Least Bit Error Rate Adaptive Multiuser Detection

Sheng Chen

Department of Electronics and Computer Science, the University of Southampton, Highfield, Southampton SO17 1BJ, U.K.

Abstract. Linear detector required at direct-sequence code division multiple access (DS-CDMA) communication systems is classically designed based on the minimum mean squares error (MMSE) criterion, which can efficiently be implemented using the standard adaptive algorithms, such as the least mean square (LMS) algorithm. As the probability distribution of the linear detector's soft output is generally non-Gaussian, the MMSE solution can be far away from the optimal minimum bit error rate (MBER) solution. Adopting a non-Gaussian approach naturally leads to the MBER linear detector. Based on the approach of Parzen window or kernel density estimation for approximating the probability density function (p.d.f.), a stochastic gradient adaptive MBER algorithm, called the least bit error rate (LBER), is developed for training a linear multiuser detector. A simplified or approximate LBER (ALBER) algorithm is particularly promising, as it has a computational complexity similar to that of the classical LMS algorithm. Furthermore, this ALBER algorithm can be extended to the nonlinear multiuser detection.

1 Introduction

DS-CDMA constitutes an attractive multiuser scheme that allows users to transmit at the same carrier frequency. However, this creates multiuser interference (MUI) which, if not controlled, can seriously degrade the quality of reception. Multiuser detection provides an effective means to combating MUI [1],[2]. In a DS-CDMA communication system, the objective of the receiver is to detect the transmitted information bits of one (at mobile-end) or many (at base station) users. The first case is concerned with communication from base station to mobile and is commonly called downlink. This chapter considers multiuser detection at mobile. In the downlink scenario, a mobile has a very limited computing power, and computational complexity is a critical issue. Typically, a linear detector is employed at the receiver to meet the strict real-time computational constraint.

A variety of linear multiuser detectors have been proposed [2]–[8]. A very simple linear detector is the matched filter (MF) with the detector weight vector being set to the user spreading code. However, multipath distortions, which are often encountered in DS-CDMA systems, can seriously degrade the performance of the MF. Another linear detector structure is called the zero-forcing (ZF) detector. As this ZF solution ignores the noise, it suffers from a serious noise enhancing problem. Moreover, other interfering user codes required in calculating the ZF solution may not be available to the detector, and it is not known how to adaptively implement the ZF solution. The most popular linear multiuser detector is the MMSE detector, as it usually provides good performance and can readily be implemented using standard adaptive filter techniques such as the LMS algorithm.

The ultimate performance criterion of a detector is its bit error rate (BER). Although for the linear detector a small mean square error (MSE) is associated with a small BER, the MMSE solution is generally not the MBER one. This is obvious, since in order for the MMSE solution to be the MBER solution the detector's soft output should be Gaussian distributed. The conditional p.d.f. of the linear detector output is however a sum of Gaussian distributions and therefore non-Gaussian. In the situation where only a few strong interfering users exist, the MMSE solution can be considerably inferior to the MBER one. A non-adaptive MBER linear multiuser detector is considered in [9] based on gradient optimization for narrow-band Gaussian CDMA channels which do not introduce intersymbol interference (ISI). There are a few adaptive MBER linear multiuser detectors in the literature [10]–[12].

The adaptive MBER algorithm given in [10] uses a difference approximation to estimate the gradient of one-sample error probability. Its main drawback is very slow convergence, particularly in the situation where the error rate is very low. Furthermore, the computational complexity of the algorithm is high and is in the order $O(M^2)$, M being the detector length. The adaptive MBER algorithm reported in [11], called the approximate MBER (AMBER) detector, is appealing due to its computational simplicity. It is a stochastic gradient algorithm that is identical to the signed-error LMS algorithm [13], except in the vicinity of the decision boundary where it is modified to continue updating the weights when the signed-error LMS algorithm would not. The AMBER algorithm therefore can continuously update when the detector weight vector has reached the regions of very low error rate.

This chapter considers the adaptive MBER algorithm based on a density approximation approach [12], called the LBER algorithm. Previous studies have shown that this LBER algorithm outperforms the AMBER algorithm in terms of convergence speed and steady-state BER misadjustment, in both the linear multiuser detection application [12] and the single-user equalization application [14],[15]. Although the computational requirement of this LBER linear detector is considerably higher than the AMBER linear detector, its complexity is still in the order $O(M)$. Furthermore, as will be shown in this chapter, a simplified LBER algorithm called the ALBER has a similar performance to the full LBER algorithm and yet has a complexity similar to the very simple LMS algorithm. An added advantage of this ALBER algorithm is that it can be extended to the nonlinear multiuser detector.

A basic assumption for a linear detector to work adequately is that the two classes of signal states related to the transmitted bit being $+1$ and -1, respectively, are linearly separable. Multipath distortions however may result in linearly inseparable situation. In the nonlinear separable case, a nonlinear multiuser detector is required to achieve good performance. Classically, training a nonlinear detector is based on the LMS algorithm. Previous work [16],[17] has shown that a nonlinear detector trained by the ALBER algorithm can closely match the theoretical optimal performance.

2 System Model

The discrete-time baseband model of the synchronous DS-CDMA downlink system supporting N users and transmitting M $(> N)$ chips per bit is depicted in Fig. 1, where $b_i(k) \in \{\pm 1\}$ denotes the k-th bit of user i, the unit-length signature sequence for user i is $\bar{\mathbf{s}}_i = [\bar{s}_{i,1} \ \cdots \ \bar{s}_{i,M}]^T$ and the transfer function associated with the channel impulse response (CIR) is

$$H(z) = \sum_{i=0}^{n_h-1} h_i z^{-i}. \tag{1}$$

The bit vector of N users at bit instant k is $\mathbf{b}(k) = [b_1(k) \ \cdots \ b_N(k)]^T$, and the received signal vector obtained by sampling at chip rate is $\mathbf{r}(k) = [r_1(k) \ \cdots \ r_M(k)]^T$. It can be shown that the baseband model for $\mathbf{r}(k)$ is

$$\mathbf{r}(k) = \mathbf{P} \begin{bmatrix} \mathbf{b}(k) \\ \mathbf{b}(k-1) \\ \vdots \\ \mathbf{b}(k-L+1) \end{bmatrix} + \mathbf{n}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k) \tag{2}$$

where L is the channel ISI span, $\bar{\mathbf{r}}(k)$ denotes the noise-free received signal vector, the white Gaussian noise vector $\mathbf{n}(k) = [n_1(k) \ \cdots \ n_M(k)]^T$ with $E[\mathbf{n}(k)\mathbf{n}^T(k)] = \sigma_n^2 \mathbf{I}$, and the $M \times LN$ system matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \mathbf{H} \begin{bmatrix} \bar{\mathbf{S}}\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} \end{bmatrix} \tag{3}$$

