Least bit error rate adaptive nonlinear equalisers for binary signalling

S. Chen, B. Mulgrew and L. Hanzo

Abstract: The paper considers the problem of constructing adaptive minimum bit error rate (MBER) neural network equalisers for binary signalling. Motivated from a kernel density estimation of the bit error rate (BER) as a smooth function of training data, a stochastic gradient algorithm called the least bit error rate (LBER) is developed for adaptive nonlinear equalisers. This LBER algorithm is applied to adaptive training of a radial basis function (RBF) equaliser in a channel intersymbol interference (ISI) plus co-channel interference setting. A simulation study shows that the proposed algorithm has good convergence speed, and a small-size RBF equaliser trained by the LBER can closely approximate the performance of the optimal Bayesian equaliser. The results also demonstrate that the standard adaptive algorithm, the least mean square (LMS), performs poorly for neural network equalisers because the minimum mean square error (MMSE) is clearly suboptimal in the equalisation setting.

1 Introduction

Equalisation techniques play a crucial role in combating distortion and interference in modern communication systems. The topic of equalisation is well researched and a variety of solutions are available [1–7]. Recently, neural networks have been used as adaptive nonlinear equalisers [8–17]. Typically, adjusting an equaliser's parameters in a sample-by-sample fashion is required in practical applications to meet real-time computational constraints, and adaptive training of neural network equalisers is usually carried out using some stochastic gradient algorithm that tries to minimise the mean square error (MSE). This kind of stochastic gradient adaptive algorithm is classically referred to as the least mean square (LMS). It is a well-known fact that adaptive training of neural network equalisers often encounters difficulties. Since the MSE surface is highly complex in the nonlinear equalisation setting, such difficulties are typically attributed to 'local minima'. However, the problem is actually more fundamental than this. Surely it is a strange situation that, on the one hand, the performance of an equaliser is determined by its bit error rate (BER) while, on the other hand, a different MSE criterion is used at the learning stage.

For linear equalisers, there is a partial relationship between the MSE and BER. A small MSE is usually associated with a small BER. However, even in the linear case, the minimum mean square error (MMSE) solution in general is not the minimum bit error rate (MBER) solution, and it is well known that the BER gap between the linear MMSE solution and linear MBER solution can be large in certain situations [18–27]. Since the BER is the true performance indicator for equalisation, recent research has derived some adaptive linear MBER equalisers [20, 24–26]. In the linear case, the properties of the MSE and BER criteria are also better known. The MSE surface is, of course, quadratic with a unique global minimum in the equaliser weight space. The BER surface is much more complicated and the global minimum solutions form a half line, one end of which approaches infinity and the other end approaches the origin. The origin of the linear equaliser weight space is the singular point (discontinuity) of the BER surface (see, for example, [28]).

For nonlinear equalisers, both the MSE and BER surfaces are highly complex and possess local minima. Furthermore, unlike the linear case, a small MSE may not correspond to a small BER. Consider, for example, the maximum a posteriori probability or Bayesian symbol-decision equaliser [7], which is optimal in terms of BER. The MMSE criterion in this case is inappropriate. This is simply because, multiplying the Bayesian equaliser by a positive constant, it remains as the optimal Bayesian equaliser while the resulting MSE value can become very large. It is worth noting that in theory, adaptive MBER training can be achieved by adjusting the equaliser's parameters only when an error occurs. Since the BER is typically very small in communication systems, such a strategy would require an extremely long training sequence and is therefore impractical. Because practical adaptive algorithms based on the MBER criterion are unavailable to date for nonlinear equalisers, adaptive training of a neural network equaliser is typically carried out using the LMS algorithm, which may sometimes lead to poor BER performance.

This paper describes the development of a stochastic gradient adaptive MBER algorithm for nonlinear equalisers with binary signalling. We adopt an approach similar to the one used in developing an adaptive MBER linear equaliser [25, 26], namely that of using a kernel density or Parzen window estimate to approximate the BER from training data and deriving a stochastic gradient adaptive algorithm. The resulting algorithm is called the LBER because of its links to the MBER criterion, in a manner analogous to the

IEE Proceedings online no. 20030284
doi:10.1049/ip-com:20030284
Paper first received 15th January 2001 and in revised form 20th February 2002

S. Chen and L. Hanzo are with the Department of Electronics and 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

LMS algorithm for the MMSE criterion. This LBER algorithm is tested with an equalisation application in the presence of channel intersymbol interfarence (ISI), additive white Gaussian noise and co-channel interference, where a radial basis function (RBF) network is trained as an adaptive equaliser. Simulation results obtained show that the LBER algorithm achieves consistent performance and has a reasonable convergence speed. A small-size RBF network trained by the LBER algorithm can closely approximate the optimal Bayesian performance. The simulation study also confirms that a neural network equaliser trained by the LMS algorithm, although converging well in the MSE, can sometimes produce a poor BER performance.

2 Equalisation problem

It is assumed that the channel is modelled as a finite impulse response filter with an additive noise source [29]. For notational simplification, it is further assumed that there exists only one relatively strong co-channel interference. Thus, the received signal at sample k is given by

$$\begin{aligned} r(k) &= \bar{r}(k) + n(k) = \bar{r}_0(k) + \bar{r}_1(k) + n(k) \\ &= \sum_{i=0}^{n_0-1} a_{0,i} b_0(k-i) + \sum_{i=0}^{n_1-1} a_{1,i} b_1(k-i) + n(k) \end{aligned} \quad (1)$$

where $n(k)$ is a white Gaussian noise with variance $E[n^2(k)] = \sigma_n^2$; $\bar{r}(k)$, $\bar{r}_0(k)$ and $\bar{r}_1(k)$ are the noise-free received signal, desired signal and interfering signal, respectively; $a_{0,i}$, $0 \le i \le n_0-1$, are the channel taps and $a_{1,i}$, $0 \le i \le n_0-1$, are the co-channel taps; the desired and interfering data $b_0(k)$ and $b_1(k)$ are binary, taking value from the set $\{\pm 1\}$, and they are uncorrelated. Let $E[b_0^2(k)] = \sigma_0^2$ and $E[b_1^2(k)] = \sigma_1^2$. The signal to interference ratio of the system is defined by

$$\mathrm{SIR} = \frac{\sigma_0^2 \sum_{i=0}^{n_0-1} a_{0,i}^2}{\sigma_1^2 \sum_{i=0}^{n_1-1} a_{1,i}^2} \quad (2)$$

the signal to noise ratio is defined by

$$\mathrm{SNR} = \frac{\sigma_0^2 \sum_{i=0}^{n_0-1} a_{0,i}^2}{\sigma_n^2} \quad (3)$$

and the signal to interference plus noise ratio is defined by

$$\mathrm{SINR} = \frac{\sigma_0^2 \sum_{i=0}^{n_0-1} a_{0,i}^2}{\sigma_n^2 + \sigma_1^2 \sum_{i=0}^{n_1-1} a_{1,i}^2} \quad (4)$$

