

# NONLINEAR AND NON-GAUSSIAN SIGNAL PROCESSING

## Adaptive minimum bit-error-rate filtering

S. Chen

**Abstract:** Adaptive filtering has traditionally been developed based on the minimum mean square error (MMSE) principle and has found ever-increasing applications in communications. The paper develops adaptive filtering based on an alternative minimum bit error rate (MBER) criterion for communication applications. It is shown that the MBER filtering exploits the non-Gaussian distribution of filter output effectively and, consequently, can provide significant performance gain in terms of smaller bit error rate (BER) over the MMSE approach. Adopting the classical Parzen window or kernel density estimation for a probability density function (pdf), a block-data gradient adaptive MBER algorithm is derived. A stochastic gradient adaptive MBER algorithm is further developed for sample-by-sample adaptive implementation of the MBER filtering. Extension of the MBER approach to adaptive nonlinear filtering is also discussed.

### 1 Introduction

Adaptive filtering has been an enabling technology for communications. Traditionally, adaptive filtering has been developed based on the Wiener or MMSE approach [1, 2]. For a communication system, however, it is the BER, not the mean square error (MSE), that really matters. In communication applications, the pdf of an adaptive filter output is generally a mixed sum of Gaussian distributions. This non-Gaussian nature can be exploited explicitly, leading to alternative approaches to the MMSE filtering. For single-user channel equalisation applications, an adaptive MBER linear equaliser and a decision feedback equaliser have been developed [3–10]. Similar approaches have been adopted for linear multi-user detection in CDMA systems [11–16]. Recently, the MBER beamforming using an antenna array for wireless communication has been considered [17–19]. These previous studies have demonstrated that the MBER approach offers potentially significant performance improvement and it provides a viable alternative to the traditional adaptive filtering based on the MMSE principle.

The main contribution of this paper is to present a unified framework for adaptive MBER filtering. A linear filtering model is given in the general communication setting, and the theoretical MBER filtering solution is derived. To effectively implement the MBER solution, the non-Gaussian probability distribution of the filter output needs to be approximated accurately, and this is achieved by adopting the classical Parzen window pdf estimate [20–22], which naturally gives rise to a block-data gradient adaptive MBER algorithm. Sample-by-sample adaptive implementation of the MBER filtering is then considered, and a stochastic gradient adaptive MBER algorithm is derived which has a similar computational complexity to the very simple least mean square (LMS) algorithm. Two applications involving channel equalisation and beamforming with an antenna array are

used to demonstrate the generality and effectiveness of adaptive MBER filtering. Simulation results obtained confirm the superior performance of the MBER filtering over the MMSE one. How to extend this adaptive MBER approach to nonlinear filtering is also discussed.

### 2 System model

Consider the general linear filter of the form

$$y(k) = \sum_{l=0}^{L-1} w_l^* x_l(k) = \mathbf{w}^H \mathbf{x}(k) \quad (1)$$

where  $L$  is the filter length,  $\mathbf{x}(k) = [x_0(k) \ x_1(k) \ \cdots \ x_{L-1}(k)]^T$  is the complex-valued filter input vector and  $\mathbf{w} = [w_0 \ w_1 \ \cdots \ w_{L-1}]^T$ , the complex-valued filter weight vector. Such a filter can be found in receivers of various communication systems. In channel equalisation, for example,  $\mathbf{x}(k)$  is generated from a tap-delay-line of the received signal. For multi-user detection in CDMA systems,  $\mathbf{x}(k)$  consists of chip rate received samples. In adaptive beamforming,  $\mathbf{x}(k)$  consists of received signals at the elements of the antenna array. Generally,  $\mathbf{x}(k)$  can be expressed as

$$\mathbf{x}(k) = \mathbf{P}\mathbf{b}(k) + \mathbf{n}(k) = \bar{\mathbf{x}}(k) + \mathbf{n}(k) \quad (2)$$

where the complex-valued Gaussian noise vector  $\mathbf{n}(k) = [n_0(k) \ n_1(k) \ \cdots \ n_{L-1}(k)]^T$  has zero mean and covariance matrix  $E[\mathbf{n}(k)\mathbf{n}^H(k)] = 2\sigma_n^2 \mathbf{I}_L$ , with  $\mathbf{I}_L$  denoting the  $L \times L$  identity matrix, the complex-valued system matrix  $\mathbf{P}$  has dimension  $L \times M$ , and the information symbol vector  $\mathbf{b}(k) = [b_0(k) \ b_1(k) \ \cdots \ b_{M-1}(k)]^T$ . For single-user applications,  $\mathbf{b}(k)$  contains current and previous  $M-1$  transmitted symbols and, for multi-user applications,  $\mathbf{b}(k)$  consists of transmitted different user symbols. Typically,  $b_i(k)$  and  $b_q(k)$  are uncorrelated if  $i \neq q$ . In this study, the modulation scheme is assumed to be binary phase and shift keying, that is,  $\mathbf{b}(k)$  is real valued with  $b_i(k) \in \{\pm 1\}$  for  $0 \leq i \leq M-1$ . The reason for considering the case of binary symbols is to simplify notations and to concentrate on the basic concepts. The approach can be extended to multi-level and complex-valued modulation schemes (see [23–25]).

The purpose of the filter (1) is to enable an estimate of the ‘desired’ symbol  $b_d(k)$ , the  $d$ th element of  $\mathbf{b}(k)$ , and this estimate is given by

$$\hat{b}_d(k) = \text{sgn}(y_R(k)) \quad (3)$$

where  $\text{sgn}(\cdot)$  denotes the sign function, and  $y_R(k) = \Re[y(k)]$  is the real part of  $y(k)$ . Note that

$$\bar{\mathbf{x}}(k) \in \mathcal{X} \triangleq \{\bar{\mathbf{x}}_q = \mathbf{P}\mathbf{b}_q, 1 \leq q \leq N_b\} \quad (4)$$

where  $N_b = 2^M$  and  $\mathbf{b}_q$ ,  $1 \leq q \leq N_b$ , are all the possible sequences of  $\mathbf{b}(k)$ . The vector set  $\mathcal{X}$  can be divided into two subsets depending on the value of  $b_d(k)$ :

$$\mathcal{X}^{(\pm)} \triangleq \{\bar{\mathbf{x}}_q^{(\pm)} \in \mathcal{X} : b_d(k) = \pm 1\} \quad (5)$$

The filter output  $y(k)$  can be expressed as

$$y(k) = \mathbf{w}^H(\bar{\mathbf{x}}(k) + \mathbf{n}(k)) = \bar{y}(k) + e(k) \quad (6)$$

where  $e(k)$  is Gaussian with zero mean and variance  $E[|e(k)|^2] = 2\sigma_n^2 \mathbf{w}^H \mathbf{w}$ , and

$$\bar{y}(k) \in \mathcal{Y} \triangleq \{\bar{y}_q = \mathbf{w}^H \bar{\mathbf{x}}_q, 1 \leq q \leq N_b\} \quad (7)$$

Thus,  $\bar{y}_R(k) = \Re[\bar{y}(k)]$  can only take value from the scalar set

$$\mathcal{Y}_R \triangleq \{\bar{y}_{R,q} = \Re[\bar{y}_q], 1 \leq q \leq N_b\} \quad (8)$$

which can be partitioned into the two subsets conditioned on the value of  $b_d(k)$ ,

$$\mathcal{Y}_R^{(\pm)} \triangleq \{\bar{y}_{R,q}^{(\pm)} \in \mathcal{Y}_R : b_d(k) = \pm 1\} \quad (9)$$

The classical Wiener solution for the linear filter (1),

$$\mathbf{w}_{\text{MMSE}} = (\mathbf{P}\mathbf{P}^H + 2\sigma_n^2 \mathbf{I}_L)^{-1} \mathbf{p}_d \quad (10)$$

where  $\mathbf{p}_d$  denotes the  $d$ th column of  $\mathbf{P}$ , is generally not the optimal MBER solution. For  $\mathbf{w}_{\text{MMSE}}$  to be an MBER solution, the conditional pdf of  $y_R(k)$  given  $b_d(k) = +1$  (or  $b_d(k) = -1$ ) should be Gaussian. However, this is obviously not the case. The so-called zero-forcing (ZF) solution  $\mathbf{w}_{\text{ZF}}$ , on the other hand, does achieve a Gaussian conditional pdf, since the combined impulse response of the ZF filter and the system matrix  $\mathbf{c}_{\text{ZF}} = \mathbf{w}_{\text{ZF}}^H \mathbf{P}$  has all zero elements except the  $d$ th element  $c_d$ . That is, the ZF filter output is

$$y(k) = c_d b_d(k) + e(k) \quad (11)$$

However, the ZF filter suffers from a problem of serious noise enhancements and, consequently, its performance is much inferior compared with the MMSE filtering, in terms of both the MSE and BER. Since the BER is the true performance indicator of the system, it is desired to consider the optimal MBER filter solution.