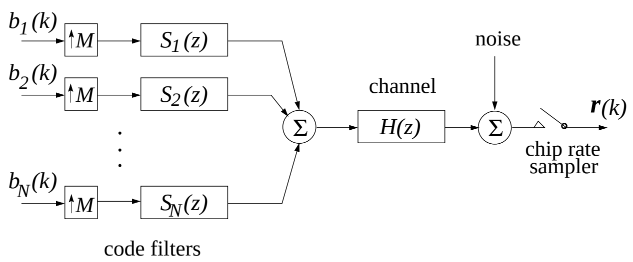

Fig. 1. Discrete-time model of synchronous CDMA downlink.

with the user signature matrix $\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1 \ \cdots \ \bar{\mathbf{s}}_N]$, the diagonal user signal amplitude matrix $\mathbf{A} = \mathrm{diag}\{A_1 \ \cdots \ A_N\}$ and the $M \times LM$ CIR matrix

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & h_0 & h_1 & \cdots & h_{n_h-1} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_{n_h-1} & 0 & \cdots & 0 \end{bmatrix}. \tag{4}$$

The channel ISI span L depends on the length of the CIR, n_h, related to the length of the chip sequence, M. For $n_h = 1$, $L = 1$; for $1 < n_h \le M$, $L = 2$; for $M < n_h \le 2M$, $L = 3$; and so on.

Let the $N_b = 2^{LN}$ possible combinations of $[\mathbf{b}^T(k) \ \cdots \ \mathbf{b}^T(k-L+1)]^T$ be

$$\mathbf{b}^{(j)} = \begin{bmatrix} \mathbf{b}^{(j)}(k) \\ \mathbf{b}^{(j)}(k-1) \\ \vdots \\ \mathbf{b}^{(j)}(k-L+1) \end{bmatrix}, \ 1 \le j \le N_b, \tag{5}$$

and $b_i^{(j)}$ the i-th element of $\mathbf{b}^{(j)}(k)$. Clearly,

$$\bar{\mathbf{r}}(k) \in \mathcal{R} \triangleq \{\bar{\mathbf{r}}_j = \mathbf{P}\mathbf{b}^{(j)}, \ 1 \le j \le N_b\}. \tag{6}$$

$\mathcal{R}$ is called the set of noise-free received signal states. For user i, it can be divided into two subsets depending on the value of $b_i(k)$

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

Consider the linear detector for user i which takes the form

$$y(k) = \mathbf{w}^T \mathbf{r}(k) = \mathbf{w}^T(\bar{\mathbf{r}}(k) + \mathbf{n}(k)) = \bar{y}(k) + e(k) \tag{8}$$

where $\mathbf{w} = [w_1 \ \cdots \ w_M]^T$ is the detector weight vector and $e(k)$ is Gaussian with zero mean and variance $\mathbf{w}^T\mathbf{w}\sigma_n^2$. The estimated $b_i(k)$ is given by

$$\hat{b}_i(k) = \mathrm{sgn}(y(k)) = \begin{cases} +1, & y(k) \ge 0, \\ -1, & y(k) < 0. \end{cases} \tag{9}$$

Obviously, the scalar $\bar{y}(k)$ can only take values from the set

$$\mathcal{Y} \triangleq \{\bar{y}_j = \mathbf{w}^T \bar{\mathbf{r}}_j, \ 1 \le j \le N_b\} \tag{10}$$

which can be divided into two subsets depending on the value of $b_i(k)$

$$\mathcal{Y}^{(\pm)} \triangleq \{\bar{y}_j^{(\pm)} \in \mathcal{Y} : \ b_i(k) = \pm 1\}. \tag{11}$$

For the linear detector (8) to perform adequately, $\mathcal{Y}^{(+)}$ and $\mathcal{Y}^{(-)}$ must be linearly separable. Otherwise, a nonlinear detector should be used. The simple

MF detector is given by $\mathbf{w}_{\rm MF} = \bar{\mathbf{s}}_i$. The most popular solution for the linear detector (8) is the MMSE one given by

$$\mathbf{w}_{\rm MMSE} = \left(\sigma_n^2 \mathbf{I} + \mathbf{P}\mathbf{P}^T\right)^{-1} \mathbf{p}_i \tag{12}$$

with $\mathbf{p}_i$ denoting the i-th column of $\mathbf{P}$. Although user i may not know the other user codes and therefore may be unable to compute $\mathbf{w}_{\rm MMSE}$ directly, adaptive implementation using the LMS does not require to know the other user codes. The ZF solution $\mathbf{w}_{\rm ZF}$ is obtained by setting $\sigma_n^2 = 0$ in (12).

3 The MBER Linear Multiuser Detector

The conditional p.d.f. of $y(k)$ given $b_i(k) = +1$ is

$$p_y(y| + 1) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \sum_{j=1}^{N_{sb}} \exp\left(- \frac{\left(y - \bar{y}_j^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^T\mathbf{w}} \right) \tag{13}$$

where $N_{sb} = N_b/2$ is the number of points in $\mathcal{Y}^{(+)}$ and $\bar{y}_j^{(+)} \in \mathcal{Y}^{(+)}$. Thus, the conditional BER of the linear detector given $b_i(k) = +1$ is

$$P_{E,+}(\mathbf{w}) = \int_{-\infty}^{0} p_y(y| + 1)\, dy = \frac{1}{N_{sb}} \sum_{j=1}^{N_{sb}} Q\left(g_{j,+}(\mathbf{w})\right) \tag{14}$$

where

$$g_{j,+}(\mathbf{w}) = \frac{\bar{y}_j^{(+)}}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} = \frac{\mathbf{w}^T\bar{\mathbf{r}}_j^{(+)}}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} = \frac{\mathrm{sgn}(b_i^{(j)})\bar{y}_j^{(+)}}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \tag{15}$$

and

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{u^2}{2}\right) du. \tag{16}$$

Similarly, the conditional p.d.f. of $y(k)$ given $b_i(k) = -1$ is

$$p_y(y| - 1) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \sum_{j=1}^{N_{sb}} \exp\left(- \frac{\left(y - \bar{y}_j^{(-)}\right)^2}{2\sigma_n^2 \mathbf{w}^T\mathbf{w}} \right) \tag{17}$$

where $\bar{y}_j^{(-)} \in \mathcal{Y}^{(-)}$, and the conditional BER given $b_i(k) = -1$ is

$$P_{E,-}(\mathbf{w}) = \int_{0}^{\infty} p_y(y| - 1)\, dy = \frac{1}{N_{sb}} \sum_{j=1}^{N_{sb}} Q\left(g_{j,-}(\mathbf{w})\right) \tag{18}$$

where

$$g_{j,-}(\mathbf{w}) = -\frac{\bar{y}_j^{(-)}}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} = \frac{\mathrm{sgn}(b_i^{(j)})\bar{y}_j^{(-)}}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}. \tag{19}$$