The equaliser considered in this study has a sample-decision finite-memory structure, and it uses the information contained in the received signal vector

$$\boldsymbol{r}(k) = [r(k) r(k-1) \cdots r(k-m+1)]^T \quad (5)$$

to produce an estimate of $b_0(k-d)$, where m is the equaliser order or memory length and d the decision delay [7]. Specifically, the equaliser is defined by

$$\begin{aligned} \hat{b}_0(k-d) &= \mathrm{sgn}(y(k)) \text{ with} \\ y(k) &= f(\boldsymbol{r}(k); \boldsymbol{w}) \end{aligned} \quad (6)$$

where $f(\cdot\,;\cdot)$ denotes the equaliser map, and $\boldsymbol{w}$ consists of all the (adjustable) parameters of the equaliser. Such an equaliser has finite 'states', since

$$\boldsymbol{b}_0(k) = [b_0(k) b_0(k-1) \cdots b_0(k-n_0-m+2)]^T \quad (7)$$

has $N_0 = 2^{m+n_0-1}$ combinations, denoted as $\boldsymbol{b}_{0,j}$, $1 \le j \le N_0$, and

$$\boldsymbol{b}_1(k) = [b_1(k) b_1(k-1) \cdots b_1(k-n_1-m+2)]^T \quad (8)$$

has $N_1 = 2^{m+n_1-1}$ combinations, denoted as $\boldsymbol{b}_{1,l}$, $1 \le l \le N_1$. From the system model (1), the received signal vector can be written as

$$\boldsymbol{r}(k) = \bar{\boldsymbol{r}}(k) + \boldsymbol{n}(k) = \bar{\boldsymbol{r}}_0(k) + \bar{\boldsymbol{r}}_1(k) + \boldsymbol{n}(k) \quad (9)$$

It is obvious that $\bar{\boldsymbol{r}}(k)$, $\bar{\boldsymbol{r}}_0(k)$ and $\bar{\boldsymbol{r}}_1(k)$ all have finite states, that is,

$$\bar{\boldsymbol{r}}_0(k) \in \mathscr{R}_0 \triangleq \{\boldsymbol{r}_{0,j},\ 1 \le j \le N_0\} \quad (10)$$

$$\bar{\boldsymbol{r}}_1(k) \in \mathscr{R}_1 \triangleq \{\boldsymbol{r}_{1,l},\ 1 \le l \le N_1\} \quad (11)$$

and

$$\begin{aligned} \bar{\boldsymbol{r}}(k) \in \mathscr{R} \triangleq \{ &\boldsymbol{r}_{0,j} + \boldsymbol{r}_{1,l}, \\ &1 \le j \le N_0 \text{ and } 1 \le l \le N_1\} \end{aligned} \quad (12)$$

Notice that the set $\mathscr{R}$ contains $N_r = N_0 \times N_1$ states, that is,

$$\bar{\boldsymbol{r}}(k) \in \mathscr{R} \triangleq \{\boldsymbol{r}_i,\ 1 \le i \le N_r\} \quad (13)$$

Denote the dth element of $\boldsymbol{b}_{0,j}$ as $b_{0,j}^{(d)}$. Since $b_{0,j}^{(d)} \in \{\pm 1\}$, the sets $\mathscr{R}_0$ and $\mathscr{R}$ can each be divided into two subsets $\mathscr{R}_0^{(\pm)}$ and $\mathscr{R}^{(\pm)}$, depending on the value of $b_{0,j}^{(d)}$. Thus $b_{0,j}^{(d)}$ is the 'class' label for $\boldsymbol{r}_{0,j} \in \mathscr{R}_0$ or $\boldsymbol{r}_{0,j} + \boldsymbol{r}_{1,l} \in \mathscr{R}$. For notational convenience, define

$$\begin{aligned} b_i^{(d)} &= b_{0,j}^{(d)}, \text{ for } 1 \le i \le N_r \\ &\text{with } j = 1 + \left\lfloor \frac{i-1}{N_1} \right\rfloor \end{aligned} \quad (14)$$

where $\lfloor \cdot \rfloor$ denotes the floor function, i.e. $\lfloor x \rfloor$ is the largest integer less than or equal to x. It can be seen that $b_i^{(d)}$ serves as the class label for $\boldsymbol{r}_i \in \mathscr{R}$.

If the channel and co-channel are known, applying the maximum *a posteriori* probability principle leads to the following optimal Bayesian equaliser [30]:

$$\begin{aligned} y_B(k) &= f_B(\boldsymbol{r}(k); \boldsymbol{w}) \\ &= \sum_{\boldsymbol{r}_{0,j} \in \mathscr{R}_0} \sum_{\boldsymbol{r}_{1,l} \in \mathscr{R}} \frac{b_{0,j}^{(d)}}{(2\pi\sigma_n^2)^{m/2} N_0 N_1} \\ &\quad \times \exp\left(-\frac{\|\boldsymbol{r}(k) - \boldsymbol{r}_{0,j} - \boldsymbol{r}_{1,l}\|^2}{2\sigma_n^2} \right) \\ &= \sum_{\boldsymbol{r}_i \in \mathscr{R}} \frac{b_i^{(d)}}{(2\pi\sigma_n^2)^{m/2} N_r} \exp\left(-\frac{\|\boldsymbol{r}(k) - \boldsymbol{r}_i\|^2}{2\sigma_n^2} \right) \end{aligned} \quad (15)$$

where equiprobable states in $\mathscr{R}_0$ and $\mathscr{R}_1$ have been assumed, and $\boldsymbol{r}_i \in \mathscr{R}$ as is defined in (13). It should be pointed out that, multiplying $y_B(k)$ by a positive constant, it remains the optimal Bayesian solution even though the resulting equaliser has a different MSE value. It is also clear that this Bayesian solution has very high computational complexity. An often-used equaliser is the linear one:

$$y_L(k) = f_L(\boldsymbol{r}(k); \boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{r}(k) \quad (16)$$

The most popular solution for the linear equaliser is the MMSE solution, which can readily be implemented adaptively using the LMS algorithm.

3 Derivation of the LBER algorithm

Consider the generic equaliser (6) where the equaliser map f is realised, for example, by a neural network. Classically, such a nonlinear equaliser is trained by adjusting $\boldsymbol{w}$ so that the MSE, $E[(b_0(k-d) - y(k))^2]$, is minimised. Typically, this is implemented adaptively using the LMS algorithm, which has a very simple form:

$$\left.\begin{aligned} &y(k) = f(\boldsymbol{r}(k);\, \boldsymbol{w}(k-1)) \\ &\boldsymbol{w}(k) = \boldsymbol{w}(k-1) + \mu(b_0(k-d) \\ &\quad -y(k))\tfrac{\partial f(\boldsymbol{r}(k);\, \boldsymbol{w}(k-1))}{\partial \boldsymbol{w}} \end{aligned}\right\} \tag{17}$$

where μ is an adaptive gain. However, minimizing the MSE does not necessarily produce a small BER. A main objective of this study is to derive an adaptive algorithm for the nonlinear equaliser (6) based on the MBER criterion.