### 3 Minimum bit error rate filtering solution

To derive the BER expression for the linear filter with a weight vector  $\mathbf{w}$ , first note that the pdf of  $y_R(k)$  is a mixed sum of Gaussian distributions:

$$p(y_R) = \frac{1}{N_b \sqrt{2\pi\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \sum_{q=1}^{N_b} \exp\left(-\frac{(y_R - \bar{y}_{R,q})^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}\right) \quad (12)$$

By exploiting the symmetric distributions of  $\mathcal{Y}_R$ , it can be shown that the BER is given by

$$P_E(\mathbf{w}) = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q(g_{q,+}(\mathbf{w})) \quad (13)$$

where  $N_{sb} = N_b/2$  is the number of the points in  $\mathcal{Y}_R^{(+)}$ ,

$$g_{q,+}(\mathbf{w}) = \frac{\text{sgn}(b_{q,d}) \bar{y}_{R,q}^{(+)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} = \frac{\text{sgn}(b_{q,d}) \Re[\mathbf{w}^H \bar{\mathbf{x}}_q^{(+)}]}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (14)$$

$\bar{y}_{R,q}^{(+)} \in \mathcal{Y}_R^{(+)}$  and  $b_{q,d}$  is the  $d$ th element of  $\mathbf{b}_q$  corresponding to the desired symbol  $b_d(k)$ . Note that the BER is invariant to a positive scaling of  $\mathbf{w}$ . Alternatively, the BER can be calculated using the other subset  $\mathcal{Y}_R^{(-)}$ . A proof of the BER formula (13) is given in the Appendix, where it can also be seen that the BER can be expressed as

$$P_E(\mathbf{w}) = \frac{1}{N_b} \sum_{q=1}^{N_b} Q(g_q(\mathbf{w})) \quad (15)$$

with

$$g_q(\mathbf{w}) = \frac{\text{sgn}(b_{q,d}) \bar{y}_{R,q}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} = \frac{\text{sgn}(b_{q,d}) \Re[\mathbf{w}^H \bar{\mathbf{x}}_q]}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (16)$$

and the calculation being over all the  $\bar{y}_{R,q} \in \mathcal{Y}_R$ .

The MBER filtering solution is then defined as

$$\mathbf{w}_{\text{MBER}} = \arg \min_{\mathbf{w}} P_E(\mathbf{w}) \quad (17)$$

Unlike the unique MMSE solution, there exists no close-form solution for  $\mathbf{w}_{\text{MBER}}$ , and there are an infinite number of MBER solutions. In fact, if  $\mathbf{w}_{\text{MBER}}$  is a MBER solution, then  $\alpha \mathbf{w}_{\text{MBER}}$  are all MBER solutions for any  $\alpha > 0$ . The gradient of  $P_E(\mathbf{w})$  with respect to  $\mathbf{w}$  is

$$\begin{aligned} \nabla P_E(\mathbf{w}) &= \frac{1}{2N_{sb} \sqrt{2\pi\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{(\bar{y}_{R,q}^{(+)})^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}\right) \\ &\quad \times \text{sgn}(b_{q,d}) \left(\frac{\bar{y}_{R,q}^{(+)}}{\mathbf{w}^H \mathbf{w}} - \bar{\mathbf{x}}_q^{(+)}\right) \end{aligned} \quad (18)$$

With the gradient, the optimisation problem (17) can be solved for iteratively using a gradient-based optimisation algorithm. It is also computationally advantageous to normalise  $\mathbf{w}$  to a unit length after every iteration, so that the gradient can be simplified as

$$\begin{aligned} \nabla P_E(\mathbf{w}) &= \frac{1}{2N_{sb} \sqrt{2\pi\sigma_n^2}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{(\bar{y}_{R,q}^{(+)})^2}{2\sigma_n^2}\right) \\ &\quad \times \text{sgn}(b_{q,d}) \left(\bar{y}_{R,q}^{(+)} \mathbf{w} - \bar{\mathbf{x}}_q^{(+)}\right) \end{aligned} \quad (19)$$

The following simplified conjugate gradient algorithm [26, 16] provides an efficient means to find a MBER solution. The algorithm is summarised:

*Initialisation:* Choose step size  $\mu > 0$  and termination scalar  $\beta > 0$ ; given  $\mathbf{w}(0)$  and  $\mathbf{d}(0) = -\nabla P_E(\mathbf{w}(0))$ ; set iteration index  $\iota = 0$ .

*Loop:* If  $\|\nabla P_E(\mathbf{w}(\iota))\| = \sqrt{(\nabla P_E(\mathbf{w}(\iota)))^H \nabla P_E(\mathbf{w}(\iota))} < \beta$ : goto *Stop*.

$$\begin{aligned}
\mathbf{w}(\iota+1) &= \mathbf{w}(\iota) + \mu \mathbf{d}(\iota) \\
\mathbf{w}(\iota+1) &= \frac{\mathbf{w}(\iota+1)}{\|\mathbf{w}(\iota+1)\|} \\
\phi_\iota &= \frac{\|\nabla P_E(\mathbf{w}(\iota+1))\|^2}{\|\nabla P_E(\mathbf{w}(\iota))\|^2} \\
\mathbf{d}(\iota+1) &= \phi_\iota \mathbf{d}(\iota) - \nabla P_E(\mathbf{w}(\iota+1))
\end{aligned}$$

$\iota = \iota + 1$ , goto *Loop*.

*Stop*:  $\mathbf{w}(\iota)$  is the solution.

At a minimum we have  $\|\nabla P_E(\mathbf{w})\| = 0$ . Hence the termination scalar  $\beta$  determines the accuracy of the solution obtained. The step size  $\mu$  controls the rate of convergence. Typically, a much larger value of  $\mu$  can be used compared to the steepest-descent gradient algorithm. As the BER surface  $P_E(\mathbf{w})$  is highly nonlinear, occasionally the search direction  $\mathbf{d}$  may no longer be a good approximation to the conjugate gradient direction or may even point to the ‘uphill’ direction, when the iteration index becomes large. It is thus advisable to periodically reset  $\mathbf{d}$  to the negative gradient in the above conjugate gradient algorithm.

#### 4 Adaptive minimum bit error rate filtering

In reality, the pdf of  $y_R(k)$  is unknown. The key to adaptive implementation of the MBER filtering solution is an effective estimate of the pdf (12). The Parzen window or kernel density estimate [20–22] is the best known method for estimating a probability distribution. The Parzen window method estimates a pdf using a window or block of  $y_R(k)$  by placing a symmetric unimodal kernel function on each  $y_R(k)$ . Kernel density estimation is capable of producing reliable pdf estimates with short data records and in particular is extremely natural when dealing with Gaussian mixtures, such as the one given in (12). In our particular application, it is obvious and natural to choose a Gaussian kernel function with a kernel width  $\rho_n \sqrt{\mathbf{w}^H \mathbf{w}}$  that is similar in form to the noise standard deviation  $\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}$ .

##### 4.1 Block-data gradient adaptive MBER algorithm

Given a block of  $K$  training samples  $\{\mathbf{x}(k), b_d(k)\}$ , a Parzen window estimate of the pdf (12) is readily given by

$$\hat{p}(y_R) = \frac{1}{K \sqrt{2\pi} \rho_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{k=1}^K \exp\left(-\frac{(y_R - y_R(k))^2}{2\rho_n^2 \mathbf{w}^H \mathbf{w}}\right) \quad (20)$$

where the radius or scaling parameter  $\rho_n$  is related to the standard deviation  $\sigma_n$  of the system noise. Accuracy analysis of the Parzen window density estimate is well documented in the literature. The pdf estimate (20) is known to possess a mean integrated square error convergence rate at order  $K^{-1}$  [20]. Some examples of accurate pdf estimates using (20) with short data records can be seen in [10, 16]. In [21], a lower bound  $\rho_n = (4/3K)^{1/5} \sigma_n$  is suggested. In practice,  $\rho_n$  can often be chosen from a large range of values.

