study are still valid. Specifically, the channel is modelled as a finite impulse response filter with an additive noise source, and the received signal at sample k is

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

where $\bar{r}(k)$ denotes the noiseless channel observation; n_a is the channel length and a_i are the channel tap weights; the Gaussian white noise $e(k)$ has a zero mean and variance $\mathrm{E}[e^2(k)] = \sigma_e^2$, and the symbol sequence $\{s(k)\}$ is independently identically distributed and has an M-PAM constellation defined by the set

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

The SNR of the system is defined as

$$\mathrm{SNR} = \mathrm{E}[\bar{r}^2(k)]/\mathrm{E}[e^2(k)] = \sigma_s^2 \left(\sum_{i=0}^{n_a-1} a_i^2 \right) / \sigma_e^2 \quad (3)$$

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

IEE Proceedings online no. 19990772

DOI: 10.1049/ip-com:19990772

Paper first received 23rd July 1998 and in revised form 25th March 1999

S. Chen is with the Communications Group, Department of Electronics & Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK

B. Mulgrew is with the Department of Electronics and Electrical Engineering, University of Edinburgh, King's Buildings, Edinburgh EH9 3JL, UK

2 Space translation and linear separability

The generic DFE, depicted in Fig. 1, uses the information present in the channel observation vector

$$\boldsymbol{r}(k) = [r(k) \cdots r(k-m+1)]^T \tag{4}$$

and the past detected symbol vector

$$\hat{\boldsymbol{s}}_b(k) = [\hat{s}(k-d-1) \cdots \hat{s}(k-d-n)]^T \tag{5}$$

to produce an estimate $\hat{s}(k-d)$ of $s(k-d)$. The integers d, m and n will be referred to as the decision delay, the feedforward and feedback orders, respectively. Without loss of generality, $d = n_a - 1$, $m = n_a$ and $n = n_a - 1$ will be used, as 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 (see lemma 1 in this Section).

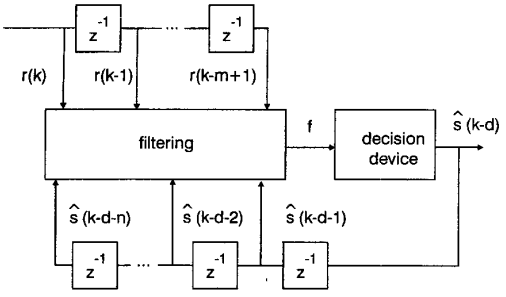

Fig. 1 *Schematic diagram of a generic decision feedback equaliser*

Applying the channel model, eqn. 1, to each element of the observation vector, eqn. 4, yields

$$\boldsymbol{r}(k) = F\boldsymbol{s}(k) + \boldsymbol{e}(k) \tag{6}$$

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

$$\begin{aligned} \boldsymbol{s}_f(k) &= [s(k) \cdots s(k-d)]^T \\ \boldsymbol{s}_b(k) &= [s(k-d-1) \cdots s(k-d-n)]^T \end{aligned} \tag{7}$$

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

$$F = [F_1 \quad F_2] \tag{8}$$

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

$$F_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix} \tag{9}$$

and

$$F_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a-1} & 0 & \ddots & \vdots \\ a_{n_a-2} & a_{n_a-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a-2} & a_{n_a-1} \end{bmatrix} \tag{10}$$

respectively. Under the assumption of correct decision feedback, that is, $\hat{\boldsymbol{s}}_b(k) = \boldsymbol{s}_b(k)$,

$$\boldsymbol{r}(k) = F_1\boldsymbol{s}_f(k) + F_2\hat{\boldsymbol{s}}_b(k) + \boldsymbol{e}(k) \tag{11}$$

Thus the decision feedback translates the original space $\boldsymbol{r}(k)$ into a new space $\boldsymbol{r}'(k)$:

$$\boldsymbol{r}'(k) \triangleq \boldsymbol{r}(k) - F_2\hat{\boldsymbol{s}}_b(k) \tag{12}$$

This property was recognised in [6, 12]. Previous research [13, 14] further pointed out that the elements of $\boldsymbol{r}'(k)$ can be computed recursively according to:

$$\begin{aligned} r'(k-i) &= z^{-1}r'(k-i+1) - a_{n_a-i}\hat{s}(k-d-1) \\ & \qquad i = m-1, \ldots, 2, 1 \\ r'(k) &= r(k) \end{aligned} \tag{13}$$

where z^{-1} is interpreted as the unit delay operator.

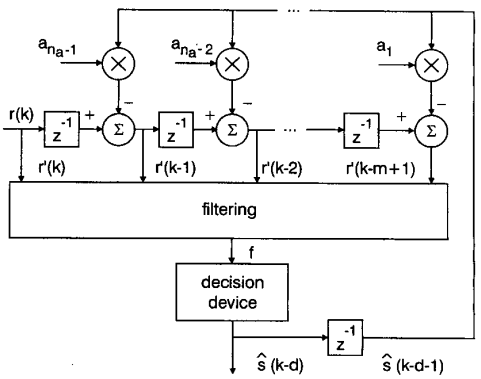

Fig. 2 *Schematic diagram of translated decision feedback equaliser*

Based on this interpretation of decision feedback, an alternative DFE structure is depicted in Fig. 2. Since a DFE is reduced to a transversal equaliser in the translated space, properties of the DFE can be studied more easily in the translated space. We have the following result of linear separability for the DFE.