3.1 Approximate bit error rate expression

The error probability of the equaliser (6) is

$$P_E(\boldsymbol{w}) = \text{Prob}\{\text{sgn}(b_0(k-d))y(k) < 0\} \tag{18}$$

Define the signed decision variable

$$y_s(k) = \text{sgn}(b_0(k-d))y(k) \tag{19}$$

and denote the probability density function (p.d.f.) of $y_s(k)$ as $p_y(y_s)$. Then

$$P_E(\boldsymbol{w}) = \int_{-\infty}^{0} p_y(y_s)\, dy_s \tag{20}$$

By linearising the equaliser around $\bar{\boldsymbol{r}}(k)$, the noise-free received signal vector, it can be approximated as

$$\begin{aligned} y(k) &= f(\bar{\boldsymbol{r}}(k) + \boldsymbol{n}(k); \boldsymbol{w}) \\ &\approx f(\bar{\boldsymbol{r}}(k); \boldsymbol{w}) + \left[\frac{\partial f(\bar{\boldsymbol{r}}(k); \boldsymbol{w})}{\partial \boldsymbol{r}}\right]^T \boldsymbol{n}(k) \end{aligned} \tag{21}$$

or

$$y(k) \approx f(\bar{\boldsymbol{r}}(k); \boldsymbol{w}) + e(k) = \bar{y}(k) + e(k) \tag{22}$$

where $e(k)$ is Gaussian with zero mean and variance

$$\begin{aligned} \rho^2 &= \text{E}\left[\left[\frac{\partial f(\bar{\boldsymbol{r}}(k);\, \boldsymbol{w})}{\partial \boldsymbol{r}}\right]^T \boldsymbol{n}(k)\boldsymbol{n}^T(k)\frac{\partial f(\bar{\boldsymbol{r}}(k);\, \boldsymbol{w})}{\partial \boldsymbol{r}}\right] \\ &= \frac{\sigma_n^2}{N_r}\sum_{i=1}^{N_r}\left[\frac{\partial f(\boldsymbol{r}_i;\, \boldsymbol{w})}{\partial \boldsymbol{r}}\right]^T \frac{\partial f(\boldsymbol{r}_i;\, \boldsymbol{w})}{\partial \boldsymbol{r}} \end{aligned} \tag{23}$$

with $\boldsymbol{r}_i \in \mathscr{R}$. Basically, the equaliser is approximated as an additive Gaussian noise model, with the 'clean' signal $\bar{y}(k)$ taking values from the finite set

$$\bar{y}(k) \in \{y_i = f(\boldsymbol{r}_i;\, \boldsymbol{w}),\quad 1 \le i \le N_r\} \tag{24}$$

The p.d.f. of $y_s(k)$ can thus be approximated by

$$\begin{aligned} p_y(y_s) &\approx \frac{1}{N_r\sqrt{2\pi}\rho} \\ &\times \sum_{i=1}^{N_r}\exp\left(-\frac{(y_s - \text{sgn}(b_i^{(d)})y_i)^2}{2\rho^2}\right) \end{aligned} \tag{25}$$

and the error probability of the equaliser is approximately

$$\begin{aligned} P_E(\boldsymbol{w}) &\approx \frac{1}{N_r\sqrt{2\pi}}\sum_{i=1}^{N_r}\int_{g_i(\boldsymbol{w})}^{\infty}\exp\left(-\frac{x_i^2}{2}\right)dx_i \\ &= \frac{1}{N_r}\sum_{i=1}^{N_r} Q(g_i(\boldsymbol{w})) \end{aligned} \tag{26}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty}\exp\left(-\frac{y^2}{2}\right)dy \tag{27}$$

and

$$g_i(\boldsymbol{w}) = \frac{\text{sgn}(b_i^{(d)})y_i}{\rho} = \frac{\text{sgn}(b_i^{(d)})f(\boldsymbol{r}_i;\, \boldsymbol{w})}{\rho} \tag{28}$$

In general the linearisation (21) is valid only for small $\boldsymbol{n}(k)$ (in some statistical sense). The level of approximation involved in (21), however, becomes insignificant in comparison with any stochastic (one-sample) approximation needed in deriving adaptive training.

3.2 Approximate minimum bit error rate solution

If the channel and co-channel are known, an approximate MBER solution for the equaliser (6) can be obtained by minimising the approximate BER expression (26). The gradient of $P_E(\boldsymbol{w})$ is approximately:

$$\begin{aligned} \nabla P_E(\boldsymbol{w}) &\approx -\frac{1}{N_r\sqrt{2\pi}}\sum_{i=1}^{N_r}\exp\left(-\frac{y_i^2}{2\rho^2}\right)\frac{\partial g_i(\boldsymbol{w})}{\partial \boldsymbol{w}} \\ &\approx -\frac{1}{N_r\sqrt{2\pi}}\sum_{i=1}^{N_r}\exp\left(-\frac{y_i^2}{2\rho^2}\right)\text{sgn}(b_i^{(d)})\frac{\partial f(\boldsymbol{r}_i;\, \boldsymbol{w})}{\partial \boldsymbol{w}} \end{aligned} \tag{29}$$

In the last approximation of (29), we have dropped the term containing $\partial\rho/\partial\boldsymbol{w}$. That is, ρ is assumed to be independent of $\boldsymbol{w}$. In general, ρ depends on the value of $\boldsymbol{w}$, unless the algorithm has already converged to the (near) optimal solution $\boldsymbol{w}_{\text{MBER}}$ and ρ has been fixed to its optimal value. Theoretical justification for this approximation still needs to be investigated, but it could be argued on the ground that this approximation is less significant than the approximation involved in deriving a stochastic one-sample adaptation.