From this estimated pdf, the estimated BER is given by

$$\hat{P}_E(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^K Q(\hat{g}_k(\mathbf{w})) \quad (21)$$

with

$$\hat{g}_k(\mathbf{w}) = \frac{\text{sgn}(b_d(k)) y_R(k)}{\rho_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (22)$$

The gradient of  $\hat{P}_E(\mathbf{w})$  is

$$\begin{aligned}
\nabla \hat{P}_E(\mathbf{w}) &= \frac{1}{2K \sqrt{2\pi} \rho_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{k=1}^K \exp\left(-\frac{y_R^2(k)}{2\rho_n^2 \mathbf{w}^H \mathbf{w}}\right) \\
&\quad \times \text{sgn}(b_d(k)) \left(\frac{y_R(k) \mathbf{w}}{\mathbf{w}^H \mathbf{w}} - \mathbf{x}(k)\right) \quad (23)
\end{aligned}$$

By substituting  $\nabla P_E(\mathbf{w})$  with  $\nabla \hat{P}_E(\mathbf{w})$  in the conjugate gradient updating mechanism, a block-data gradient adaptive MBER algorithm is readily obtained. The step size  $\mu$  and radius parameter  $\rho_n$  are the two algorithm parameters that need to be chosen.

##### 4.2 Stochastic gradient adaptive MBER algorithm

In the Parzen window estimate (20), the kernel width  $\rho_n \sqrt{\mathbf{w}^H \mathbf{w}}$  depends on the filter weight vector  $\mathbf{w}$ . In a general density estimate, there is no reason why the kernel width should be chosen in such a way except that we notice the dependency of the noise standard deviation on  $\mathbf{w}$  in the true density (12). However, the BER is invariant to  $\mathbf{w}^H \mathbf{w}$ . To fully take advantage of this fact, a constant width  $\rho_n$  in density estimate can be used. One advantage of using a constant width  $\rho_n$ , rather than a variable one,  $\rho_n \sqrt{\mathbf{w}^H \mathbf{w}}$ , in the density estimate is that the gradient of the resulting estimated BER has a much simpler form, which leads to considerable reduction in computational complexity. This is particularly relevant in the derivation of stochastic gradient updating mechanisms. Adopting this approach, an alternative Parzen window estimate of the true pdf (12) is given by

$$\bar{p}(y_R) = \frac{1}{K \sqrt{2\pi} \rho_n} \sum_{k=1}^K \exp\left(-\frac{(y_R - y_R(k))^2}{2\rho_n^2}\right) \quad (24)$$

and an approximation of the BER is

$$\hat{P}_E(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^K Q(\tilde{g}_k(\mathbf{w})) \quad (25)$$

with

$$\tilde{g}_k(\mathbf{w}) = \frac{\text{sgn}(b_d(k)) y_R(k)}{\rho_n} \quad (26)$$

This approximation is valid provided that the width  $\rho_n$  is chosen appropriately.

To derive a sample-by-sample adaptive algorithm, consider a single-sample estimate of  $p(y_R)$ , namely

$$\tilde{p}(y_R, k) = \frac{1}{\sqrt{2\pi} \rho_n} \exp\left(-\frac{(y_R - y_R(k))^2}{2\rho_n^2}\right) \quad (27)$$

Conceptually, from this one-sample pdf ‘estimate’, we have a one-sample or instantaneous BER ‘estimate’  $\tilde{P}_E(\mathbf{w}, k)$ . Using the instantaneous stochastic gradient of

$$\nabla \tilde{P}_E(\mathbf{w}, k) = -\frac{\text{sgn}(b_d(k))}{2\sqrt{2\pi} \rho_n} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2}\right) \mathbf{x}(k) \quad (28)$$

gives rise to the following stochastic gradient adaptive MBER algorithm:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\text{sgn}(b_d(k))}{2\sqrt{2\pi} \rho_n} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2}\right) \mathbf{x}(k) \quad (29)$$

This adaptive algorithm has a similar computational complexity to the very simple LMS algorithm, which is given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(b_d(k) - y(k))\mathbf{x}^*(k) \quad (30)$$

It is interesting to see some analogy between the traditional adaptive filtering approach based on the MMSE criterion and the proposed adaptive MBER filtering approach. The second-order statistics required to compute the Wiener solution can be estimated using a block of samples, and, by considering a single-sample estimate, a stochastic gradient adaptive MMSE algorithm, namely the LMS, is derived. The pdf required to determine the MBER solution can be approximated with a kernel density estimate based on a block of samples, and by considering a single-sample density estimate, a stochastic gradient adaptive MBER algorithm is formulated. The adaptive gain  $\mu$  and kernel width  $\rho_n$  for the adaptive algorithm (29) should be chosen appropriately to ensure good performance in terms of convergence speed and steady-state BER misadjustment. Note that the adaptive algorithm (29) belongs to the general stochastic gradient-based adaptive algorithm investigated in [27]. Therefore, the results of convergence analysis presented in [27] can readily be applied here.

## 5 Application examples

The effectiveness of the proposed adaptive MBER filtering approach is demonstrated using two applications.

### 5.1 Single-user channel equalisation

In the single-user communication system involving a dispersive channel, the received signal sample can be

expressed as

$$x(k) = \sum_{i=0}^{n_a-1} a_i b(k-i) + n(k) \quad (31)$$

where  $n_a$  is the length of the channel impulse response,  $a_i$  are complex-valued channel taps, and  $\{b(k)\}$  is the transmitted data symbol sequence. The linear equaliser at the receiver is a linear filter in the form of (1), where  $\mathbf{x}(k) = [x(k) \ x(k-1) \ \dots \ x(k-L+1)]^T$  with  $L$  known as the equaliser order. The  $L \times M$  system matrix  $\mathbf{P}$  in (2) has the form

$$\mathbf{P} = \begin{bmatrix} a_0 & a_1 & \dots & a_{n_a-1} & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_{n_a-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \dots & \ddots & 0 \\ 0 & \dots & 0 & a_0 & a_1 & \dots & a_{n_a-1} \end{bmatrix} \quad (32)$$

with  $M = L + n_a - 1$ , and the symbol vector  $\mathbf{b}(k) = [b(k) \ b(k-1) \ \dots \ b(k-L-n_a+2)]^T$ . The equaliser provides an estimate  $\hat{b}(k-d)$  of the transmitted symbol  $b(k-d)$ , where the integer  $d$  is called the equaliser delay. The system signal to noise ratio (SNR) is defined as  $\text{SNR} = \mathbf{a}^H \mathbf{a} / 2\sigma_n^2$ , where  $\mathbf{a} = [a_0 \ a_1 \ \dots \ a_{n_a-1}]^T$  is the channel tap vector.

In the simulation study, the following three-tap channel,

$$\mathbf{a}^T = [-0.5 + j0.4 \quad 0.7 + j0.6 \quad 0.4 + j0.3] \quad (33)$$

and a five-tap ( $L = 5$ ) equaliser were used. For this example, it was found that the optimal equaliser delay is  $d = 3$ . Figure 1 compares the BER performance of the MMSE equaliser with that of the MBER one. Table 1 lists the MMSE and MBER solutions,  $\mathbf{w}_{\text{MMSE}}$  and  $\mathbf{w}_{\text{MBER}}$ ,

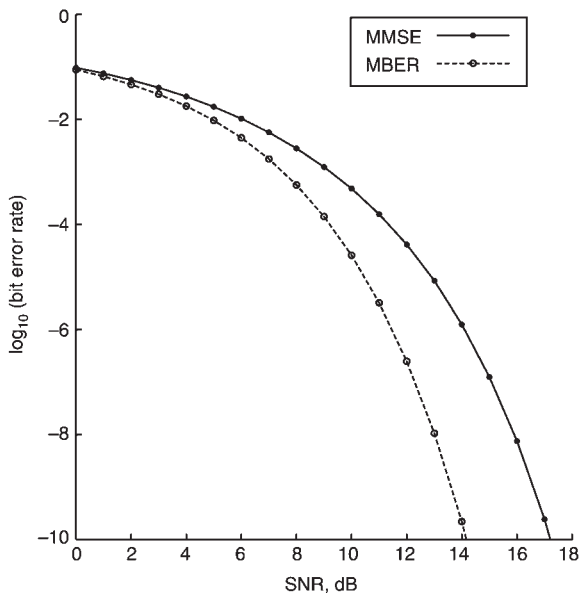


Fig. 1 Bit error rate performance comparison of the MMSE and MBER equalisers

Table 1: Weight vectors and BERs of the MMSE and MBER equalisers given SNR = 12 dB

	MMSE	MBER
$\mathbf{w}$	$-0.162742 + j0.084336$	$-0.008256 - j0.008298$
	$0.082974 + j0.365218$	$0.131712 + j0.175459$
	$0.635619 + j0.194423$	$0.636324 + j0.349129$
	$-0.338643 + j0.022754$	$-0.535888 + j0.292630$
	$0.151703 + j0.010665$	$0.212291 - j0.083981$
BER	$4.15 \times 10^{-5}$	$2.49 \times 10^{-7}$