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

$$R' \triangleq \{\boldsymbol{r}'_j = F_1\boldsymbol{s}_{f,j}, \ 1 \le j \le N_f\} \tag{14}$$

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

$$R^{(i)} \triangleq \{\boldsymbol{r}'_j \in R' | s(k-d) = s_i\} \qquad 1 \le i \le M \tag{15}$$

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

The proof of this lemma can be found in [14]. Lemma 1 shows that the mapping F_1: $\boldsymbol{r}' = F_1\boldsymbol{s}_f$ maps linearly separable sets in the $\boldsymbol{s}_f$-space onto linearly separable sets in the $\boldsymbol{r}'$-space. This is in contrast to the case of an equaliser without decision feedback, where the mapping F: $\boldsymbol{r} = F\boldsymbol{s}$ maps a large space $\boldsymbol{s}$ onto a smaller space $\boldsymbol{r}$. States which are linearly separable in the $\boldsymbol{s}$-space will not necessarily be linearly separable in the $\boldsymbol{r}$-space (see Appendix of [16]). Notice that we do not specify how $\boldsymbol{r}(k)$ and $\hat{\boldsymbol{s}}_b(k)$ are combined here and, therefore, the results are valid for any DFE. It should be emphasised that, even though $R^{(i)}$, $1 \le i \le M$, are linearly separable, the optimal decision boundary will generally be nonlinear (the Bayesian DFE [7]). However, linear separability of the channel states related to the different decisions is a highly desirable property to have because equalisation performance in this case is generally much better than that of the nonlinear separable case.

A simple example taken from [14] is used to illustrate the space translation property of decision feedback. Consider the two-tap channel

$$\boldsymbol{a} = [a_0 \quad a_1]^T = [0.5 \quad 1.0]^T \text{ with 2-PAM symbols} \tag{16}$$

and the DFE with $d = 1$, $m = 2$ and $n = 1$. The set of 8-channel states in the original observation space $\boldsymbol{r}(k)$ is depicted in Fig. 3. The decision feedback $s(k-2)$ corresponds to a space translation, the effect of which is illustrated in Fig. 3. It can be seen that decision feedback effectively 'merges' channel states, and this simplifies the decision process. This space translation property was adopted in [17] to derive a concise version of the Bayesian DFE. Iltis [18] has developed an importance sampling technique for evaluating the performance of the Bayesian equaliser, valid only for the case of linearly separable channel states. Lemma 1 shows that this importance sampling technique can readily be applied to evaluate the performance of the Bayesian DFE [Note 1].

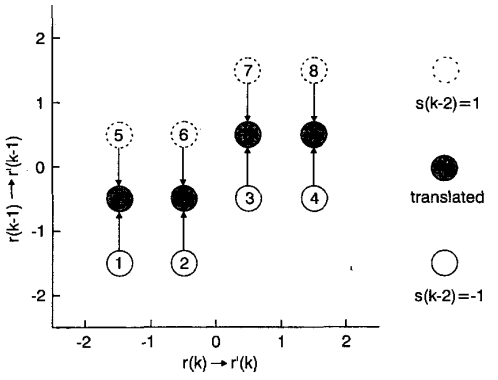

Fig. 3 *Illustration of effect of decision feedback $s(k-2)$ on channel states for channel $\boldsymbol{a} = [0.5\ 1.0]^T$ with a 2-PAM constellation*

3 Linear-combiner DFE

The linear-combiner DFE is based on a linear filtering of $\boldsymbol{r}(k)$ and $\hat{\boldsymbol{s}}_b(k)$ given by

$$f(\boldsymbol{r}(k), \hat{\boldsymbol{s}}_b(k)) = \boldsymbol{w}^T \boldsymbol{r}(k) + \boldsymbol{b}^T \hat{\boldsymbol{s}}_b(k) \tag{17}$$

where

$$\begin{aligned} \boldsymbol{w} &= [w_0 \cdots w_{m-1}]^T \\ \boldsymbol{b} &= [b_1 \cdots b_n]^T \end{aligned} \tag{18}$$

are the coefficients of the feedforward and feedback filters, respectively. Since the linear-combiner DFE is a special case of the generic DFE depicted in Fig. 1, by performing the translation eqn. 12, it is reduced to the equivalent linear equaliser:

$$f'(\boldsymbol{r}'(k)) = \boldsymbol{w}^T \boldsymbol{r}'(k) \tag{19}$$

The decision boundary of this equivalent equaliser consists of $M-1$ parallel hyperplanes defined by: $\{\boldsymbol{r}'\colon \boldsymbol{w}^T\boldsymbol{r}' = 2i - M\}$, $1 \le i \le M-1$. These hyperplanes can always be designed properly to separate the M subsets of the translated channel states $R^{(i)}$, $1 \le i \le M$. One of the hyperplanes, $\{\boldsymbol{r}'\colon \boldsymbol{w}^T\boldsymbol{r}' = 0\}$, passes through the origin of the $\boldsymbol{r}'(k)$-space. Obviously, there must exist an MSER solution $\boldsymbol{w}_{opt}$ for the structure in eqn. 19. The usual MMSE linear-combiner DFE, however, is not this MSER solution.