Because of the symmetric distribution of $\mathcal{Y}$, $P_{E,-}(\mathbf{w}) = P_{E,+}(\mathbf{w})$. Thus, the BER of the linear detector with the weight vector $\mathbf{w}$ is

$$P_E(\mathbf{w}) = P_{E,+}(\mathbf{w}) = \frac{1}{N_{sb}} \sum_{j=1}^{N_{sb}} Q\left(g_{j,+}(\mathbf{w})\right). \tag{20}$$

It is seen that the BER can be evaluated based on a single subset $\mathcal{Y}^{(+)}$ (or $\mathcal{Y}^{(-)}$). Also the BER is invariant to a positive scaling of the weight vector, that is, the BER depends on the direction of $\mathbf{w}$ only.

The MBER solution is defined as

$$\mathbf{w}_{\mathrm{MBER}} = \arg\min_{\mathbf{w}} P_E(\mathbf{w}). \tag{21}$$

The gradient of $P_E(\mathbf{w})$ with respect to $\mathbf{w}$ is

$$\nabla P_E(\mathbf{w}) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n} \sum_{j=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_j^{(+)}\right)^2}{2\sigma_n^2 \mathbf{w}^T\mathbf{w}} \right) \times$$
$$\mathrm{sgn}(b_i^{(j)}) \left(\frac{\bar{y}_j^{(+)}\mathbf{w}}{\left(\mathbf{w}^T\mathbf{w}\right)^{3/2}} - \frac{\bar{\mathbf{r}}_j^{(+)}}{\sqrt{\mathbf{w}^T\mathbf{w}}} \right). \tag{22}$$

With the gradient, the optimization problem (21) can be solved for iteratively using a conjugated gradient algorithm [18],[12] with a resetting of the search direction to the negative gradient $-\nabla P_E(\mathbf{w})$ every J iterations. It is computationally advantageous to normalize $\mathbf{w}$ to a unit-length after every iteration, so that the gradient can be simplified as

$$\nabla P_E(\mathbf{w}) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n} \sum_{j=1}^{N_{sb}} \exp\left(-\frac{\left(\bar{y}_j^{(+)}\right)^2}{2\sigma_n^2} \right) \mathrm{sgn}(b_i^{(j)}) \left(\bar{y}_j^{(+)}\mathbf{w} - \bar{\mathbf{r}}_j^{(+)} \right). \tag{23}$$

It is obvious that, if $\mathcal{Y}^{(-)}$ is used for the BER evaluation, one only needs to substitute $\bar{y}_j^{(+)}$ and $\bar{\mathbf{r}}_j^{(+)}$ by $\bar{y}_j^{(-)}$ and $\bar{\mathbf{r}}_j^{(-)}$ in the gradient formula.

Unlike the MMSE solution (12), there exists no close-form solution for $\mathbf{w}_{\mathrm{MBER}}$. For $\mathbf{w}_{\mathrm{MMSE}}$ to achieve the MBER, the p.d.f. (13) or (17) must be Gaussian. If the number of users is large, this p.d.f. will be close to Gaussian, and the BER difference between $\mathbf{w}_{\mathrm{MMSE}}$ and $\mathbf{w}_{\mathrm{MBER}}$ is expected to be small, if any gap exists. If however, there exist only a few interferers, it is likely that the BER difference between $\mathbf{w}_{\mathrm{MMSE}}$ and $\mathbf{w}_{\mathrm{MBER}}$ will be large. Whether the MMSE solution can achieve a BER close to the MBER also depends on the CIR. Before turning to the adaptive MBER algorithm, consider the following simple example, which provides some insights to the MBER solutions.

Example 1. This is a simplest system with two equal-power users and two chips per bit. The two chip codes are $(+1,+1)$ and $(+1,-1)$, respectively, and

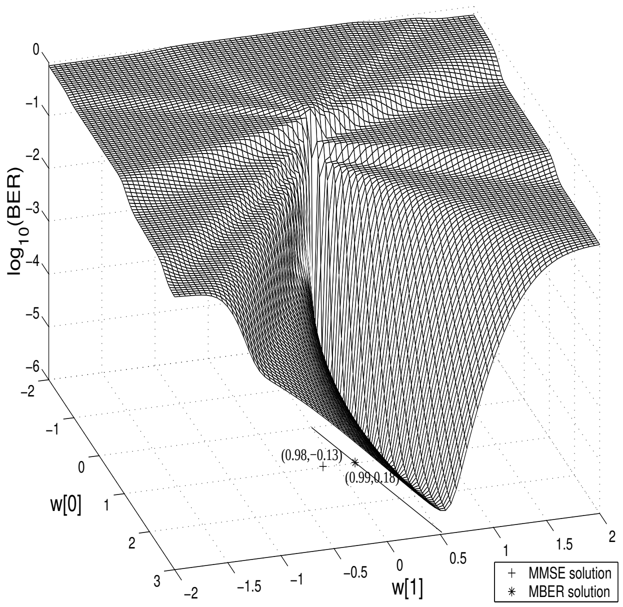

Fig. 2. Bit error rate surface of user-one detector for a simple two-user system with two chips per bit and SNR_1=25 dB given in Example 1.

the transfer function of the CIR at chip rate is $H(z) = 1 + 0.8z^{-1} + 0.6z^{-2}$. The signal to noise ratio for user 1 is $\mathrm{SNR}_1 = 25$ dB. The $\log_{10}(\mathrm{BER})$ surface for user 1 is plotted in Fig. 2. Some general observations can be drawn from Fig. 2. The MBER solutions form a half line in the weight space, with one end of the line approaching the original and the other end approaching infinity. Any point in this half line is a global MBER solution. The point marked in the MBER solution half line is the unit-length one. The origin of the weight space is the singular (discontinuity) point of the BER surface. The MMSE solution for this example is also depicted in Fig. 2. For the MMSE solution, $\log_{10}(\mathrm{BER}) = -3.88$, while for a MBER one, $\log_{10}(\mathrm{BER}) = -5.56$.