The weight vector has been normalised to a unit length

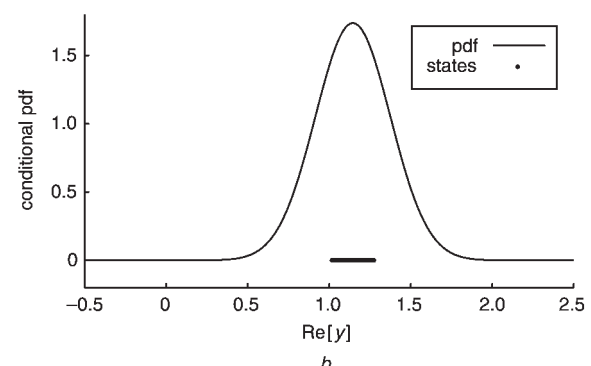
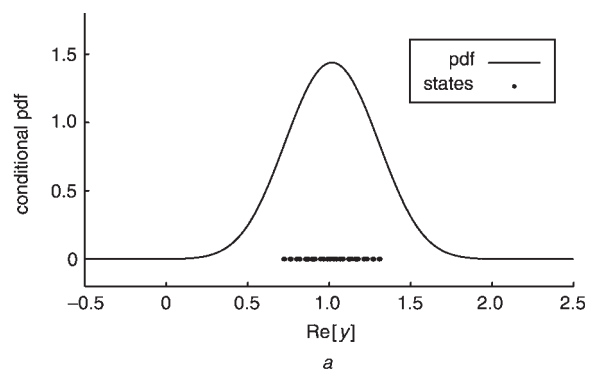


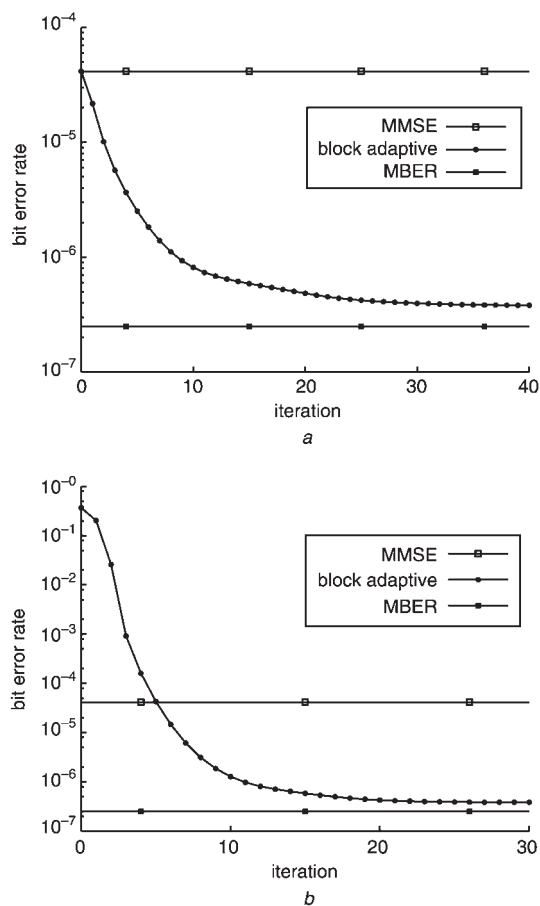
Fig. 2 Conditional probability density functions and subsets  $\mathcal{Y}_R^{(+)}$  of the MMSE and MBER equalisers for SNR = 12 dB

The equaliser weight vector has been normalised to a unit length  
a MMSE  
b MBER



together with the associated BERs given SNR = 12 dB. From Table 1, it can be seen that  $w_{\text{MMSE}}$  and  $w_{\text{MBER}}$  are very different. The conditional pdf of the real part of the MMSE equaliser output, given  $b(k-d) = +1$  and with a SNR = 12 dB, is compared with that of the MBER equaliser in Fig. 2. For this example, the subset  $\mathcal{Y}_R^{(+)}$  contains  $N_{sb} = 64$  points. In Fig. 2 the equaliser weight vector has been normalised to a unit length, so that the BER is mainly determined by the minimum distance from the subset  $\mathcal{Y}_R^{(+)}$  to the decision threshold  $y_R = 0$ . It can be seen from Fig. 2 that the minimum distance between  $y_R = 0$  and  $\mathcal{Y}_R^{(+)}$  for the MMSE equaliser is smaller than that for the MBER equaliser. Also, the density distribution of  $y_R(k)$  for the MMSE equaliser is broader than that for the MBER equaliser. This means that the noise  $e(k)$  at the MMSE equaliser output has a larger variance. These two factors explain why the MMSE equaliser has a higher BER than the MBER equaliser.

The performance of the block-data based adaptive MBER algorithm employing the conjugate gradient updating mechanism, as described in Section 4.1, was next studied. Figure 3 illustrates the convergence rate of the algorithm under SNR = 12 dB and given two different initial weight vector conditions, where the block size was  $K = 200$ . From Fig. 3, it can be seen that the convergence speed of this block-data gradient adaptive MBER algorithm is rapid. The performance of the stochastic gradient adaptive MBER algorithm discussed in Section 4.2 was then investigated. Figure 4 shows the learning curves of the algorithm



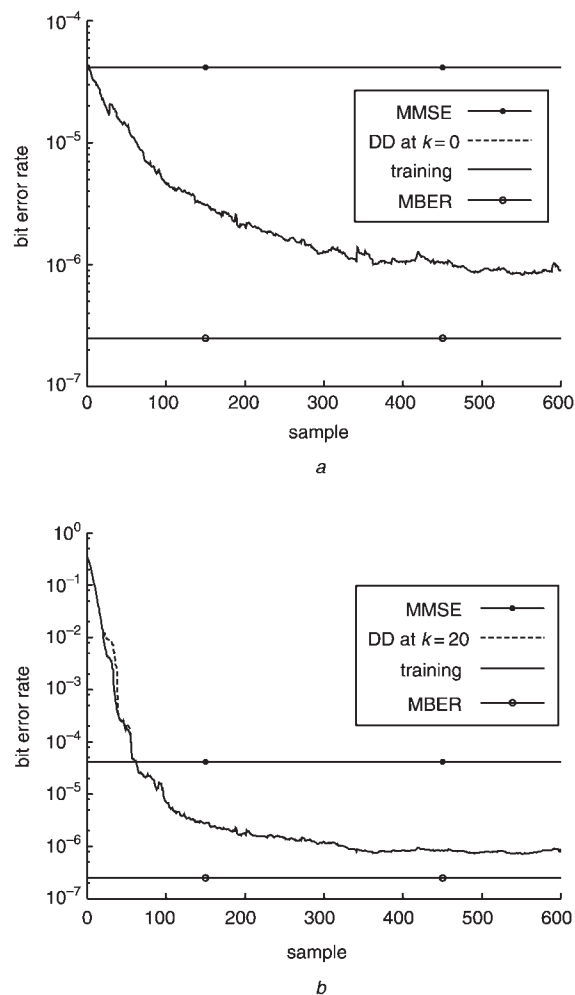
**Fig. 3** Convergence rate of the block-data gradient adaptive MBER algorithm for the equalisation example given SNR = 12 dB, and with a block size  $K = 200$ , step size  $\mu = 0.9$  and square width  $\rho_n^2 = 3\sigma_n^2 \approx 0.14$

a  $w(0) = w_{\text{MMSE}}$   
b  $w(0) = [0.1 + j0.0 \quad 0.1 + j0.0 \quad 0.1 + j0.0 \quad 0.1 + j0.0 \quad 0.1 + j0.0]^T$

averaged over 30 runs, given SNR = 12 dB and two different initial weight vector conditions. From Fig. 4, it can be seen that this stochastic gradient adaptive MBER algorithm has a reasonable convergence rate. There are in fact two learning curves in each of Figs. 4a and 4b, corresponding to training and decision directed (DD) adaptation in which  $b(k-d)$  is substituted by the equaliser's estimate  $\hat{b}(k-d)$ . In Fig. 4a, the initial BER is well below the level of  $10^{-4}$ , and the two learning curves are indistinguishable. It can also be seen that, once the BER is below a certain level (typically 0.01), decision-directed adaptation can be applied with little performance degradation, as is confirmed in Fig. 4b.