3.1 MMSE linear-combiner DFE

The Wiener solution for the linear-combiner DFE is well known (e.g. [11]). Let $\hat{\boldsymbol{w}}$ and $\hat{\boldsymbol{b}}$ be the MMSE solutions of $\boldsymbol{w}$ and $\boldsymbol{b}$. It can readily be shown that

$$\begin{bmatrix} \hat{\boldsymbol{w}} \\ \hat{\boldsymbol{b}} \end{bmatrix} = \begin{bmatrix} \mathcal{T} \\ -F_2^T \mathcal{T} \end{bmatrix} \boldsymbol{c} \tag{20}$$

where

$$\boldsymbol{c} = \sigma_s^2 [a_{n_a-1} \quad a_{n_a-2} \quad \cdots \quad a_0]^T \tag{21}$$

and

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

Here

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

with

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

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

$$\hat{\boldsymbol{w}}^T \boldsymbol{r}(k) + \hat{\boldsymbol{b}}^T \hat{\boldsymbol{s}}_b(k) = \hat{\boldsymbol{w}}^T \boldsymbol{r}'(k) \tag{25}$$

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 $\hat{\boldsymbol{w}}$.

Lemma 2: In the noise-free case,

$$\hat{\boldsymbol{w}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \dfrac{1}{a_0} \end{bmatrix}^T \tag{26}$$

This result can be derived by setting $\sigma_e^2 \to 0$ in eqn. 20, but an alternative proof is given in [14]. In the limit case of SNR $\to \infty$, the hyperplanes of the MMSE solution are always orthogonal to the last axis of the $\boldsymbol{r}'(k)$-space, which cannot be the optimal solution of eqn. 19 for any channel. Consider the example given in Fig. 3. 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, which is very different from the MMSE solution. The true optimal Bayesian decision boundary in the asymptotic case is also illustrated in Fig. 4.

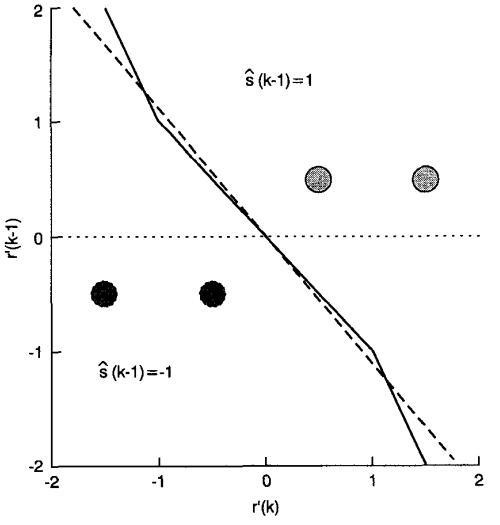

Fig. 4 *Asymptotic decision boundaries corresponding to large SNR for channel $\boldsymbol{a} = [0.5\ 1.0]^T$ with a 2-PAM constellation and decision feedback*
——— optimal Bayesian
— — — best linear approximation
- - - - Wiener solution

Note 1: CHEN, S.: 'Importance sampling simulation for evaluating the lower-bound BER of the Bayesian DFE', submitted to *IEEE Trans. Commun.*, 1998

When the noise is added, the hyperplanes of the MMSE linear decision boundary will rotate and are no longer orthogonal to the axis $r'(k-d)$. Consider the example of Fig. 4 again. When SNR $\rightarrow$ 0, the Wiener decision boundary will rotate towards the line with a slope -2 ($\hat{w}_0/\hat{w}_1 = 2$), and there is no difference between the MMSE and MSER solutions. However, for meaningful SNRs, the difference between the MMSE decision boundary and the best linear boundary can be large. For example, given SNR = 15dB, the Wiener decision boundary is the line with a slope of -0.28, but the best linear decision boundary obtained by minimising the SER has a slope of -1.03. In general the MMSE solution is different from the MSER solution, and searching for the latter is worthwhile at least for certain channels.

3.2 MSER linear-combiner DFE

For the given channel model $\boldsymbol{a} = [a_0 \ldots a_{n_a-1}]^T$ and the noise variance σ_e^2, the following lemma shows how to compute the SER of the linear-combiner DFE.