4 Adaptive MBER Algorithms

The p.d.f. of $y(k)$ is explicitly given by

$$p_y(y) = \frac{1}{N_b\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\sum_{j=1}^{N_b}\exp\left(-\frac{(y-\bar{y}_j)^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{24}$$

and the BER can alternatively be expressed as

$$P_E(\mathbf{w}) = \frac{1}{N_b}\sum_{j=1}^{N_b} Q\left(g_j(\mathbf{w})\right) \tag{25}$$

where $\bar{y}_j \in \mathcal{Y}$ and

$$g_j(\mathbf{w}) = \frac{\mathrm{sgn}(b_i^{(j)})\bar{y}_j}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}. \tag{26}$$

In reality, the p.d.f. of $y(k)$ is unknown. A widely used approach to approximate a p.d.f. is known as the kernel density estimate [19]–[21].

Given a block of K training samples $\{\mathbf{r}(k), b_i(k)\}$, a kernel density or Parzen window estimate of the p.d.f. (24) is given by:

$$\hat{p}_y(y) = \frac{1}{K\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\sum_{k=1}^{K}\exp\left(-\frac{(y-y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{27}$$

where the kernel width ρ_n is related to the noise standard deviation σ_n [20]. From this estimated p.d.f., the estimated BER is given by

$$\hat{P}_E(\mathbf{w}) = \frac{1}{K}\sum_{k=1}^{K} Q\left(\hat{g}_k(\mathbf{w})\right) \tag{28}$$

with

$$\hat{g}_k(\mathbf{w}) = \frac{\mathrm{sgn}(b_i(k))y(k)}{\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}. \tag{29}$$

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

$$\nabla\hat{P}_E(\mathbf{w}) = \frac{1}{K\sqrt{2\pi}\rho_n}\sum_{k=1}^{K}\exp\left(-\frac{y^2(k)}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right)\times$$
$$\mathrm{sgn}(b_i(k))\left(\frac{y(k)\mathbf{w}}{(\mathbf{w}^T\mathbf{w})^{3/2}} - \frac{\mathbf{r}(k)}{\sqrt{\mathbf{w}^T\mathbf{w}}}\right). \tag{30}$$

By substituting $\nabla P_E(\mathbf{w})$ with $\nabla\hat{P}_E(\mathbf{w})$ in the conjugate gradient updating mechanism, a block-data adaptive algorithm can readily be obtained [12].

4.1 Least Bit Error Rate Algorithm

To derive a sample-by-sample adaptive algorithm, consider a single-sample estimate of $p_y(y)$

$$\hat{p}_y(y,k) = \frac{1}{\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\exp\left(-\frac{(y-y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right). \tag{31}$$

Using the instantaneous stochastic gradient

$$\nabla\hat{P}_E(\mathbf{w},k) = \frac{\mathrm{sgn}(b_i(k))}{\sqrt{2\pi}\rho_n}\exp\left(-\frac{y^2(k)}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right)\left(\frac{y(k)\mathbf{w}}{(\mathbf{w}^T\mathbf{w})^{3/2}} - \frac{\mathbf{r}(k)}{\sqrt{\mathbf{w}^T\mathbf{w}}}\right), \tag{32}$$

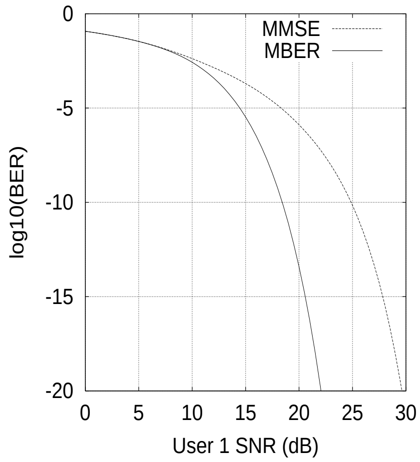

Fig. 3. Linear detector bit error rates for user 1 of Example 2. SNR_i, $1 \le i \le 4$, are identical.

with a re-scaling to ensure $\mathbf{w}^T\mathbf{w} = 1$, gives rise to the LBER algorithm:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\mathrm{sgn}(b_i(k))}{\sqrt{2\pi}\rho_n} \exp\left(-\frac{y^2(k)}{2\rho_n^2}\right) \left(\mathbf{r}(k) - y(k)\mathbf{w}(k)\right), \quad (33)$$

$$\mathbf{w}(k+1) = \frac{\mathbf{w}(k+1)}{\sqrt{\mathbf{w}^T(k+1)\mathbf{w}(k+1)}}, \quad (34)$$

where the adaptive gain μ and the kernel width ρ_n are the two algorithm parameters that need to be set appropriately.

4.2 Approximate Least Bit Error Rate Algorithm

In the kernel density estimate (27), a variable width $\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}$ is used. This is because the true standard deviation of $y(k)$ is $\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}$, which depends on the detector weight vector. If an approximation is made by using a constant width ρ_n in a kernel density estimate, computational complexity can be reduced considerably. Formally, this is to use the kernel density estimate

$$\tilde{p}_y(y) = \frac{1}{K\sqrt{2\pi}\rho_n} \sum_{k=1}^{K} \exp\left(-\frac{(y - y(k))^2}{2\rho_n^2}\right) \quad (35)$$

as an approximation of the true density (24), and to use

$$\tilde{P}_E(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^{K} Q\left(\tilde{g}_k(\mathbf{w})\right) \quad (36)$$

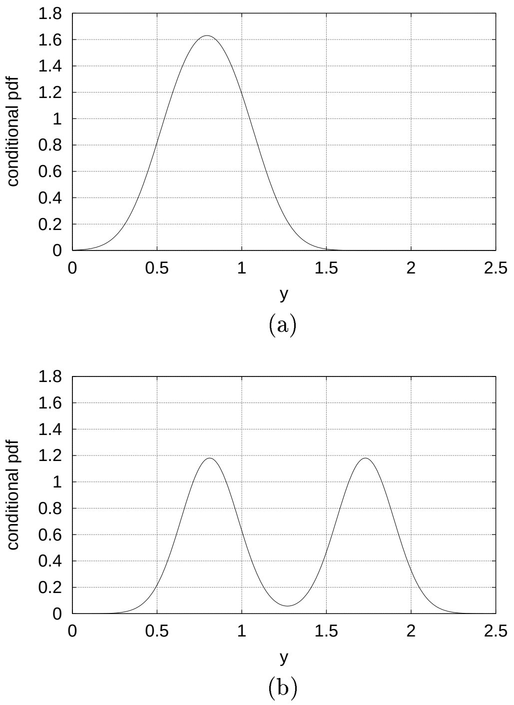

Fig. 4. Conditional probability density function $p_y(y|+1)$: (a) the MMSE detector and (b) the MBER detector, for user 1 of Example 2. $\text{SNR}_i = 16$ dB, $1 \leq i \leq 4$, and the weight vector is normalized to a unit-length.

with

$$\tilde{g}_k(\mathbf{w}) = \frac{\text{sgn}(b_i(k))y(k)}{\rho_n} \tag{37}$$

as a BER estimate. The gradient of $\tilde{P}_E(\mathbf{w})$ has a much simpler form.