## 5.2 Adaptive beamforming assisted receiver

The ever-increasing demand for mobile communication capacity has motivated the employment of space division multiple access for the sake of improving the achievable spectral efficiency. A particular approach that has shown real promise in achieving substantial capacity enhancements is the use of adaptive antenna arrays. Adaptive beamforming is capable of separating signals transmitted on the same carrier frequency, provided that they are separated

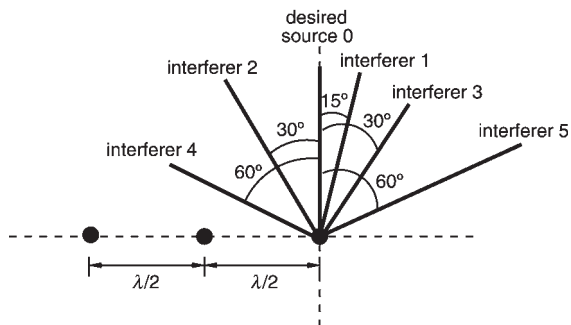


**Fig. 4** Learning curves of the stochastic gradient adaptive MBER algorithm averaged over 30 runs for the equalisation example given SNR = 12 dB, where DD denotes decision directed adaptation with  $\hat{b}(k-d)$  substituting  $b(k-d)$

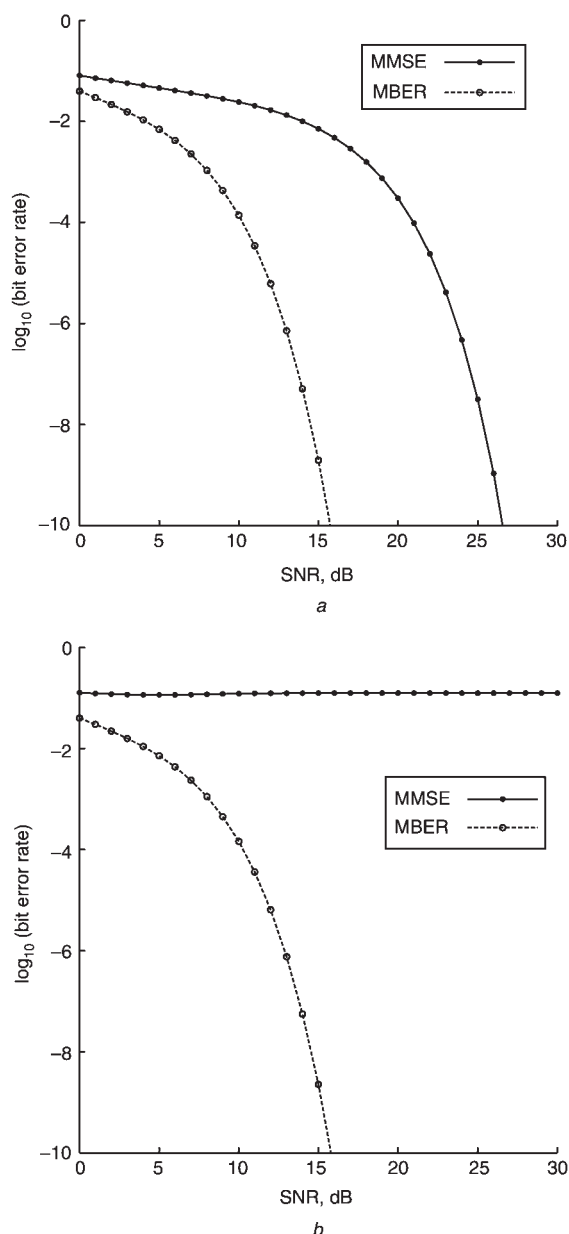
In a, the two learning curves corresponding to training and DD adaptation are indistinguishable

a  $w(0) = w_{\text{MMSE}}$ , step size  $\mu = 0.1$  and square width  $\rho_n^2 = 3\sigma_n^2 \approx 0.14$   
b  $w(0) = [0.1 + j0.0 \quad 0.1 + j0.0 \quad 0.1 + j0.0 \quad 0.1 + j0.0 \quad 0.1 + j0.0]^T$ , step size  $\mu = 0.3$  and square width  $\rho_n^2 = 4\sigma_n^2 \approx 0.19$

in the spatial domain. Consider the system that supports  $M$  users (sources) which transmit on the same carrier frequency  $\omega = 2\pi f$ , and assume that the channel is narrowband which does not induce intersymbol



**Fig. 5** Locations of the desired source and the interfering sources with respect to the three-element linear antenna array having  $\lambda/2$  element spacing, where  $\lambda$  is the wavelength



**Fig. 6** Bit error rate performance comparison of the MMSE and MBER beamformers

*a*  $SIR_i = 0$  dB for  $1 \leq i \leq 5$

*b*  $SIR_i = 0$  dB for  $i = 1, 3, 4, 5$ , and  $SIR_2 = -6$  dB

interference. The linear antenna array considered consists of  $L$  uniformly spaced elements, and the signals received by the  $L$ -element antenna array can be expressed in the form of (2), where the  $L \times M$  system matrix  $\mathbf{P}$  is defined by

$$\mathbf{P} = [A_0 s_0 \ A_1 s_1 \ \cdots \ A_{M-1} s_{M-1}] \quad (34)$$

$A_i^2$  denotes the signal power of user  $i$ , and the steering vector for source  $i$  is given by

$$s_i = [\exp(j\omega t_0(\theta_i)) \ \exp(j\omega t_1(\theta_i)) \ \cdots \ \exp(j\omega t_{L-1}(\theta_i))]^T \quad (35)$$

with  $t_l(\theta_i)$  being the relative time delay at array element  $l$  for source  $i$  and  $\theta_i$  the direction of arrival for source  $i$ . The transmitted user symbol vector is  $\mathbf{b}(k) = [b_0(k) \ b_1(k) \ \cdots \ b_{M-1}(k)]^T$ . Without any loss of generality, source 0 is assumed to be the desired user and the rest of the sources are the interfering users. The desired user's signal to noise ratio is defined as  $SNR = A_0^2/2\sigma_n^2$  and the desired signal to interference ratio with respect to interfering user  $i$  is defined as  $SIR_i = A_0^2/A_i^2$  for  $1 \leq i \leq M-1$ . The beamformer at the receiver is a linear filter in the form of (1) with  $d = 0$  in the decision rule (3).

The simulation example used consisted of six signal sources and a three-element antenna array. Figure 5 shows the locations of the desired source and the interfering sources graphically. Figure 6 compares the BER performance of the MMSE beamforming solution with that of the MBER one under two different conditions: (a) the desired user and all the five interfering sources had equal power, and (b) the desired user and the interfering sources 1, 3, 4, 5 had equal power, but the interfering source 2 had 6 dB higher power than the desired user. Note that when the 2nd interfering user's power is increased by 6 dB, the MMSE beamformer's performance breaks down, while the performance of the MBER beamformer remains almost unchanged. Thus, the MBER beamformer is robust to the so-called near-far effect. Tables 2 and 3 list the MMSE and MBER solutions together with the associated BERs given a fixed  $SNR = 14$  dB and under the two given SIR conditions, respectively. Examining the two weight vectors for the

**Table 2: Weight vectors and BERs of the MMSE and MBER beamformers given  $SNR = 14$  dB and  $SIR_i = 0$  dB for  $1 \leq i \leq 5$**

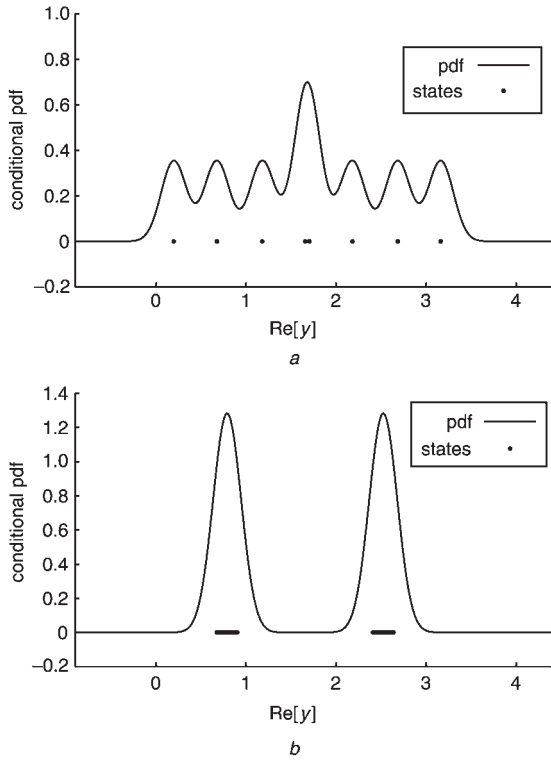
	MMSE	MBER
$\mathbf{w}$	0.170769 - j0.053050	0.448072 - j0.060545
	0.186144 - j0.000000	0.783087 - j0.035493
	0.170769 + j0.053050	0.425536 + j0.000000
BER	$1.00 \times 10^{-2}$	$5.40 \times 10^{-8}$

The weight vector has been normalised to a unit length

**Table 3: Weight vectors and BERs of the MMSE and MBER beamformers given  $SNR = 14$  dB,  $SIR_i = 0$  dB for  $i = 1, 3, 4, 5$ , and  $SIR_2 = -6$  dB**