The following iterative gradient algorithm can be used to arrive at an approximate MBER solution. Given an initial $\boldsymbol{w}(0)$, at the lth iteration, the algorithm computes:

$$\left.\begin{aligned} &y_i(l) = f(\boldsymbol{r}_i;\, \boldsymbol{w}(l-1)),\quad 1 \le i \le N_r \\ &\nabla P_E(\boldsymbol{w}(l)) = -\frac{1}{N_r\sqrt{2\pi}\rho}\sum_{i=1}^{N_r} \\ &\times\exp\left(-\frac{y_i^2(l)}{2\rho^2}\right)\text{sgn}(b_i^{(d)})\frac{\partial f(\boldsymbol{r}_i;\, \boldsymbol{w}(l-1))}{\partial \boldsymbol{w}} \\ &\boldsymbol{w}(l) = \boldsymbol{w}(l-1) - \mu\nabla P_E(\boldsymbol{w}(l)) \end{aligned}\right\} \tag{30}$$

Although the 'variance' ρ^2 could iteratively be calculated by

$$\begin{aligned} \rho^2(l) &= \frac{\sigma_n^2}{N_r}\sum_{i=1}^{N_r}\left[\frac{\partial f(\boldsymbol{r}_i;\, \boldsymbol{w}(l-1))}{\partial \boldsymbol{r}}\right]^T \\ &\times \frac{\partial f(\boldsymbol{r}_i;\, \boldsymbol{w}(l-1))}{\partial \boldsymbol{r}} \end{aligned} \tag{31}$$

for numerical and convergence considerations, it is preferred to fix ρ^2 to an appropriately chosen constant. Thus, ρ^2 is considered as an algorithm parameter that requires tuning.

Notice that if the equaliser (6) is linear, all the approximations done so far can be made exactly, and the variance of $e(k)$ is $\rho^2 = \sigma_n^2\boldsymbol{w}^T\boldsymbol{w}$. In this case, we arrive at the exact MBER solution for the linear equaliser (16) [25, 26].

3.3 Block-data gradient algorithm

In practice, the set $\mathscr{R}$ is unknown. The key to developing an effective adaptive algorithm is the p.d.f. $p_y(y_s)$ of the signed decision variable $y_s(k)$. Kernel density or Parzen window estimation is known to produce reliable p.d.f. estimates with short data records and in particular is extremely natural in dealing with Gaussian mixtures [31, 32]. Given a block of K training samples $\{\boldsymbol{r}(k), b_0(k-d)\}$, a kernel density estimate of the p.d.f. is

$$\hat{p}_y(y_s) = \frac{1}{K\sqrt{2\pi}\rho} \times \sum_{k=1}^{K} \exp\left(-\frac{(y_s - \text{sgn}(b_0(k-d))y(k))^2}{2\rho^2}\right) \quad (32)$$

From the estimated error probability

$$\hat{P}_E(w) = \int_{-\infty}^{0} \hat{p}_y(y_s)\, dy_s \quad (33)$$

$\nabla \hat{P}_E(w)$ can be calculated:

$$\nabla \hat{P}_E(w) = -\frac{1}{K\sqrt{2\pi}\rho} \sum_{k=1}^{K} \exp\left(-\frac{y^2(k)}{2\rho^2}\right) \times \text{sgn}(b_0(k-d)) \frac{\partial f(r(k); w)}{\partial w} \quad (34)$$

Thus a block adaptive gradient algorithm can be derived. At the lth iteration, the algorithm computes:

$$\left.\begin{aligned} y_l(k) &= f(r(k); w(l-1)), \quad 1 \le k \le K \\ \nabla \hat{P}_E(w(l)) &= -\frac{1}{K\sqrt{2\pi}\rho} \sum_{k=1}^{K} \\ &\times \exp\left(-\frac{y^2(k)}{2\rho^2}\right) \text{sgn}(b_0(k-d)) \frac{\partial f(r(k); w(l-1))}{\partial w} \\ w(l) &= w(l-1) - \mu \nabla \hat{P}_E(w(l)) \end{aligned}\right\} \quad (35)$$

3.4 Stochastic gradient algorithm

Our real aim is to develop a stochastic gradient adaptive algorithm with sample-by-sample updating, in a similar manner to the LMS (17). The LMS algorithm is derived from its related ensemble gradient algorithm by replacing the ensemble average of the gradient with a single-data-point estimate of the gradient. Adopting a similar strategy, at sample k, a single-data-point estimate of the p.d.f. is

$$\hat{p}_y(y_s, k) = \frac{1}{\sqrt{2\pi}\rho} \times \exp\left(-\frac{(y_s - \text{sgn}(b_0(k-d))y(k))^2}{2\rho^2}\right) \quad (36)$$

Using the instantaneous or stochastic gradient,

$$\nabla \hat{P}_E(k; w) = -\frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{y^2(k)}{2\rho^2}\right) \times \text{sgn}(b_0(k-d)) \frac{\partial f(r(k); w)}{\partial w} \quad (37)$$

a stochastic gradient algorithm is given by

$$\left.\begin{aligned} y(k) &= f(r(k); w(k-1)) \\ w(k) &= w(k-1) + \frac{\mu}{\sqrt{2\pi}\rho} \exp\left(-\frac{y^2(k)}{2\rho^2}\right) \\ &\times \text{sgn}(b_0(k-d)) \frac{\partial f(r(k); w(k-1))}{\partial w} \end{aligned}\right\} \quad (38)$$

assuming that an appropriate ρ has been chosen. The influence of ρ on algorithm performance will be investigated in a simulation study. Following similar reasoning to the LMS for the MMSE criterion, the algorithm (38) is called the LBER algorithm.

4 Simulation study

To test the LBER algorithm for adaptive training of neural network equalisers, the RBF network equaliser of the form

$$y_{RBF}(k) = f_{RBF}(r(k); w) = \sum_{j=1}^{n_c} \alpha_j \exp\left(-\frac{\|r(k) - c_j\|^2}{\tilde{\sigma}_j}\right) \quad (39)$$

is used. The equaliser parameter vector w thus contains all the RBF weights α_j, widths $\tilde{\sigma}_j$ and centres c_j. The dimension of w is therefore $n_c \times (m+2)$. The derivatives of the equaliser output with respect to the equaliser's parameters in this case are given by

$$\left.\begin{aligned} \frac{\partial f_{RBF}}{\partial \alpha_j} &= \exp\left(-\frac{\|r(k) - c_j\|^2}{\tilde{\sigma}_j}\right) \\ \frac{\partial f_{RBF}}{\partial \tilde{\sigma}_j} &= \alpha_j \exp\left(-\frac{\|r(k) - c_j\|^2}{\tilde{\sigma}_j}\right) \frac{\|r(k) - c_j\|^2}{\tilde{\sigma}_j^2} \\ \frac{\partial f_{RBF}}{\partial c_j} &= 2\alpha_j \exp\left(-\frac{\|r(k) - c_j\|^2}{\tilde{\sigma}_j}\right) \frac{r(k) - c_j}{\tilde{\sigma}_j} \end{aligned}\right\} \quad (40)$$

$$1 \le j \le n_c$$

For comparison purposes, both the LMS and LBER algorithms are used to train the RBF equaliser (39), with the adaptive gain μ given in the form $\mu_k = \mu_0 k^{-1/4}$, where μ_0 is an appropriately chosen constant. For the LBER algorithm, the value of ρ^2 also needs to be determined.