Adopting this approach, an approximate LBER algorithm is obtained:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\text{sgn}(b_i(k))}{\sqrt{2\pi}\rho_n} \exp\left(-\frac{y^2(k)}{2\rho_n^2}\right) \mathbf{r}(k). \tag{38}$$

For this ALBER algorithm, there is no need to normalize $\mathbf{w}$ after each update, and the algorithm has a similar complexity to the LMS algorithm.

5 Simulation Study

The adaptive MBER algorithms discussed in the previous section are investigated using computer simulation.

Example 2. This is a 4-user system with 8 chips per bit. The four user code sequences are $(+1,+1,+1,+1,-1,-1,-1,-1)$, $(+1,-1,+1,-1,-1,+1,-1,$

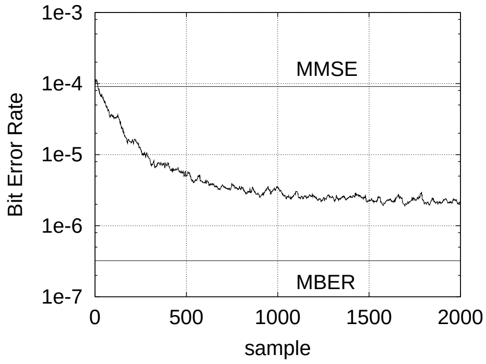

Fig. 5. Learning curves (solid: training and dashed: decision directed adaptation) of the LBER algorithm for user 1 of Example 2. $\text{SNR}_i = 16$ dB, $1 \le i \le 4$, $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$, adaptive gain $\mu = 0.05$ and width $\rho_n^2 = 4\sigma_n^2$. Two curves are indistinguishable.

$+1)$, $(+1,+1,-1,-1,-1,-1,+1,+1)$ and $(+1,-1,-1,+1,-1,+1,+1,-1)$, respectively. The four users have equal signal power and the transfer function of the CIR at chip rate is

$$H(z) = 0.4 + 0.7z^{-1} + 0.4z^{-2}. \tag{39}$$

The linear detector for user 1 is considered. Fig. 3 compares the BER performance of the MMSE detector with that of the MBER detector. The MBER solution is obtained using the conjugate gradient algorithm with a periodic resetting of search direction. It can be seen that for this case the BERs of the two detectors are considerably different. Given the user 1 SNR to be $\text{SNR}_1 = 16$ dB, Fig. 4 depicts the conditional p.d.f.s, $p_y(y| + 1)$, for the

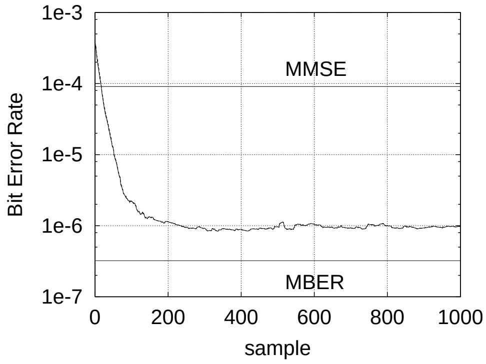

Fig. 6. Learning curves (solid: training and dashed: decision directed adaptation) of the LBER algorithm for user 1 of Example 2. $\text{SNR}_i = 16$ dB, $1 \le i \le 4$, $\mathbf{w}(0) = [0.01\ 0.01\ 0.01\ -0.01\ -0.01\ -0.01\ -0.01\ 0.01]^T$, adaptive gain $\mu = 0.05$ and width $\rho_n^2 = 4\sigma_n^2$. Two curves are indistinguishable.

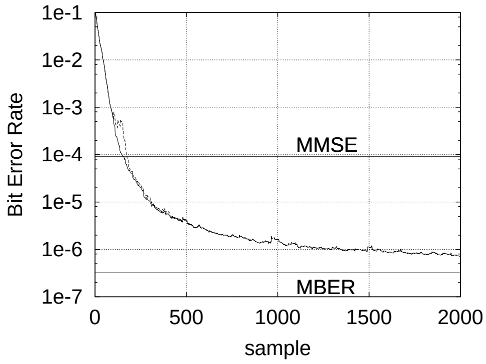

Fig. 7. Learning curves (solid: training and dashed: first 80-sample training followed by decision directed adaptation) of the LBER algorithm for user 1 of Example 2. $\mathrm{SNR}_i = 16$ dB, $1 \le i \le 4$, $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MF}}$, adaptive gain $\mu = 0.05$ and width $\rho_n^2 = 2\sigma_n^2$.

MMSE and MBER detectors. It can be see that the conditional p.d.f. of the MMSE detector resembles a Gaussian. This is not surprising. The conditional p.d.f. $p_y(y|+1)$ is a sum of Gaussian distributions. The MMSE solution can be viewed to set the parameter vector $\mathbf{w}$ of $p_y(y|+1)$ in such a way that this non-Gaussian distribution looks like a Gaussian distribution. The non-Gaussian nature of $p_y(y|+1)$ is clearly demonstrated in the case of the MBER solution. The two adaptive MBER algorithms, the LBER and ALBER, are next studied. In the investigation, all the results are averaged over 100 runs.

Fig. 5 shows the learning curves of the LBER algorithm, given $\mathrm{SNR}_1 = 16$ dB and $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, where the adaptive gain $\mu = 0.05$ and the width parameter $\rho_n^2 = 4\sigma_n^2 \approx 0.1$. There are in fact two indistinguishable learning curves in Fig. 5, the solid one indicates the training performance, and the dashed one is the decision-directed performance with $b_1(k)$ being substituted by the detector's decisions $\hat{b}_1(k)$. It is well known that the BER surface is highly complicated and may contain local minima. Our experience has suggested that initializing $\mathbf{w}(0)$ to the MMSE solution is generally a bad choice for the LBER algorithm, as the convergence speed is generally slow and the steady-state BER misadjustment is relatively large for this initial condition. The training and decision-directed learning curves for the LBER algorithm given $\mathbf{w}(0) = [0.01\ 0.01\ 0.01\ -0.01\ -0.01\ -0.01\ -0.01\ 0.01]^T$ are shown in Fig. 6. It can be seen that for this choice of $\mathbf{w}(0)$ the algorithm has a much faster convergence rate and a smaller steady-state BER misadjustment. Unlike the MMSE solution, the MF is known to a detector and can conveniently be used as $\mathbf{w}(0)$. With $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MF}}$, the two learning curves of the LBER algorithm are illustrated in Fig. 7. It can be seen that the algorithm performs better with the MF solution as $\mathbf{w}(0)$ than with the MMSE solution as $\mathbf{w}(0)$.