Lemma 3: Let $l = M/2 + 1$. The SER $P_E(\boldsymbol{w})$ of the linear-combiner DFE, with the weight vector $\boldsymbol{w}$ subject to the constraint

$$\boldsymbol{w}^T \boldsymbol{a}_{rev} = \sum_{i=0}^{m-1} w_i a_{n_a-1-i} = 1 \tag{27}$$

is given by

$$P_E(\boldsymbol{w}) = \sum_{\boldsymbol{r}_j^{(l)} \in R^{(l)}} \left(\frac{M}{N_f} \cdot Q\left(\frac{\rho_{j,1}}{\sigma_e}\right) + \frac{M-2}{N_f} \cdot Q\left(\frac{\rho_{j,2}}{\sigma_e}\right) \right) \tag{28}$$

where

$$\boldsymbol{a}_{rev} = [a_{n_a-1} \cdots a_1 a_0]^T \tag{29}$$

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \tag{30}$$

$$\rho_{j,1} + \rho_{j,2} = \frac{2}{\|\boldsymbol{w}\|} \qquad \rho_{j,1} = \frac{|(\boldsymbol{r}_j^{(l)} - \boldsymbol{v})^T \boldsymbol{w}|}{\|\boldsymbol{w}\|} \tag{31}$$

and $\boldsymbol{v}$ can be any point in the hyperplane $\boldsymbol{w}^T\boldsymbol{r}' = 0$. Since this hyperplane passes through the origin of the $\boldsymbol{r}'(k)$ space, we can always choose $\boldsymbol{v} = \boldsymbol{0}$.

The derivation of this SER expression is given in the Appendix (Section 7.1). $R^{(l)}$ is the subset of channel states related to $s(k-d) = s_l = 1$, and the number of states in $R^{(l)}$ is $N_f/M = M^{n_a-1}$. Obviously, the MMSE solution does not minimise $P_E(\boldsymbol{w})$. Notice that the elements of $\boldsymbol{w}$ are not linearly independent. The constraint eqn. 27 is introduced to express the SER neatly in the form of eqn. 28, and it does not change the SER. It is worth pointing out that the low noise Wiener solution, eqn. 26, satisfies the constraint eqn. 27. The following algorithm can be employed to obtain the optimal weight vector $\boldsymbol{w}_{opt}$ for the MSER linear-combiner DFE.

Algorithm:

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

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

Step 3. Solve the optimisation problem,

$$\min_{\boldsymbol{w}} P_E(\boldsymbol{w}), \quad \text{subject to } \boldsymbol{w}^T \boldsymbol{a}_{rev} = 1 \tag{32}$$

to obtain a $\boldsymbol{w}_{opt}$.

In the above algorithm, only step 1 involves channel observations. Once estimates of the channel model and noise variance are obtained, the optimisation eqn. 32 is carried out without involving any channel observation. This off-line optimisation problem can be solved, for example, using the augmented Lagrangian method [19], and an algorithm is given in the Appendix (Section 7.2). Computational complexity of this MSER linear-combiner DFE is much more than that of the standard MMSE linear-combiner DFE. However, the performance gain can justify the increase in computation. Some of the channel states $\boldsymbol{r}_j^{(l)}$ are far away from the decision hyperplanes and contribute little to the SER. Computational requirements can be reduced 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 subset $R^{(2)}$ of the two states.

4 Numerical examples

Three examples were used to compare the MSER and MMSE solutions of the linear-combiner DFE. The optimal weight vector $\boldsymbol{w}_{opt}$ for the linear-combiner DFE was obtained using the algorithm described in the preceding Section. All the SERs were evaluated with detected symbols being fed back. The first example was the two-tap channel with 2-PAM symbols defined in eqn. 16. Fig. 5 compares the SERs of the MSER linear-combiner DFE with those of the MMSE linear-combiner DFE for a range of SNR conditions. For this example, the MSER linear-combiner DFE is superior and, at the SER of 10^{-4}, it has an SNR gain of ~2dB over the Wiener solution.

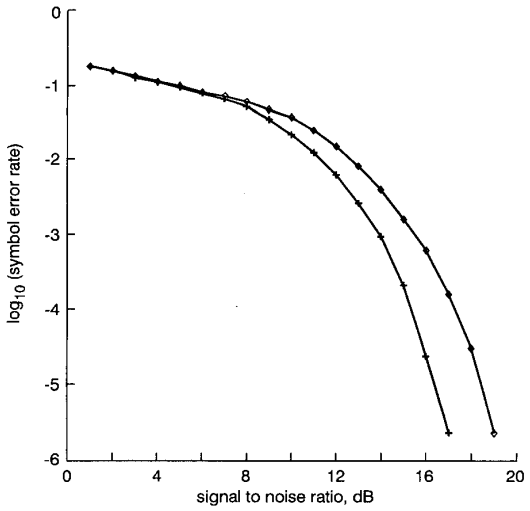

Fig. 5 *Performance comparison for channel $\boldsymbol{a} = [0.5\ 1.0]^T$ and 2-PAM symbols with detected symbols being fed back*
MMSE/MSER: MMSE/MSER linear-combiner DFEs
—◇— MMSE
—+— MSER

The second example was a 5-tap channel with the 2-PAM constellation:

$$\boldsymbol{a} = [0.227 \quad 0.466 \quad 0.688 \quad 0.466 \quad 0.227]^T \quad \text{with 2-PAM symbols} \tag{33}$$

The structure of the DFE was chosen to be $d = 4$, $m = 5$ and $n = 4$. The SERs of the MSER 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 MSER linear-combiner DFE is significantly better than that of the MMSE solution. At the SER of 10^{-4}, the MSER solution has an SNR gain of ~1dB over the MMSE solution.