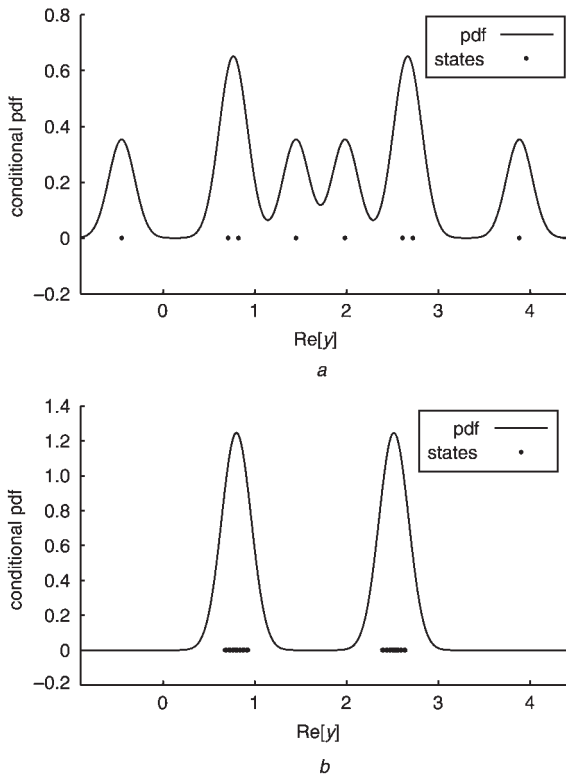
	MMSE	MBER
$\mathbf{w}$	0.159405 + j0.009740	0.450894 - j0.057414
	0.120395 - j0.000000	0.783001 - j0.030264
	0.159405 - j0.009740	0.423547 - j0.000000
BER	$1.25 \times 10^{-1}$	$5.60 \times 10^{-8}$

The weight vector has been normalised to a unit length



**Fig. 7** Conditional probability density functions and subsets  $\mathcal{Y}_R^{(+)}$  of the MMSE and MBER beamformers given SNR = 14 dB and  $SIR_i = 0$  dB for  $1 \leq i \leq 5$

The beamformer weight vector has been normalised to a unit length  
a MMSE  
b MBER

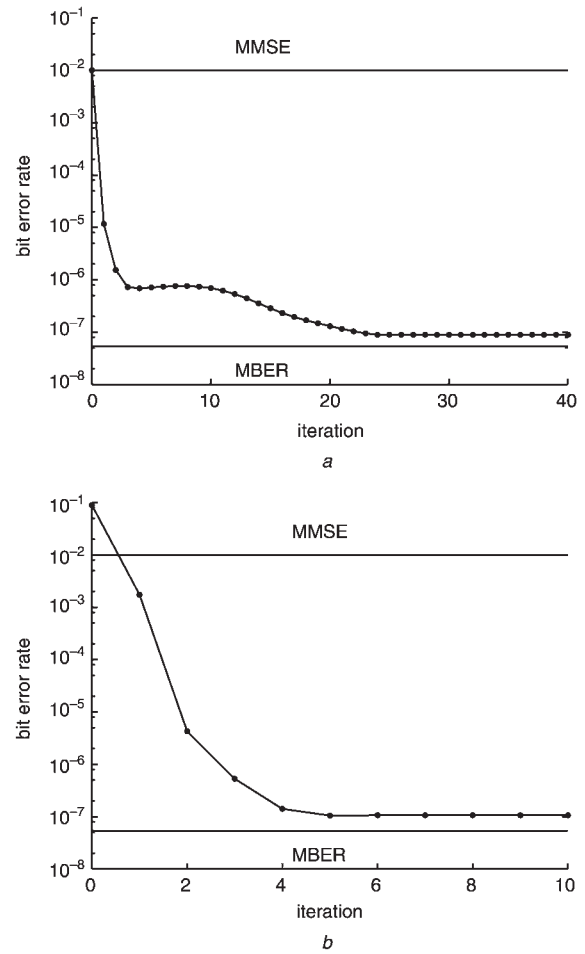


**Fig. 8** Conditional probability density functions and subsets  $\mathcal{Y}_R^{(+)}$  of the MMSE and MBER beamformers given SNR = 14 dB,  $SIR_i = 0$  dB for  $i = 1, 3, 4, 5$ , and  $SIR_2 = -6$  dB

The beamformer weight vector has been normalised to a unit length  
a MMSE  
b MBER

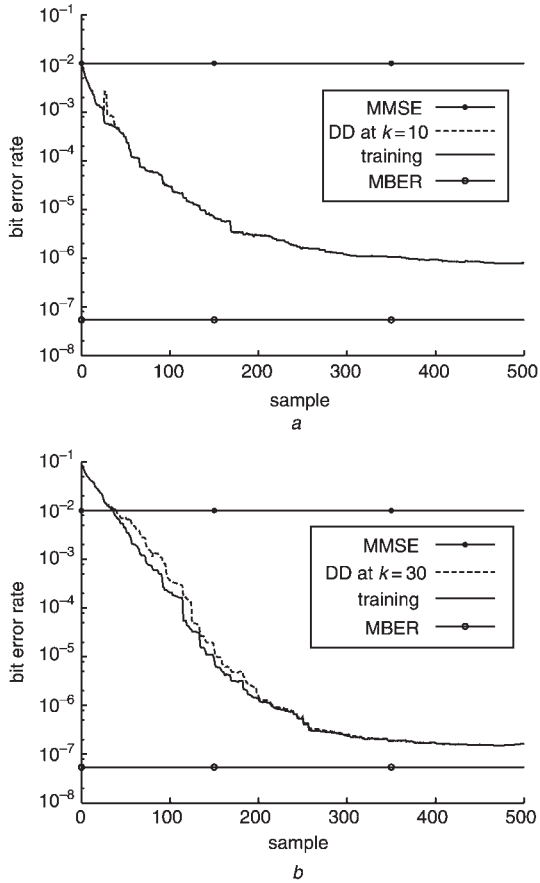
MBER solution under the two different conditions, they are very similar. However, differences between the two weight vectors of the MMSE solution can be clearly seen from Tables 2 and 3. Figures 7 and 8 depict the conditional pdfs of the MMSE and MBER beamformers given  $b_0(k) = +1$  together with the associated subsets  $\mathcal{Y}_R^{(+)}$ , under the same two conditions as given in Tables 2 and 3. Again, in these two figures, the beamformer weight vector had been normalised to a unit length. It is interesting to see that, given SNR = 14 dB,  $SIR_i = 0$  dB for  $i = 1, 3, 4, 5$  and  $SIR_2 = -6$  dB, the resulting  $\mathcal{Y}_R^{(-)}$  and  $\mathcal{Y}_R^{(+)}$  for the MMSE beamforming becomes linearly inseparable. There are  $N_{sb} = 32$  points in  $\mathcal{Y}_R^{(+)}$ , and a cluster of four points is on the wrong side of the decision boundary  $y_R = 0$  for the MMSE beamforming, giving rise to a high BER floor  $4/32 = 0.125$ .

Performance of the block-data gradient adaptive MBER algorithm portrayed in Section 4.1 was next tested. Figure 9 illustrates the convergence rates of the algorithm given SNR = 14 dB and  $SIR_i = 0$  dB for  $1 \leq i \leq 5$ , and with the two different initial weight vectors. It can be seen that this block-data adaptive MBER algorithm generally converges rapidly. As the BER surface is highly complicated, the initial condition will influence convergence rate. It has been found out in a variety of applications that the MMSE solution  $\mathbf{w}_{MMSE}$  is typically not a good initial choice for the adaptive MBER algorithm in terms of convergence rate. Performance



**Fig. 9** Convergence rate of block-data gradient adaptive MBER algorithm for the beamforming example given SNR = 14 dB and  $SIR_i = 0$  dB for  $1 \leq i \leq 5$ , and with a block size  $K = 200$ , step size  $\mu = 0.6$  and a square width  $\rho_n^2 = 4\sigma_n^2 \approx 0.08$

a  $\mathbf{w}(0) = \mathbf{w}_{MMSE}$   
b  $\mathbf{w}(0) = [0.1 + j0.0 \quad 0.1 + j0.0 \quad 0.1 + j0.0]^T$



**Fig. 10** Learning curves of stochastic gradient adaptive MBER algorithm averaged over 30 runs for the beamforming example given  $\text{SNR} = 14 \text{ dB}$  and  $\text{SIR}_i = 0 \text{ dB}$  for  $1 \leq i \leq 5$ , where DD denotes decision directed adaptation with  $\hat{b}_0(k)$  substituting  $b_0(k)$   
a  $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$ , step size  $\mu = 0.03$  and square width  $\rho_n^2 = 2\sigma_n^2 \approx 0.04$   
b  $\mathbf{w}(0) = [0.1 + j0.0 \ 0.1 + j0.0 \ 0.1 + j0.0]^T$ , step size  $\mu = 0.02$  and square width  $\rho_n^2 = 4\sigma_n^2 \approx 0.08$

of the stochastic gradient adaptive MBER algorithm described in Section 4.2 was also investigated. Figure 10 shows the learning curves of the algorithm under the same conditions of Fig. 9, where DD denotes decision-directed adaptation with  $\hat{b}_0(k)$ , substituting  $b_0(k)$  as the desired response. It can be seen that this stochastic gradient adaptive MBER algorithm has a reasonable convergence speed. Note that the steady-state BER misadjustment is higher when the initial weight vector is set to  $\mathbf{w}_{\text{MMSE}}$ , compared with the other initial weight condition.