In all simulations, the first $n_c/2$ data points that belong to the class $+1$ and the first $n_c/2$ data points that belong to the class -1 are used as initial centres. The initial weights are set to $\pm 1/[n_c(2\pi\sigma_n^2)^{m/2}]$ accordingly. All the widths are initially set to $8\sigma_n^2$. Two kinds of BER are used in the simulation investigation, the true BER that is computed using Monte Carlo simulation with a sufficiently long test sequence and the estimated BER calculated using the approximate BER expression (26) with $g_i(w) = \text{sgn}(b_i^{(d)})y_i/\tilde{\rho}$. The value of $\tilde{\rho}$ is so chosen that the estimated BER agrees with the true one. This $\tilde{\rho}$ should not be confused with the algorithm parameter ρ used in the LBER. The estimated BER is used to illustrate the learning rate of an adaptive algorithm, since computing the true BER of the adaptive RBF equaliser at each sample k would be computationally too demanding.

Example 1. The transfer functions of the channel and co-channel were respectively $A_0(z) = 0.5 + 1.0z^{-1}$ and $A_1(z) = \lambda(1.0 + 0.5z^{-1})$. The value of λ was set to give an SIR = 12 dB. The equaliser order was chosen to be $m = 2$ and the decision delay $d = 1$. The sets $\mathscr{R}_0$ and $\mathscr{R}_1$ each had eight points, and the set $\mathscr{R}$ had 64 states. Fig. 1 shows the sets $\mathscr{R}_0$ and $\mathscr{R}$, together with the decision boundaries of the linear MMSE and optimal Bayesian equalisers given SNR = 20 dB (SINR = 11.36 dB). Given this noise and interference condition, RBF equalisers with four and six centres were trained by the LMS and LBER algorithms, respectively. At each sample k, the estimated BER was calculated for an equaliser with $w(k)$, and this resulted in the learning rates plotted in Fig. 2 for the corresponding adaptive equalisers, where the results were averaged over 100 runs. The LBER algorithm had $\rho^2 = 20\sigma_n^2$ and $\mu_0 = 0.15$ for the four-centre RBF equaliser, and $\rho^2 = 2\sigma_n^2$ and $\mu_0 = 0.1$ for the six-centre RBF equaliser; while the LMS algorithm had $\mu_0 = 0.5$ for the four-centre RBF and $\mu_0 = 0.4$ for the six-centre RBF. For LMS training, the MSE for an equaliser with $w(k)$ was also calculated using a block of 100 test samples, and this produced the learning

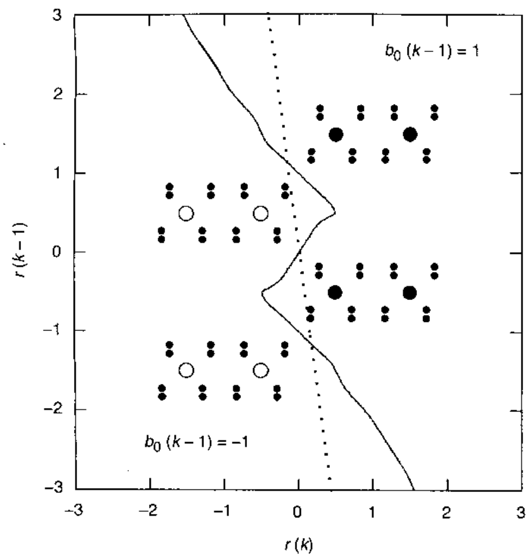

Fig. 1 *Sets of $\mathcal{R}_0^{(+)}$ (●), $\mathcal{R}_0^{(-)}$ (○) and $\mathcal{R}$ (•) together with the two decision boundaries (dotted: linear MMSE; solid: optimal) for example 1*
SIR = 12dB; SNR = 20dB (SINR = 11.36dB)

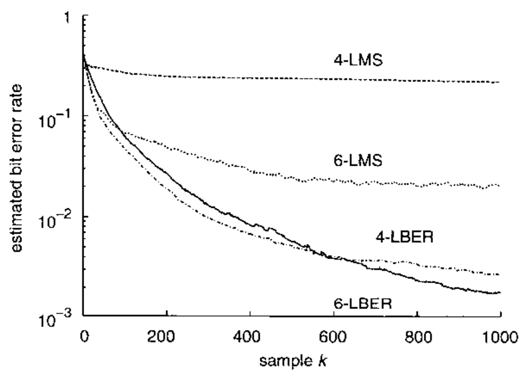

Fig. 2 *Convergence rates in terms of the estimated BER, averaged over 100 runs, for example 1*
SIR = 12dB and SNR = 20dB
4-LMS: the 4-centre RBF trained by the LMS with $\mu_0 = 0.5$;
6-LMS: the 6-centre RBF trained by the LMS with $\mu_0 = 0.4$;
4-LBER: the 4-centre RBF trained by the LBER with $\mu_0 = 0.15$ and $\rho^2 = 20\sigma_n^2$;
6-LBER: the 6-centre RBF trained by the LBER with $\mu_0 = 0.1$ and $\rho^2 = 2\sigma_n^2$

rates in terms of the MSE given in Fig. 3, where again the results were averaged over 100 runs.

It was found that the estimated BERs of an RBF equaliser with LMS training varied greatly for different runs. For the four-centre RBF with LMS training, on average, the estimated BER was hardly reduced, even though there was around 7 dB reduction in the MSE, as can be seen in Figs. 2 and 3. For the six-centre RBF with the LMS training, in some runs the BERs were close to that obtained by the LBER training and for other runs the BERs were very poor, resulting in, on average, a BER larger than that of the linear MMSE equaliser. In contrast, examining the MSE learning rate in Fig. 3 closely shows that the corresponding MSE reductions were consistent in different runs. This is not surprising, since the LMS is

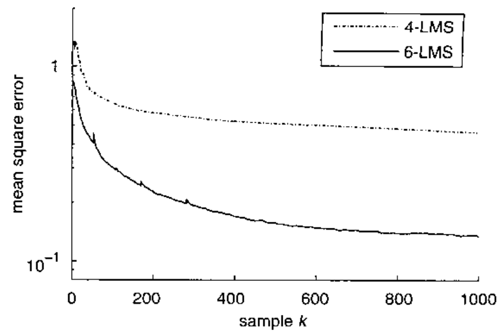

Fig. 3 *Convergence rates in terms of the MSE, averaged over 100 runs, for example 1*
SIR = 12dB and SNR = 20dB
4-LMS: the 4-centre RBF trained by the LMS with $\mu_0 = 0.5$;
6-LMS: the 6-centre RBF trained by the LMS with $\mu_0 = 0.4$