The learning curves of the ALBER algorithm with $\mathbf{w}(0)$ set to the MMSE solution and MF one are depicted in Figs. 8 and 9, respectively. Comparing

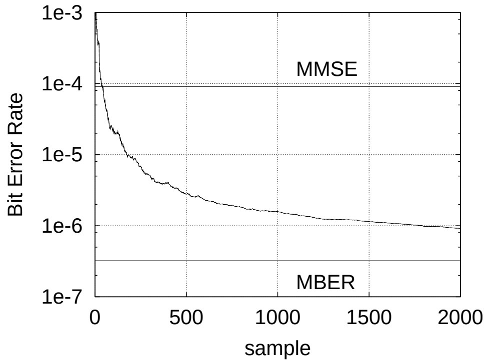

Fig. 8. Learning curves (solid: training and dashed: decision directed adaptation) of the ALBER algorithm for user 1 of Example 2. $\mathrm{SNR}_i = 16$ dB, $1 \le i \le 4$, $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, adaptive gain $\mu = 0.2$ and width $\rho_n^2 = 25\sigma_n^2$. Two curves are indistinguishable.

Figs. 8 and 9 with Figs. 5 and 7, it can be seen that this approximate LBER algorithm does not seem to result in performance degradation. In fact, the ALBER algorithm appears to be less sensitive to the choice of initial condition, and the algorithm with $\mathbf{w}(0)$ set to the MMSE solution performs better than the LBER algorithm with the same initial condition. The ALBER algorithm has an added advantage of simpler computational requirements.

Example 3. This example investigate the near-far effect to the adaptive MBER algorithm. The system has two users with the two user chip codes given by $(+1, +1, -1, -1)$ and $(+1, -1, -1, +1)$, respectively. The transfer

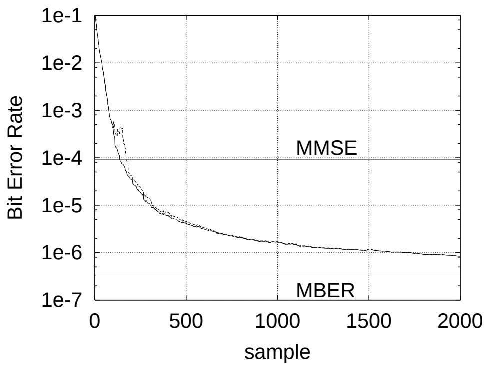

Fig. 9. Learning curves (solid: training and dashed: first 80-sample training followed by decision directed adaptation) of the ALBER algorithm for user 1 of Example 2. $\mathrm{SNR}_i = 16$ dB, $1 \le i \le 4$, $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MF}}$, adaptive gain $\mu = 0.5$ and width $\rho_n^2 = 25\sigma_n^2$.

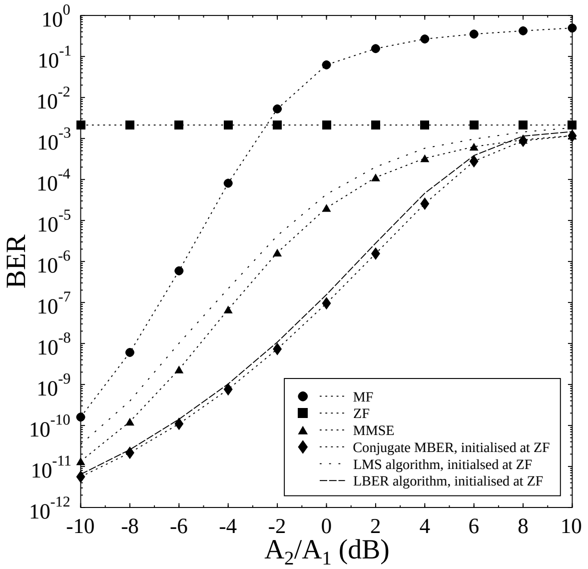

Fig. 10. Comparison of bit error rates of various linear detectors for user 1 of Example 3. $\mathrm{SNR}_1 = 16$ dB with varying SIR.

function of the CIR at chip rate is given by

$$H(z) = 1.0 + 0.25z^{-1} + 0.5z^{-3}. \tag{40}$$

The linear detector for user 1 is considered. With a fixed user 1 signal power A_1 and a fixed $\mathrm{SNR}_1 = 16$ dB, the interfering user 2 power A_2 is varied to provide different desired signal to interference ratios (SIR). Fig. 10 summarizes the performance of various detectors.

Example 4. The settings of this example are identical to Example 2, except that the three-path channel

$$H(z) = h_0 + h_1z^{-1} + h_2z^{-2} \tag{41}$$

is a Rayleigh fading one with the normalized Doppler frequency $f_d = 7.69 \times 10^{-7}$. Transmission is organized into frames, and a frame consists of 37 training bits and 112 data bits. Decision-directed adaptation is employed during data transmission. The channel is assumed to be frame faded, that is, the channel is kept constant within a frame. A typical set of the channel taps is shown in Fig. 11. Linear detector for user 1 is considered. Fig. 12 depicts the performance of the two adaptive algorithms, the LBER and LMS, in comparison with the benchmark of the theoretical MMSE performance.

6 Extension to the nonlinear multiuser detector

If we are not restricting to the class of linear detector, then the optimal detector for the system model described in Section 2 can easily be shown to

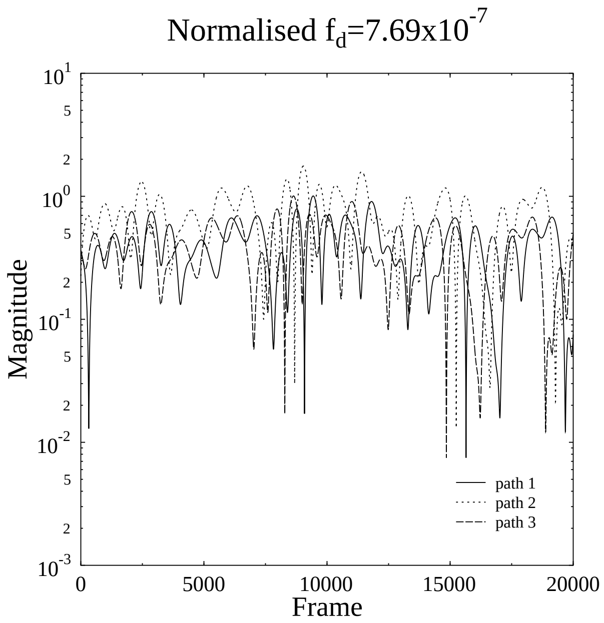

Fig. 11. A typical set of channel fading paths for Example 4.