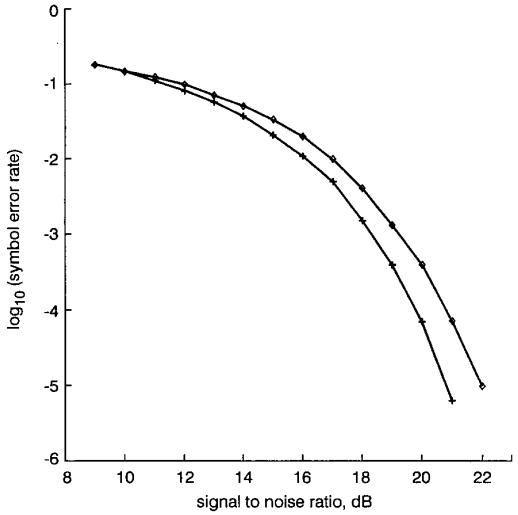

Fig. 6 *Performance comparison for channel* $\boldsymbol{a} = [0.227\ 0.466\ 0.688\ 0.466\ 0.227]^T$ *and 2-PAM symbols with detected symbols being fed back*
MMSE/MSER: MMSE/MSER linear-combiner DFEs
—◇— MMSE
—+— MSER

The third example was a 3-tap channel with the 4-PAM constellation:

$$\boldsymbol{a} = [0.3482 \quad 0.8704 \quad 0.3482]^T \quad \text{with 4-PAM symbols} \quad (34)$$

The structural parameters of the DFE were set to $d = 2$, $m = 3$ and $n = 2$. The SERs of the MSER and MMSE linear-combiner DFEs with detected symbols being fed back are depicted in Fig. 7. Again, the MSER solution is superior and has an SNR gain over 1dB at the SER of 10^{-4}, compared with the MMSE solution.

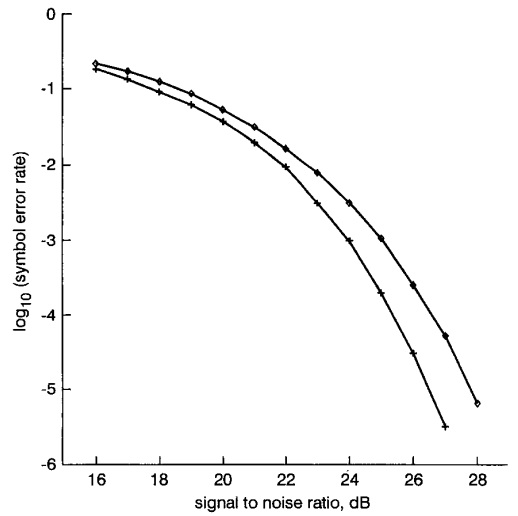

Fig. 7 *Performance comparison for channel* $\boldsymbol{a} = [0.3482\ 0.8704\ 0.3482]^T$ *and 4-PAM symbols with detected symbols being fed back*
MMSE/MSER: MMSE/MSER linear-combiner DFEs
—◇— MMSE
—+— MSER

5 Conclusions

We have derived an SER expression of the linear-combiner DFE for the general M-PAM constellation. This is made possible by utilising a geometric translation property of the decision feedback in the DFE structure. Basically, the decision feedback performs a space translation that maps the DFE onto an equivalent transversal equaliser in the translated observation space and, furthermore, the subsets of translated channel states corresponding to the different decisions are always linearly separable. In particular, viewed from the translated observation space, the linear-combiner DFE is reduced to a linear equaliser and, moreover, the hyperplanes of the Wiener solution under very low noise conditions are orthogonal to the last axis of the translated space. This shows that the MMSE solution does not achieve the full performance potential of the linear-combiner DFE structure. An algorithm is proposed to obtain the MSER solution by minimising the SER criterion. Numerical examples have been included to illustrate the better performance of the MSER linear-combiner DFE over the MMSE solution for certain channels. A drawback of this MSER solution is a significant increase in computational complexity compared with the Wiener solution.

The algorithm presented in this paper for obtaining the MSER solution is an off-line algorithm. For communication links, practical application of this algorithm is limited to the initial set-up of the DFE. This MSER linear-combiner DFE in its present form is more suited for data storage systems, as in many commercial disk drives the equalisers are trained at the factory floor and then are 'frozen' before shipping. Ongoing research will investigate how to implement this MSER linear-combiner DFE adaptively, so that it can be applied to fast time-varying channels.