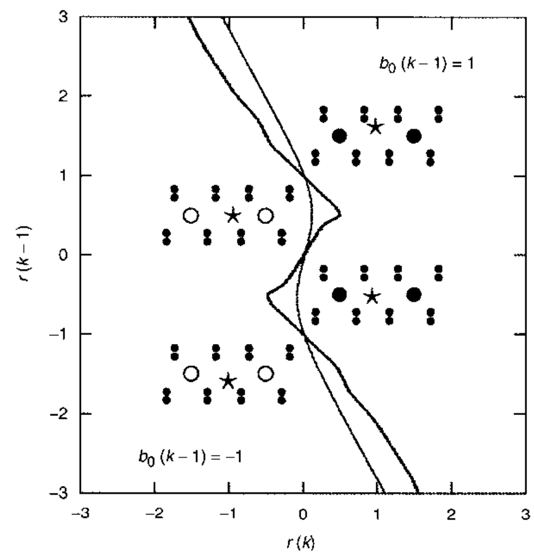

Fig. 4 *Comparison of two decision boundaries (thin solid: adaptive RBF equaliser, thick solid: optimal) for example 1*
SIR = 12dB and SNR = 20dB
The adaptive RBF equaliser has four centres and is trained by the LBER algorithm. The stars indicate the final centre positions

designed to minimise the MSE not the BER. In comparison, in terms of BER performance, LBER training was found to produce consistent results in different runs, and the six-centre RBF equaliser with LBER training converged consistently very close to the optimal performance of the 64-state Bayesian equaliser.

Typical decision boundaries of the four-centre and six-centre RBF equalisers trained by the LBER algorithm are compared with the optimal Bayesian boundary in Figs. 4 and 5, respectively. The (true) BERs of the four-centre RBF equaliser trained by the LBER together with those of the linear MMSE and optimal Bayesian equalisers are depicted in Fig. 6 as functions of SINRs. The BERs of the six-centre RBF equaliser trained by the LBER are not shown here, as they are almost indistinguishable from the optimal performance. Fig. 7 depicts the same BERs of the three equalisers as functions of SNRs. The influence of the algorithm parameter ρ^2 on the performance of the LBER algorithm was also investigated. Fig. 8 shows the BERs of the four-centre RBF equaliser trained by the LBER algorithm with a

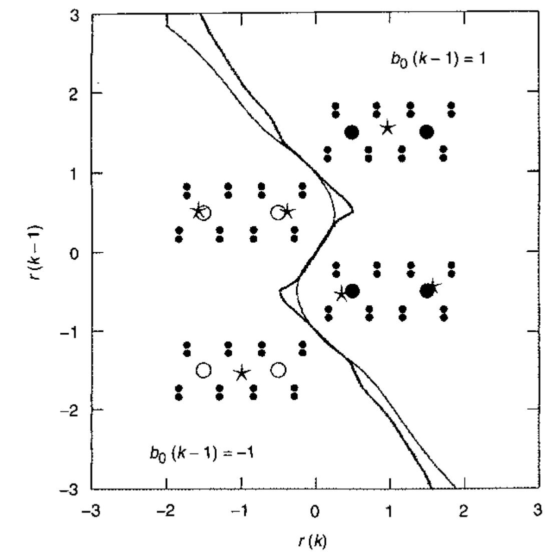

Fig. 5 *Comparison of two decision boundaries (thin solid: adaptive RBF equaliser, thick solid: optimal) for example 1*
SIR = 12 B and SNR = 20 dB
The adaptive RBF equaliser has six centres and is trained by the LBER algorithm. The stars indicate the final centre positions

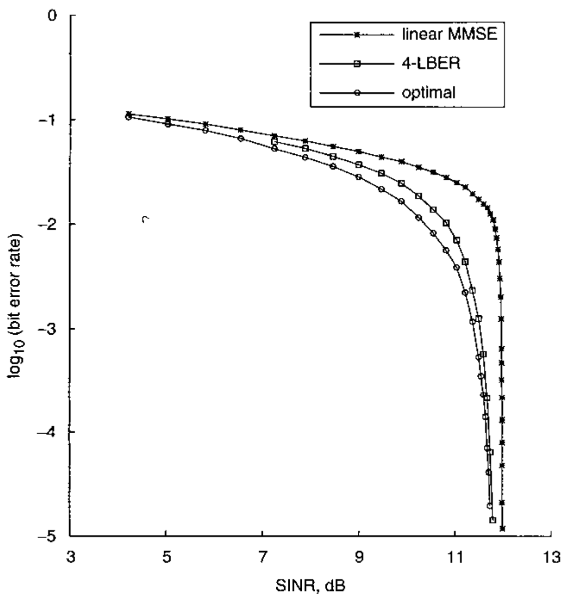

Fig. 6 *Performance comparison of three equalisers in terms of BER against SINR for example 1*
SIR = 12 dB
The adaptive RBF equaliser has four centres and is trained by the LBER algorithm

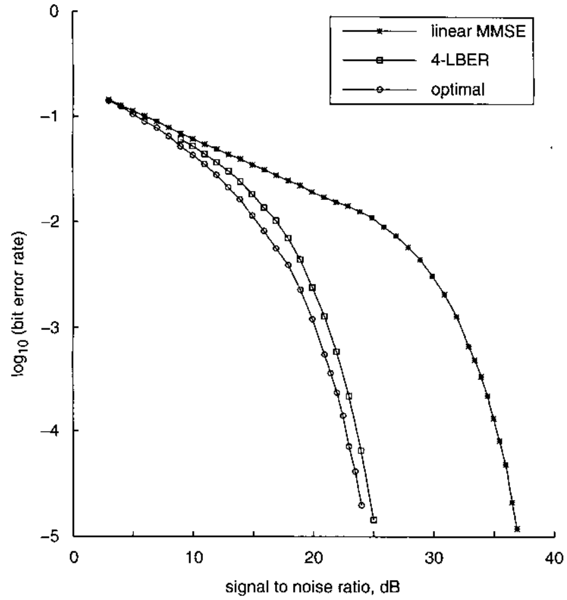

Fig. 7 *Performance comparison of three equalisers in terms of BER against SNR for example 1*
SIR = 12 dB
The adaptive RBF equaliser has four centres and is trained by the LBER algorithm

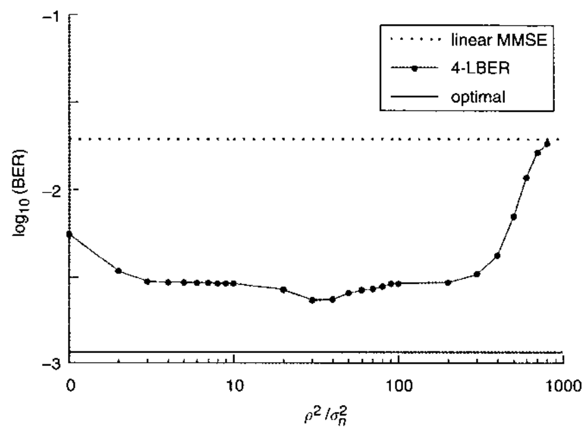

Fig. 8 *Influence of ρ^2 on the performance of the LBER algorithm for example 1 with SIR = 12 dB and SNR = 20 dB*
The adaptive RBF equaliser has four centres and the algorithm has a fixed μ_0

range of ρ^2 in a given noise and interference condition, where it can be seen that the algorithm performance is not overly sensitive to ρ^2 over a large range of values.