## 6 Extension to nonlinear filtering

For a linear filter to work satisfactorily in a communication application, an implicit assumption is that  $\mathcal{X}^{(+)}$  and  $\mathcal{X}^{(-)}$  defined in (5) are linearly separable. That is, there exists a weight vector  $\mathbf{w}$  such that the two resulting scalar sets  $\mathcal{Y}_R^{(+)}$  and  $\mathcal{Y}_R^{(-)}$  are completely separated by the decision threshold  $y_R = 0$ . Otherwise nonlinear filtering is required. Examples of such a nonlinear filtering includes nonlinear single-user equalisation and nonlinear multiuser detection [28–31]. If the linear restriction is removed, it can readily be shown that the true optimal filtering solution in terms of BER is the maximum a posteriori probability or Bayesian one, which is formulated as

$$y_B(k) = \frac{1}{N_b(2\pi\sigma_n^2)^L} \sum_{q=1}^{N_b} \text{sgn}(b_{q,d}) \exp\left(\frac{-\|\mathbf{x}(k) - \bar{\mathbf{x}}_q\|^2}{2\sigma_n^2}\right) \quad (36)$$

where  $\bar{\mathbf{x}}_q \in \mathcal{X}$  and  $\text{sgn}(b_{q,d})$  acts as a class label. Note that  $y_B(k)$  is real-valued due to its pdf interpretation, but  $\mathbf{x}(k)$  and  $\bar{\mathbf{x}}_q$  are complex-valued. Because the number of vector states  $N_b$  is generally very large, this optimal Bayesian filtering solution is computationally very expensive. In an adaptive implementation, all the states  $\bar{\mathbf{x}}_q$  have to be identified by some means.

Consider the general nonlinear filter, which takes the form

$$y(k) = f(\mathbf{x}(k); \mathbf{w}) \quad (37)$$

where the nonlinear map  $f$  is generally complex-valued and is realised, for example, by a neural network, and the vector  $\mathbf{w}$  consists of all the adjustable parameters of the nonlinear filter. Classically, adaptive training of such a nonlinear structure is based on the MMSE principle. For example, sample-by-sample adaptation is typically implemented using the LMS algorithm

$$\left. \begin{aligned} y(k) &= f(\mathbf{x}(k); \mathbf{w}(k-1)) \\ \mathbf{w}(k) &= \mathbf{w}(k-1) + \mu(b_d(k) - y(k)) \frac{\partial f^*(\mathbf{x}(k); \mathbf{w}(k-1))}{\partial \mathbf{w}} \end{aligned} \right\} \quad (38)$$

However, the true performance criterion of the system is the BER, and ideally the system design should be based on minimising the BER.

By linearising the nonlinear filter (37) around  $\bar{\mathbf{x}}(k)$ , it can be approximated as

$$y(k) \approx \bar{y}(k) + e(k) \quad (39)$$

where

$$\bar{y}(k) = f(\bar{\mathbf{x}}(k); \mathbf{w}) \quad (40)$$

can only take the value from the set

$$\tilde{\mathcal{Y}} \triangleq \{\bar{y}_q = f(\bar{\mathbf{x}}_q; \mathbf{w}), 1 \leq q \leq N_b\} \quad (41)$$

and

$$e(k) = \left[ \frac{\partial f(\bar{\mathbf{x}}(k); \mathbf{w})}{\partial \mathbf{x}} \right]^H \mathbf{n}(k) \quad (42)$$

is Gaussian with zero mean and variance

$$\tilde{\rho}_n^2(\mathbf{w}) = \frac{2\sigma_n^2}{N_b} \sum_{q=1}^{N_b} \left[ \frac{\partial f(\bar{\mathbf{x}}_q; \mathbf{w})}{\partial \mathbf{x}} \right]^H \frac{\partial f(\bar{\mathbf{x}}_q; \mathbf{w})}{\partial \mathbf{x}} \quad (43)$$

The pdf of  $y_R(k)$  can then be approximated by

$$p(y_R) \approx \frac{1}{N_b \sqrt{2\pi\tilde{\rho}_n^2(\mathbf{w})}} \sum_{q=1}^{N_b} \exp\left(-\frac{(y_R - \bar{y}_{R,q})^2}{2\tilde{\rho}_n^2(\mathbf{w})}\right) \quad (44)$$

and the BER of the nonlinear filter is

$$P_E(\mathbf{w}) \approx \frac{1}{N_b} \sum_{q=1}^{N_b} Q(g_q(\mathbf{w})) \quad (45)$$

with

$$g_q(\mathbf{w}) = \frac{\text{sgn}(b_{q,d})\bar{y}_{R,q}}{\tilde{\rho}_n(\mathbf{w})} = \frac{\text{sgn}(b_{q,d})f_R(\bar{\mathbf{x}}_q; \mathbf{w})}{\tilde{\rho}_n(\mathbf{w})} \quad (46)$$

Using the kernel density estimate in the form of (24) with a constant  $\rho_n^2$  to approximate the density (44) naturally leads to a block-data based gradient adaptive near the MBER algorithm for training the nonlinear filter (37). This can be further simplified to give rise to a stochastic gradient adaptive near MBER algorithm in the form:



$$\left. \begin{aligned} y(k) &= f(\mathbf{x}(k); \mathbf{w}(k-1)) \\ \mathbf{w}(k) &= \mathbf{w}(k-1) + \mu \frac{\text{sgn}(b_d(k))}{2\sqrt{2\pi\rho_n}} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2}\right) \frac{\partial f_R(\mathbf{x}(k); \mathbf{w}(k-1))}{\partial \mathbf{w}} \end{aligned} \right\} \quad (47)$$

for a sample-by-sample adaptation. The derivative  $\partial f_R / \partial \mathbf{w}$  depends on the particular nonlinear map employed. Previous studies [32–34] have applied this adaptive near MBER nonlinear filtering approach to equalisation and multiuser detection applications.

## 7 Conclusions

A general adaptive filtering technique has been proposed for applications to communication systems based on the novel MBER principle. It has been demonstrated that the MBER filtering is capable of achieving significant performance gains in terms of reduced BER over the traditional MMSE filtering. Adaptive implementation of the proposed MBER filtering has been developed based on the classical Parzen window estimation for the pdf of the filter's output. A block-data based conjugate gradient adaptive MBER algorithm has been shown to converge rapidly and requires a reasonably small data block size to accurately approximate the theoretical MBER solution. An LMS-style stochastic gradient adaptive MBER algorithm has been shown to perform well, and the algorithm has similar computational requirements to the low-complexity LMS algorithm. Extension of this adaptive MBER filtering approach to nonlinear filtering has been discussed.

## 8 Acknowledgment

Original contributions of Professor B. Mulgrew to the topic of adaptive MBER equalisation are gratefully acknowledged.