6 References

1 QURESHI, S.U.H.: 'Adaptive equalization', *Proc. IEEE*, 1985, **73**, (9), pp. 1349–1387
2 PROAKIS, J.G.: 'Digital commnnications' (McGraw-Hill, New York, 1995, 3rd edn.)
3 MOON, J.: 'The role of SP in data-storage systems', *IEEE Signal Process. Mag.*, 1998, **15**, (4), pp. 54–72
4 PROAKIS, J.G.: 'Equalization techniques for high-density magnetic recording', *IEEE Signal Process. Mag.*, 1998, **15**, (4), pp. 73–82
5 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. I, Commun. Speech Vis.*, 1990, **137**, (4), pp. 221–225
6 WILLIAMSON, D., KENNEDY, R.A., and PULFORD, G.W.: 'Block decision feedback equalization', *IEEE Trans. Commun.*, 1992, **40**, (2), pp. 255–264
7 CHEN, S., MULGREW, B., and McLAUGHLIN, S.: 'Adaptive Bayesian equaliser with decision feedback', *IEEE Trans. Signal Process.*, 1993, **41**, (9), pp. 2918–2927
8 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
9 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
10 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
11 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
12 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
13 CHEN, S., CHNG, E.S., MULGREW, B., and GIBSON, G.: 'Minimum-BER linear-combiner DFE'. Proceedings of ICC'96, Dallas, TX, 1996, Vol. 2, pp. 1173–1177

14 CHEN, S., MULGREW, B., CHNG, E.S., and GIBSON, G.: 'Space translation properties and the minimum-BER linear-combiner DFE', *IEE Proc. Commun.*, 1998, **145**, (5), pp. 316–322
15 YEH, C.C., and BARRY, J.R.: 'Approximate minimum bit-error rate equalization for pulse-amplitude and quadrature-amplitude modulation'. Proceedings of ICC'98, Atlanta, USA, 1998, Vol. 1, pp. 16–20
16 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
17 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
18 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
19 BAZARAA, M.S., SHERALI, H.D., and SHETTY, C.M.: 'Nonlinear programming: Theory and algorithms (John Wiley, New York, 1993)

7 Appendixes

7.1 Derivation of SER expression

Consider the linear-combiner DFE, eqn. 19. The $M-1$ hyperplanes $\{\boldsymbol{r}': \boldsymbol{w}^T\boldsymbol{r}' = 2i - M\}$, $1 \le i \le M-1$, partition the m-dimensional $\boldsymbol{r}'$-space into M regions:

$$Z^{(i)} \triangleq \{\boldsymbol{r}' : \hat{s}(k-d) = s_i\} \qquad 1 \le i \le M \tag{35}$$

The SER of the linear-combiner DFE is a function of $\boldsymbol{w}$ and can be expressed as

$$P_E(\boldsymbol{w}) = \sum_{i=1}^{M} \sum_{\boldsymbol{r}_j^{(i)} \in R^{(i)}} \beta_j^{(i)} \int_{\boldsymbol{r}' \ni Z^{(i)}} p_{\boldsymbol{r}'}\left(\boldsymbol{r}' | \boldsymbol{r}_j^{(i)}\right) d\boldsymbol{r}' \tag{36}$$

where $p_{\boldsymbol{r}'}(\boldsymbol{r}'|\boldsymbol{r}_j^{(i)})$ is the probability density function of $\boldsymbol{r}'(k)$ conditional on the received channel state being $\boldsymbol{r}_j^{(i)}$, $\beta_j^{(i)}$ is the *a priori* probability of $\boldsymbol{r}_j^{(i)}$ and $\ni$ denotes 'not in'. Taking into account the fact of symmetry and equiprobable states, eqn. 36 is reduced to

$$P_E(\boldsymbol{w}) = \frac{2}{N_f} \sum_{i=\frac{M}{2}+1}^{M} \sum_{\boldsymbol{r}_j^{(i)} \in R^{(i)}} P_e\left(\boldsymbol{r}_j^{(i)}\right) \tag{37}$$

where

$$P_e\left(\boldsymbol{r}_j^{(i)}\right) \triangleq \int_{\boldsymbol{r}' \ni Z^{(i)}} p_{\boldsymbol{r}'}\left(\boldsymbol{r}' | \boldsymbol{r}_j^{(i)}\right) d\boldsymbol{r}' \tag{38}$$

is the conditional error probability when the received channel state is $\boldsymbol{r}_j^{(i)} \in R^{(i)}$.

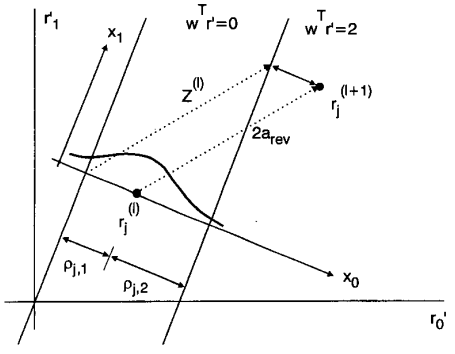

Fig. 8 *Computation of conditional error probability*

Consider the subset of channel states $R^{(l)}$, where $l = M/2 + 1$. $R^{(l)}$ is separated from other subsets by two hyperplanes $\boldsymbol{w}^T\boldsymbol{r}' = 0$ and $\boldsymbol{w}^T\boldsymbol{r}' = 2$. Referring to Fig. 8, an orthogonal transformation $\boldsymbol{x} = L\boldsymbol{r}'$ can be constructed which rotates the bases so that one of the transformed bases, say x_0, is parallel to $\boldsymbol{w}$, the normal of the decision hyperplanes. Since $LL^T = I$ and the white noise $e(k)$ has a Gaussian distribution, the conditional error probability $P_e(\boldsymbol{r}_j^{(l)})$ can be computed as