Example 2. The channel transfer function was $A_0(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$, and the co-channel transfer function was $A_1(z) = \lambda\,(0.6 + 0.8z^{-1})$ with the value of λ chosen to give an SIR = 20 dB. The equaliser order was set to $m = 4$ and decision delay $d = 1$. For this example, the channel state set $\mathscr{R}_0$ had 64 points and the co-channel state set $\mathscr{R}_1$ had 32 points. Thus, the number of states in $\mathscr{R}$ was 2048, and it was computationally too demanding to implement the optimal Bayesian equaliser. Given SNR = 18 dB (SINR = 15.88 dB), RBF equalisers with 16 and 32 centres were trained by the LMS and LBER algorithms, respectively. The learning rates in terms of the estimated BER are plotted in Fig. 9 for the respective adaptive equalisers, where the results were averaged over 100 runs. The LBER algorithm had $\rho^2 = 10\sigma_n^2$ and $\mu_0 = 0.3$ for the 16-centre RBF equaliser, and $\rho^2 = 8\sigma_n^2$ and $\mu_0 = 0.3$ for the 32-centre RBF equaliser; while the LMS algorithm had $\mu_0 = 0.3$ for both the 16-centre and 32-centre RBF equalisers. For the LMS training, the MSE convergence

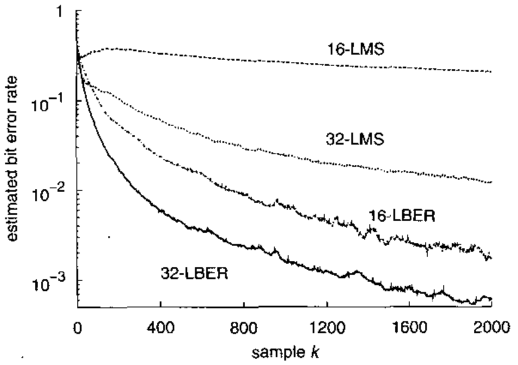

Fig. 9 *Convergence rates in terms of the estimated BER, averaged over 100 runs, for example 2*
SIR = 20 dB and SNR = 18 dB
16-LMS: the 16-centre RBF trained by the LMS with $\mu_0 = 0.3$;
32-LMS: the 32-centre RBF trained by the LMS with $\mu_0 = 0.3$;
16-LBER: the 16-centre RBF trained by the LBER with $\mu_0 = 0.3$ and $\rho^2 = 10\sigma_n^2$;
32-LBER: the 32-centre RBF trained by the LBER with $\mu_0 = 0.3$; $\rho^2 = 8\sigma_n^2$

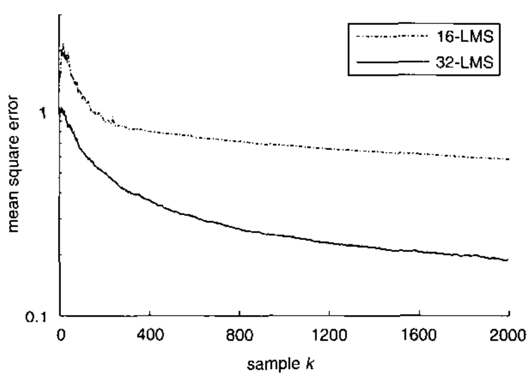

Fig. 10 *Convergence rates in terms of the MSE, averaged over 100 runs, for example 2*
SIR = 20 dB and SNR = 18 dB
16-LMS: the 16-centre RBF trained by the LMS with $\mu_0 = 0.3$;
32-LMS: the 32-centre RBF trained by the LMS with $\mu_0 = 0.3$

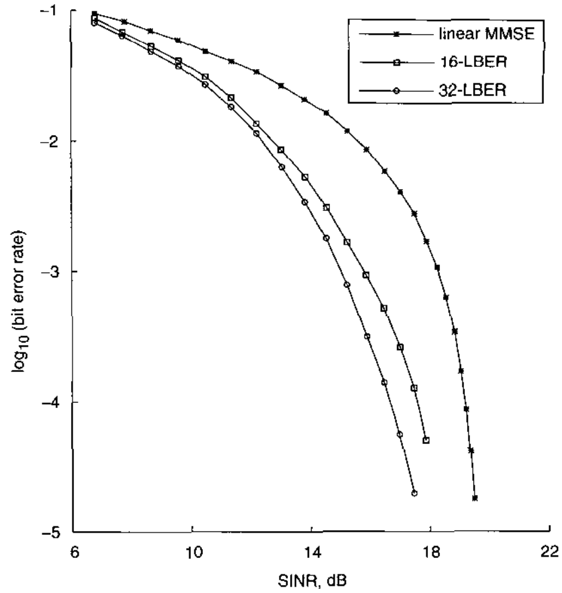

Fig. 11 *Performance comparison of three equalisers in terms of BER against SINR for example 2*
SIR = 20 dB
Adaptive RBF equalisers have 16 and 32 centres, respectively, and are trained by the LBER algorithm

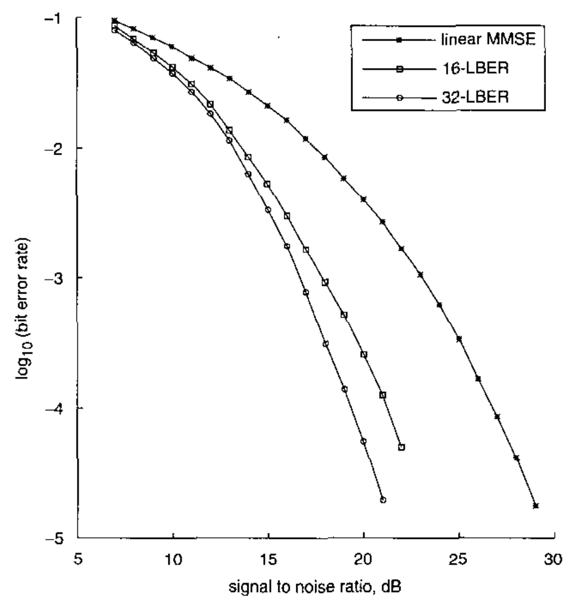

Fig. 12 *Performance comparison of three equalisers in terms of BER against SNR for example 2*
SIR = 20 dB
Adaptive RBF equalisers have 16 and 32 centres, and are trained by the LBER algorithm

performance, averaged over 100 runs, are also given in Fig. 10.