be the following Bayesian detector [22]

$$y_B(k) = f_B(\mathbf{r}(k)) = \sum_{j=1}^{N_b} \beta_j b_i^{(j)} \exp\left(-\frac{\|\mathbf{r}(k) - \bar{\mathbf{r}}_j\|^2}{2\sigma_n^2}\right) \tag{42}$$

where $\bar{\mathbf{r}}_j \in \mathcal{R}$ and β_j is a positive constant incorporating *a priori* probability of $\bar{\mathbf{r}}_j$. Since all the states in $\mathcal{R}$ are equiprobable, all the β_j have the same value. The hard decision is made by quantizing $y_B(k)$ according to the rule (9). Although this Bayesian detector provides the best performance, it is computationally very expensive. Furthermore, the set of the channel states $\bar{\mathbf{r}}_j$ are unknown and have to be learned by some means.

Consider the general nonlinear detector for user i, which has the form

$$y(k) = f(\mathbf{r}(k); \mathbf{w}) \tag{43}$$

where the detector map f is realized for example by a neural network and the vector $\mathbf{w}$ consists of all the adjustable parameters of the detector. Classically, adaptive training of such a nonlinear structure is based on the MMSE principle and typically implemented using the LMS algorithm

$$\left.\begin{array}{l} y(k) = f(\mathbf{r}(k); \mathbf{w}(k-1)) \\ \mathbf{w}(k) = \mathbf{w}(k-1) + \mu(b_i(k) - y(k)) \frac{\partial f(\mathbf{r}(k); \mathbf{w}(k-1))}{\partial \mathbf{w}} \end{array}\right\} \tag{44}$$

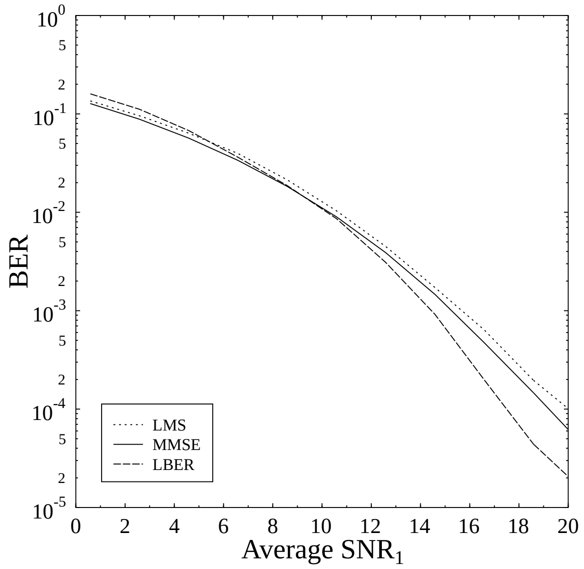

Fig. 12. Performance comparison of three user-1 detectors for Example 4. For the two adaptive algorithms, $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MF}}$.

However, the true performance criterion should be the BER, and we consider how to construct an adaptive MBER algorithm.

By linearizing the detector (43) around $\bar{\mathbf{r}}(k)$, it can be approximated as

$$y(k) \approx \bar{y}(k) + e(k) \tag{45}$$

where

$$\bar{y}(k) = f(\bar{\mathbf{r}}(k); \mathbf{w}) \tag{46}$$

can only take the value from the set

$$\mathcal{Y} \stackrel{\triangle}{=} \{\bar{y}_j = f(\bar{\mathbf{r}}_j; \mathbf{w}),\ 1 \le j \le N_b\} \tag{47}$$

and

$$e(k) = \left[\frac{\partial f(\bar{\mathbf{r}}(k); \mathbf{w})}{\partial \mathbf{r}}\right]^T \mathbf{n}(k) \tag{48}$$

is Gaussian with zero mean and variance

$$\rho_n^2(\mathbf{w}) = \frac{\sigma_n^2}{N_b} \sum_{j=1}^{N_b} \left[\frac{\partial f(\bar{\mathbf{r}}_j; \mathbf{w})}{\partial \mathbf{r}}\right]^T \frac{\partial f(\bar{\mathbf{r}}_j; \mathbf{w})}{\partial \mathbf{r}} \tag{49}$$

The p.d.f. of $y(k)$ can then be approximated by

$$p_y(y) \approx \frac{1}{N_b\sqrt{2\pi\rho_n^2(\mathbf{w})}} \sum_{j=1}^{N_b} \exp\left(-\frac{(y-\bar{y}_j)^2}{2\rho_n^2(\mathbf{w})}\right) \tag{50}$$

and the BER of the detector is approximately

$$P_E(\mathbf{w}) \approx \frac{1}{N_b} \sum_{j=1}^{N_b} Q(g_j(\mathbf{w})) \tag{51}$$

where

$$g_j(\mathbf{w}) = \frac{\mathrm{sgn}(b_i^{(j)})\bar{y}_j}{\rho_n(\mathbf{w})} = \frac{\mathrm{sgn}(b_i^{(j)}) f(\bar{\mathbf{r}}_j; \mathbf{w})}{\rho_n(\mathbf{w})}. \tag{52}$$

Using the kernel density estimate (35) with a constant ρ_n^2 to approximate the density (50) naturally leads to the ALBER algorithm

$$\left.\begin{array}{l} y(k) = f(\mathbf{r}(k); \mathbf{w}(k-1)) \\ \mathbf{w}(k) = \mathbf{w}(k-1) + \mu \frac{\mathrm{sgn}(b_i(k))}{\sqrt{2\pi}\rho_n} \exp\left(-\frac{y^2(k)}{2\rho_n^2}\right) \frac{\partial f(\mathbf{r}(k); \mathbf{w}(k-1))}{\partial \mathbf{w}} \end{array}\right\} \tag{53}$$

for training the nonlinear detector (43). The derivative $\frac{\partial f}{\partial \mathbf{w}}$ depends on the particular detector map employed. For example, consider the radial basis function (RBF) detector of the form

$$y(k) = f_{RBF}(\mathbf{r}(k); \mathbf{w}) = \sum_{j=1}^{n_c} \alpha_j \exp\left(-\frac{\|\mathbf{r}(k) - \mathbf{c}_j\|^2}{\tilde{\sigma}_j}\right). \tag{54}$$