## 9 References

- Widrow, B., and Stearns, S.D.: 'Adaptive signal processing' (Prentice-Hall, Englewood Cliffs, NJ, 1985)
- Haykin, S.: 'Adaptive filter theory' (Prentice-Hall, Upper Saddle River, NJ, 1996, 3rd edn.)
- Shamash, E., and Yao, K.: 'On the structure and performance of a linear decision feedback equalizer based on the minimum error probability criterion'. Proc. ICC, 1974, pp. 25F1–25F5
- Chen, S., Chng, E.S., Mulgrew, B., and Gibson, G.: 'Minimum-BER linear-combiner DFE'. Proc. ICC, Dallas, Texas, 1996, vol. 2, pp. 1173–1177
- Yeh, C.C., and Barry, J.R.: 'Approximate minimum bit-error rate equalization for binary signaling'. Proc. ICC, Montreal, Canada, 1997, vol. 2, pp. 1095–1099
- 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
- Chen, S., and Mulgrew, B.: 'The minimum-SER linear-combiner decision feedback equalizer'. *IEE Proc., Commun.*, 1999, **146**, (6), pp. 347–353
- Mulgrew, B., and Chen, S.: 'Stochastic gradient minimum-BER decision feedback equalisers'. Proc. IEEE Symposium on Adaptive systems for signal processing, communication and control, Lake Louise, Alberta, Canada, 1–4 Oct. 2000, pp. 93–98
- 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
- Mulgrew, B., and Chen, S.: 'Adaptive minimum-BER decision feedback equalisers for binary signalling'. *Signal Process.*, 2001, **81**, (7), pp. 1479–1489
- Mandayam, N.B., and Aazhang, B.: 'Gradient estimation for sensitivity analysis and adaptive multiuser interference rejection in code-division multiple-access systems'. *IEEE Trans. Commun.*, 1997, **45**, (7), pp. 848–858
- Yeh, C.C., Lopes, R.R., and Barry, J.R.: 'Approximate minimum bit-error rate multiuser detection'. Proc. Globecom, Sydney, Australia, Nov. 1998, pp. 3590–3595
- Wang, X.F., Lu, W.S., and Antoniou, A.: 'Constrained minimum-BER multiuser detection'. Proc. ICASSP, Phoenix, USA, 14–18 May 1999, vol. 5, pp. 2603–2606

- Psaromiligkos, I.N., Batalama, S.N., and Pados, D.A.: 'On adaptive minimum probability of error linear filter receivers for DS-CDMA channels'. *IEEE Trans. Commun.*, 1999, **47**, (7), pp. 1092–1102
- Chen, S., Samingan, A.K., Mulgrew, B., and Hanzo, L.: 'Adaptive minimum-BER linear multiuser detection'. Proc. ICASSP, Salt Lake City, Utah, USA, 7–11 May 2001, vol. 4, pp. 2253–2256
- 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
- Chen, S., Ahmad, N.N., and Hanzo, L.: 'Smart beamforming for wireless communications: a novel minimum bit error rate approach'. Proc. 2nd IMA Int. Conf. on mathematics in communications, Lancaster, UK, 16–18 Dec. 2002
- Chen, S., Hanzo, L., and Ahmad, N.N.: 'Adaptive minimum bit error rate beamforming assisted receiver for wireless communications'. Proc. ICASSP, Hong Kong, China, 6–10 April 2003, vol. IV, pp. 640–643
- Chen, S., Ahmad, N.N., and Hanzo, L.: 'Adaptive minimum bit error rate beamforming', submitted to *IEEE Trans. Wireless Communications*, 2002
- Parzen, E.: 'On estimation of probability density function and mode'. *Ann. Math. Stat.*, 1962, **33**, pp. 1066–1076
- Silverman, B.W.: 'Density estimation' (Chapman Hall, London, 1996)
- Bowman, A.W., and Azzalini, A.: 'Applied smoothing techniques for data analysis' (Oxford University Press, Oxford, 1997)
- Chen, S., Mulgrew, B., and Hanzo, L.: 'Stochastic least-symbol-error-rate adaptive equalization for pulse-amplitude modulation'. Proc. ICASSP, Orlando, Florida, USA, 13–17 May 2002, vol. 3, pp. 2629–2632
- Chen, S.: 'Minimum symbol-error-rate equalisation'. Proc. EPSRC/IEE Non-linear and non-Gaussian signal processing workshop, Peebles, Scotland, 8–9 July 2002
- Chen, S., Hanzo, L., and Mulgrew, B.: 'Adaptive minimum symbol-error-rate decision feedback equalization for multi-level pulse-amplitude modulation', submitted to *IEEE Trans. Signal Processing*, 2002
- Bazaraa, M.S., Sherali, H.D., and Shetty, C.M.: 'Nonlinear programming: theory and algorithms' (John Wiley, New York, 1993)
- Sharma, R., Sethares, W.A., and Buckle, J.A.: 'Asymptotic analysis of stochastic gradient-based adaptive filtering algorithms with general cost functions'. *IEEE Trans. Signal Process.*, 1996, **44**, (9), pp. 2186–2194
- Chen, S., Mulgrew, B., and Grant, P.M.: 'A clustering technique for digital communications channel equalisation using radial basis function networks'. *IEEE Trans. Neural Netw.*, 1993, **4**, (4), pp. 570–579
- Chen, S., Mulgrew, B., and McLaughlin, S.: 'Adaptive Bayesian equaliser with decision feedback'. *IEEE Trans. Signal Process.*, 1993, **41**, (9), pp. 2918–2927
- 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
- Chen, S., Samingan, A.K., and Hanzo, L.: 'Support vector machine multiuser receiver for DS-CDMA signals in multipath channels'. *IEEE Trans. Neural Netw.*, 2001, **12**, (3), pp. 604–611
- Chen, S., Mulgrew, B., and Hanzo, L.: 'Adaptive least error rate algorithm for neural network classifier'. Proc. IEEE Workshop on Neural networks for signal processing, Falmouth, MA, USA, 10–12 Sept. 2001, pp. 223–232
- Chen, S., Mulgrew, B., and Hanzo, L.: 'Least bit error rate adaptive nonlinear equalizers for binary signalling'. *IEE Proc., Commun.*, 2003, **150**, (1), pp. 29–36
- Chen, S.: 'Least bit error rate adaptive multiuser detection', in Wang, L.P. (Ed.): 'Soft computing in communications' (Springer Verlag, 2003), pp. 389–408

## 10 Appendix

The conditional pdf of  $y_R(k)$  given  $b_d(k) = +1$  is

$$p(y_R|+) = \frac{1}{N_{sb} \sqrt{2\pi\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{(y_R - \bar{y}_{R,q}^{(+)})^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}\right) \quad (48)$$

where  $\bar{y}_{R,q}^{(+)} \in \mathcal{Y}_R^{(+)}$ . Thus the conditional BER of the linear filter (6) given  $b_d(k) = +1$  is

$$P_{E,+}(\mathbf{w}) = \int_{-\infty}^0 p(y_R|+) dy_R = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q(g_{q,+}(\mathbf{w})) \quad (49)$$

with

$$g_{q,+}(\mathbf{w}) = \frac{\bar{y}_{R,q}^{(+)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} = \frac{\text{sgn}(b_{q,d}) \bar{y}_{R,q}^{(+)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \\ = \frac{\text{sgn}(b_{q,d}) \Re[\mathbf{w}^H \bar{\mathbf{x}}_q^{(+)}]}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (50)$$

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{u^2}{2}\right) du \quad (51)$$

Similarly, the conditional pdf of  $y_R(k)$  given  $b_d(k) = -1$  is

$$p(y_R|-) = \frac{1}{N_{sb} \sqrt{2\pi\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \sum_{q=1}^{N_{sb}} \exp\left(-\frac{(y_R - \bar{y}_{R,q}^{(-)})^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}\right) \quad (52)$$

where  $\bar{y}_{R,q}^{(-)} \in \mathcal{Y}_R^{(-)}$ , and the conditional BER given  $b_d(k) = -1$  is

$$P_{E,-}(\mathbf{w}) = \int_0^\infty p(y_R|-) dy_R = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q(g_{q,-}(\mathbf{w})) \quad (53)$$

with

$$g_{q,-}(\mathbf{w}) = -\frac{\bar{y}_{R,q}^{(-)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} = \frac{\text{sgn}(b_{q,d}) \bar{y}_{R,q}^{(-)}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \\ = \frac{\text{sgn}(b_{q,d}) \Re[\mathbf{w}^H \bar{\mathbf{x}}_q^{(-)}]}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (54)$$

Because of the symmetric distribution of  $\mathcal{Y}_R$ ,  $P_{E,-}(\mathbf{w}) = P_{E,+}(\mathbf{w})$ . This proves that the BER of the linear filter with a weight vector  $\mathbf{w}$  is given by

$$P_E(\mathbf{w}) = \frac{1}{2} P_{E,+}(\mathbf{w}) + \frac{1}{2} P_{E,-}(\mathbf{w}) \\ = \frac{1}{N_{sb}} \sum_{q=1}^{N_{sb}} Q(g_{q,+}(\mathbf{w})) \quad (55)$$

It is also obvious that the BER can be expressed as

$$P_E(\mathbf{w}) = \frac{1}{N_b} \sum_{q=1}^{N_b} Q(g_q(\mathbf{w})) \quad (56)$$

where

$$g_q(\mathbf{w}) = \frac{\text{sgn}(b_{q,d}) \bar{y}_{R,q}}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} = \frac{\text{sgn}(b_{q,d}) \Re[\mathbf{w}^H \bar{\mathbf{x}}_q]}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (57)$$

and the calculation is over all the  $\bar{y}_{R,q} \in \mathcal{Y}_R$ .