$$\begin{aligned} P_e(\boldsymbol{r}_j^{(l)}) &= \int_{\rho_{j,1}}^{\infty} p_x(x_0)dx_0 \int_{-\infty}^{\infty} p_x(x_1)dx_1 \\ &\quad \cdots \int_{-\infty}^{\infty} p_x(x_{m-1})dx_{m-1} \\ &\quad + \int_{\rho_{j,2}}^{\infty} p_x(x_0)dx_0 \int_{-\infty}^{\infty} p_x(x_1)dx_1 \\ &\quad \cdots \int_{-\infty}^{\infty} p_x(x_{m-1})dx_{m-1} \\ &= \int_{\rho_{j,1}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{x^2}{2\sigma_e^2}\right) dx \\ &\quad + \int_{\rho_{j,2}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{x^2}{2\sigma_e^2}\right) dx \\ &\triangleq Q\left(\frac{\rho_{j,1}}{\sigma_e}\right) + Q\left(\frac{\rho_{j,2}}{\sigma_e}\right) \end{aligned} \tag{39}$$

where $\rho_{j,1}$ and $\rho_{j,2}$ are the Euclidean distances between $\boldsymbol{r}_j^{(l)}$ and the hyperplanes $\boldsymbol{w}^T\boldsymbol{r}' = 0$ and $\boldsymbol{w}^T\boldsymbol{r}' = 2$, respectively. It can easily be seen that

$$\rho_{j,1} + \rho_{j,2} = \frac{2}{\|\boldsymbol{w}\|} \qquad \rho_{j,1} = \frac{|(\boldsymbol{r}_j^{(l)} - \boldsymbol{v})^T\boldsymbol{w}|}{\|\boldsymbol{w}\|} \tag{40}$$

and $\boldsymbol{v}$ can be any point in the hyperplane $\boldsymbol{w}^T\boldsymbol{r}' = 0$.

From eqns. 9, 14 and 15, it is obvious that the subset $R^{(i+1)}$ is a translation of $R^{(i)}$:

$$\begin{aligned} R^{(i+1)} &= R^{(i)} + (s_{i+1} - s_i)[a_{n_a-1} \cdots a_1 a_0]^T \\ &= R^{(i)} + 2\boldsymbol{a}_{rev} \end{aligned} \tag{41}$$

where $\boldsymbol{a}_{rev} = [a_{n_a-1} \ldots a_1 a_0]^T$. Notice that the elements of $\boldsymbol{w}$ are linearly dependent. Specifically, if we impose the following constraint:

$$\boldsymbol{w}^T\boldsymbol{a}_{rev} = \sum_{i=0}^{m-1} w_i a_{n_a-1-i} = 1 \tag{42}$$

the $(i+1)$th hyperplane is the translation of the ith hyperplane by the amount $2\boldsymbol{a}_{rev}$. As illustrated in Fig. 8, it becomes evident that

$$P_e(\boldsymbol{r}_j^{(l+i)}) = P_e(\boldsymbol{r}_j^{(l)}) \qquad 1 \le i \le \frac{M}{2} - 2 \tag{43}$$

and

$$P_e(\boldsymbol{r}_j^{(M)}) = Q\left(\frac{\rho_{j,1}}{\sigma_e}\right) \tag{44}$$

Thus the SER of the linear-combiner DFE is given by

$$\begin{aligned} &P_E(\boldsymbol{w}) \\ &= \sum_{\boldsymbol{r}_j^{(l)} \in R^{(l)}} \left(\frac{M}{N_f} \cdot Q\left(\frac{\rho_{j,1}}{\sigma_e}\right) + \frac{M-2}{N_f} \cdot Q\left(\frac{\rho_{j,2}}{\sigma_e}\right)\right) \end{aligned} \tag{45}$$

The feedforward weights of the equaliser are subject to the constraint eqn. 42.

7.2 Algorithm for solving the optimisation problem

Define the augmented Lagrangian function

$$\tilde{P}_E(\boldsymbol{w}, \lambda, \mu) = P_E(\boldsymbol{w}) + \lambda(\boldsymbol{w}^T \boldsymbol{a}_{rev} - 1) + \mu(\boldsymbol{w}^T \boldsymbol{a}_{rev} - 1)^2 \quad (46)$$

The following algorithm [19] can be used to solve the optimisation problem, eqn. 32.

Initialisation. Choose λ, $\mu > 0$ and $\boldsymbol{w}(0)$; give a termination scalar $\varepsilon > 0$; set $t = 1$.