For the 16-centre RBF equaliser with the LMS training, the estimated BER on average was extremely poor, even though the corresponding MSE reduction was reasonably large. For the 32-centre RBF equaliser with the LMS training, the estimated BER on average was no better than the performance of the linear MMSE equaliser. Again, it was found that the estimated BERs varied greatly for different runs. In some runs, the 32-centre RBF equaliser with the LMS training were as good as the LBER training, but for other runs the results were very poor. The LBER algorithm was seen to produce consistent results. The BERs of the linear MMSE equaliser as functions of SINR and SNR are compared with those of the adaptive LBER 16-centre and 32-centre RBF equalisers in Figs. 11 and 12, respectively.

5 Conclusions

A novel adaptive stochastic gradient algorithm called the LBER has been developed for nonlinear equalisers. This algorithm has a similar simplicity to the LMS but is directly linked to the MBER criterion, the real goal of equalisation. This LBER algorithm has been applied to train an adaptive RBF equaliser in the presence of channel ISI, co-channel interference and additive Gaussian noise. Simulation results have shown that the LBER algorithm achieves consistent performance and has a good convergence speed. In particular, a small adaptive RBF equaliser trained by the LBER can closely approximate the optimal Bayesian performance. The results also confirm that an adaptive nonlinear equaliser trained by the LMS algorithm may produce a poor BER performance, even though it converges consistently in the MSE. Further research is warranted to investigate theoretical convergence analysis of the LBER algorithm.

6 References

1 QURESHI, S.U.H.: 'Adaptive equalisation', *Proc. IEEE*, 1985, **73**, (9), pp. 1349–1387

2 CHANG, R.W., and HANCOCK, J.C.: 'On receiver structures for channels having memory', *IEEE Trans. Inf. Theory*, 1966, **12**, (4), pp. 463-468

3 ABEND, K., HARLEY, T.J. Jr., FRICHMAN, B.D., and GUMACOS, C.: 'On optimum receivers for channels having memory', *IEEE Trans. Inf. Theory*, 1968, **14**, pp. 818–819

4 ABEND, K., and FRITCHMAN, B.D.: 'Statistical detection for communication channels with intersymbol interference', *Proc. IEEE*, 1970, **58**, (5), pp. 779–785

5 FORNEY, G.D.: 'Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference', *IEEE Trans. Inf. Theory*, 1972, **18**, (3), pp. 363–378

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 GIBSON, G.J., SIU, S., CHEN, S., COWAN, C.F.N., and GRANT, P.M.: 'The application of nonlinear architectures to adaptive channel equalisation'. Proceedings of ICC'90, Atlanta, USA, 1990, pp. 312.8.1.–312.8.5

9 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

10 CHEN, S., GIBSON, G.J., COWAN, C.F.N., and GRANT, P.M.: 'Adaptive equalization of finite non-linear channels using multilayer perceptrons', *Signal Process.*, 1990, **20**, (2), pp. 107–119

11 CHEN, S., GIBSON, G.J., and COWAN, C.F.N.: 'Adaptive channel equalization using a polynomial-perceptron structure', *IEE Proc. I, Commun. Speech Vis.*, 1990, **137**, (5), pp. 257–264

12 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

13 CHEN, S., GIBSON, G.J., COWAN, C.F.N., and GRANT, P.M.: 'Reconstruction of binary signals using an adaptive radial-basis-function equalizer', *Signal Process.*, 1991, **22**, (1), pp. 77–93

14 CHEN, S., McLAUGHLIN, S., and MULGREW, B.: 'Complex-valued radial basis function network, Part II: application to digital communications channel equalisation', *Signal Process.*, 1994, **36**, pp. 175–188

15 XIANG, Z.J., BI, G.G., and LE-NGOC, T.: 'Polynomial perceptrons and their applications to fading channel equalisation and co-channel interference suppression', *IEEE Trans. Signal Process.*, 1994, **42**, pp. 2470–2480

16 CHA, I., and KASSAM, S.A.: 'Channel equalisation using adaptive complex radial basis function networks', *IEEE J. Sel. Areas Commun.*, 1995, **13**, (1), pp. 122–131

17 CHANG, C.H., SIU, S., and WEI, C.H.: 'A polynomial-perceptron based decision feedback equalizer with a robust learning algorithm', *Signal Process.*, 1995, **47**, pp. 145–158

18 SHAMASH, E., and YAO, K.: 'On the structure and performance of a linear decision feedback equaliser based on the minimum error probability criterion'. Proceedings of ICC'74, 1974, pp. 25F1–25F5

19 CHEN, S., CHNG, E.S., MULGREW, B., and GIBSON, G.: 'Minimum-BER linear-combiner DFE'. Proceedings of ICC'96, Dallas, Texas, 1996, Vol. 2, pp. 1173–1177

20 YEH, C.C., and BARRY, J.R.: 'Approximate minimum bit-error rate equalization for binary signaling'. Proceedings of ICC'97, Montreal, Canada, 1997, Vol. **2**, pp. 1095–1099

21 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

22 CHEN, S., and MULGREW, B.: 'The minimum-SER linear-combiner decision feedback equalizer', *IEE Proc., Commun.*, 1999, **146**, (6), pp. 347–353

23 CHEN, S., GUNN, S., and HARRIS, C.J.: 'Decision feedback equaliser design using support vector machines', *IEE Proc., Vis. Image Signal Process.*, 2000, **147**, (3), pp. 213–219

24 YEH, C.C., and BARRY, J.R.: 'Adaptive minimum bit-error rate equalization for binary signaling', *IEEE Trans. Commun.*, 2000, **48**, (7), pp. 1226–1235

25 MULGREW, B., and CHEN, S.: 'Stochastic gradient minimum-BER decision feedback equalisers'. Proceedings of IEEE Symp. on adaptive systems for signal processing, communication and control, Lake Louise, Alberta, Canada, 1–4 Oct. 2000, pp. 93–98

26 MULGREW, B., and CHEN, S.: 'Adaptive minimum-BER decision feedback equalisers for binary signalling', *Signal Process.*, 2001, **81**, (7), pp. 1479–1489

27 ALTEKAR, S.A., VITYAEV, A.E., and WOLF, J.K.: 'Decision-feedback equalization via separating hyperplanes', *IEEE Trans. Commun.*, 2001, **49**, (3), pp. 480–486

28 CHEN, S., SAMINGAN, A.K., MULGREW, B., and HANZO, L.: 'Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels', *IEEE Trans. Signal Process.*, 2001, **49**, (6), pp. 1240–1247

29 PROAKIS, J.G.: 'Digital communications' (McGraw-Hill, New York, 1995, 3rd edn)

30 CHEN, S., McLAUGHLIN, S., MULGREW, B., and GRANT, P.M.: 'Bayesian decision feedback equaliser for overcoming co-channel interference', *IEE Proc., Commun.*, 1996, **143**, pp. 219–225

31 SILVERMAN, B.W.: 'Density estimation' (Chapman Hall, London, 1996)

32 BOWMAN, A.W., and AZZALINI, A.: 'Applied smoothing techniques for data analysis' (Oxford University Press, 1997)