Loop. Solve the unconstrained optimisation problem

$$\boldsymbol{w}(t) = \min_{\boldsymbol{w}} \tilde{P}_E(\boldsymbol{w}, \lambda, \mu) \quad (47)$$

If $|\boldsymbol{w}^T(t)\boldsymbol{a}_{rev} - 1| < \varepsilon$: goto *Stop*;

Else if $|\boldsymbol{w}^T(t)\boldsymbol{a}_{rev} - 1| > 0.25|\boldsymbol{w}^T(t-1)\boldsymbol{a}_{rev} - 1|$: $\mu = 10.0\mu$, goto *Loop*;

Else if $|\boldsymbol{w}^T(t)\boldsymbol{a}_{rev} - 1| \leq 0.25|\boldsymbol{w}^T(t-1)\boldsymbol{a}_{rev} - 1|$: $\lambda = \lambda + 2\mu(\boldsymbol{w}^T(t)\boldsymbol{a}_{rev} - 1)$, $t = t + 1$, goto *Loop*.

Stop. $\boldsymbol{w}(t)$ is the solution.

The unconstrained optimisation problem, eqn. 47, is solved using a simplified conjugate gradient method. For convenience, drop λ and μ in $\tilde{P}_E$, and define the gradient vector

$$\nabla \tilde{P}_E(\boldsymbol{w}) = \left[\frac{\partial \tilde{P}_E(\boldsymbol{w})}{\partial w_0} \cdots \frac{\partial \tilde{P}_E(\boldsymbol{w})}{\partial w_{m-1}}\right]^T = \nabla P_E(\boldsymbol{w}) + \lambda \boldsymbol{a}_{rev} + 2\mu(\boldsymbol{w}^T \boldsymbol{a}_{rev} - 1)\boldsymbol{a}_{rev} \quad (48)$$

Initialisation. Choose a small step size $\alpha > 0$ and a termination scalar $\beta > 0$; given $\boldsymbol{w}(1)$ and $\boldsymbol{d}(1) = -\nabla \tilde{P}_E(\boldsymbol{w}(1))$; set $j = 1$.

Loop. If $\|\nabla \tilde{P}_E(\boldsymbol{w}(j))\| < \beta$: goto *Stop*.

$\boldsymbol{w}(j+1) = \boldsymbol{w}(j) + \alpha \boldsymbol{d}(j)$,

$\phi_j = \|\nabla \tilde{P}_E(\boldsymbol{w}(j+1))\|^2 / \|\nabla \tilde{P}_E(\boldsymbol{w}(j))\|^2$

$\boldsymbol{d}(j+1) = \phi_j \boldsymbol{d}(j) - \nabla \tilde{P}_E(\boldsymbol{w}(j+1))$, $j = j + 1$, goto *Loop*.

Stop. $\boldsymbol{w}(j)$ is the solution.

The derivatives of $P_E(\boldsymbol{w})$ with respect to w_i, $0 \leq i \leq m-1$, are

$$\frac{\partial P_E(\boldsymbol{w})}{\partial w_i} = \sum_{\boldsymbol{r}_j^{(l)} \in R^{(l)}} \left(\frac{M}{N_f} \cdot \frac{\partial Q\left(\frac{\rho_{j,1}}{\sigma_e}\right)}{\partial w_i} + \frac{M-2}{N_f} \cdot \frac{\partial Q\left(\frac{\rho_{j,2}}{\sigma_e}\right)}{\partial w_i} \right) \quad (49)$$

with

$$\frac{\partial Q\left(\frac{\rho_{j,1}}{\sigma_e}\right)}{\partial w_i} = \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{\rho_{j,1}^2}{2\sigma_e^2}\right) \operatorname{sgn}\left((\boldsymbol{r}_j^{(l)} - \boldsymbol{v})^T \boldsymbol{w}\right) \times \left(\frac{w_i (\boldsymbol{r}_j^{(l)} - \boldsymbol{v})^T \boldsymbol{w}}{\|\boldsymbol{w}\|^3} - \frac{r_{j,i}^{(l)} - v_i}{\|\boldsymbol{w}\|} \right) \quad (50)$$

and

$$\frac{\partial Q\left(\frac{\rho_{j,2}}{\sigma_e}\right)}{\partial w_i} = \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{\rho_{j,2}^2}{2\sigma_e^2}\right) \times \left(\frac{2w_i}{\|\boldsymbol{w}\|^3} - \operatorname{sgn}\left((\boldsymbol{r}_j^{(l)} - \boldsymbol{v})^T \boldsymbol{w}\right) \times \left(\frac{w_i (\boldsymbol{r}_j^{(l)} - \boldsymbol{v})^T \boldsymbol{w}}{\|\boldsymbol{w}\|^3} - \frac{r_{j,i}^{(l)} - v_i}{\|\boldsymbol{w}\|} \right) \right) \quad (51)$$

where $l = M/2 + 1$, sgn(·) is the signum function, v_i and $r_{j,i}^{(l)}$ are the ith elements of $\boldsymbol{v}$ and $\boldsymbol{r}_j^{(l)}$, respectively, $\boldsymbol{v}$ is any point in the hyperplane $\boldsymbol{w}^T\boldsymbol{r}' = 0$, and $\rho_{j,1}$ and $\rho_{j,2}$ are defined in eqn. 31.