The parameter vector $\mathbf{w}$ contains all the RBF weights α_j, widths $\tilde{\sigma}_j$ and centers $\mathbf{c}_j$. The dimension of $\mathbf{w}$ is therefore $n_c \times (M+2)$. The derivatives $\frac{\partial f_{RBF}}{\partial \mathbf{w}}$ are given in the forms

$$\left.\begin{array}{l} \frac{\partial f_{RBF}}{\partial \alpha_j} = \exp\left(-\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\tilde{\sigma}_j}\right) \\ \frac{\partial f_{RBF}}{\partial \tilde{\sigma}_j} = \alpha_j \exp\left(-\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\tilde{\sigma}_j}\right) \frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\tilde{\sigma}_j^2} \\ \frac{\partial f_{RBF}}{\partial \mathbf{c}_j} = 2\alpha_j \exp\left(-\frac{\|\mathbf{r}(k)-\mathbf{c}_j\|^2}{\tilde{\sigma}_j}\right) \frac{\mathbf{r}(k)-\mathbf{c}_j}{\tilde{\sigma}_j} \end{array}\right\} \quad 1 \le j \le n_c. \tag{55}$$

Example 5. The settings of this example are identical to Example 2, except that this time the detector for user 2 is considered. For user 2, $\mathcal{R}^{(+)}$ and $\mathcal{R}^{(-)}$ are almost linearly inseparable, and a linear detector has poor BER performance, as is shown in Fig. 13. For this example, the linear MMSE solution and the linear MBER one produce the same BER. The BERs of the optimal Bayesian detector is also shown in Fig. 13. Note that in this example the number of channel states $N_b = 256$, and the Bayesian detector is highly complex. The performance of the 64-center RBF detector trained by the ALBER algorithm (53) is depicted in Fig. 13. It can be seen that the performance of this ALBER RBF detector is very close to the full optimal Bayesian performance. Interestingly, in the simulation it is observed that the same 64-center RBF detector under the identical conditions but trained by the LMS algorithm (44), although converged well in the MSE, often results in a BER near 0.5. This confirms with the results given in [16],[17].

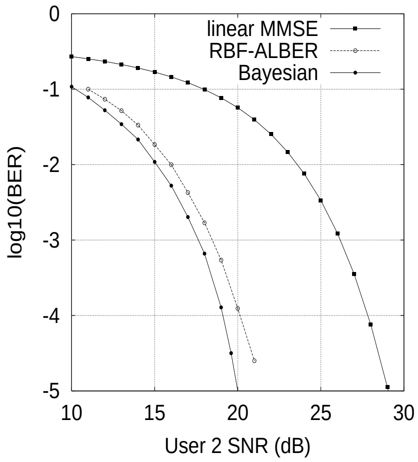

Fig. 13. Linear and nonlinear detector bit error rates for user 2 of Example 5. SNR_i, $1 \le i \le 4$, are identical. The RBF detector has 64 centers.

7 Conclusions

Adaptive multiuser detection has been considered based on the principle of minimizing the BER. It has clearly been demonstrated that, even for the linear detector, the MBER solution can be considerably better than the classical MMSE solution at least for certain situations. A fully adaptive MBER approach has been developed for training the linear detector. In particular, the ALBER algorithm has a computational complexity similar to that of the very simple LMS algorithm. Furthermore, it has been shown how to extend the adaptive MBER approach to the nonlinear multiuser detector.

Acknowledgement

Figs. 2, 10, 11 and 12 were provided by Mr. A.K. Samingan.

References

1. Prasad, R. (1996) *CDMA for Wireless Personal Communications.* Artech House, Inc.
2. Verdú, S. (1998) *Multiuser Detection.* Cambridge University Press
3. Xie, Z., Short, R.T., Rushforth, C.K. (1990) A family of suboptimum detectors for coherent multiuser communications. *IEEE J. Selected Areas in Communications*, **8**, 683–690
4. Madhow, U., Honig, M.L. (1994) MMSE interference suppression for direct-sequence spread-spectrum CDMA. *IEEE Trans. Communications*, **42**, 3178–3188

5. Miller, S.L. (1995) An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection. *IEEE Trans. Communications*, **43**, 1746–1755
6. Moshavi, S. (1996) Multi-user detection for DS-CDMA communications. *IEEE Communications Magazine*, **34**, 124–136
7. Poor, H.V., Verdú, S. (1997) Probability of error in MMSE multiuser detection. *IEEE Trans. Information Theory*, **43**, 858–871
8. Woodward, G., Vucetic, B.S. (1998) Adaptive detection for DS-CDMA. *Proc. IEEE*, **86**, 1413–1434
9. Mandayam, N.B., Aazhang, B. (1997) Gradient estimation for sensitivity analysis and adaptive multiuser interference rejection in code-division multi-access systems. *IEEE Trans. Communications*, **45**, 848–858
10. Psaromiligkos, I.N., Batalama, S.N., Pados, D.A. (1999) On adaptive minimum probability of error linear filter receivers for DS-CDMA channels. *IEEE Trans. Communications*, **47**, 1092–1102
11. Yeh, C.C., Lopes, R.R., Barry, J.R. (1998) Approximate minimum bit-error rate multiuser detection. In: *Proc. Globecom'98* (Sydney, Australia), Nov. 1998, 3590–3595
12. Chen, S., Samingan, A.K., Mulgrew, M., Hanzo, L. (2001) Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels. *IEEE Trans. Signal Processing*, **49**, 1240–1247
13. Sharma, R., Sethares, W.A., Bucklew, J.A. (1996) Asymptotic analysis of stochastic gradient-based adaptive filtering algorithms with general cost functions. *IEEE Trans. Signal Processing*, **44**, 2186–2194
14. Mulgrew, B., Chen, S. (2000) Stochastic gradient minimum-BER decision feedback equalisers. In: *Proc. IEEE Symp. Adaptive Systems for Signal Processing, Communication and Control* (Lake Louise, Alberta, Canada), Oct.1-4, 2000, 93–98
15. Mulgrew, B., Chen, S. (2001) Adaptive minimum-BER decision feedback equalisers for binary signalling. *Signal Processing*, **81**, 1479–1489
16. Chen, S., Samingan, A.K., Hanzo, L. (2000) Adaptive near minimum error rate training for neural networks with application to multiuser detection in CDMA communication systems. submitted to *IEEE Trans. Neural Networks*
17. Chen, S., Mulgrew, B., Hanzo, L. (2001) Adaptive least error rate algorithm for neural network classifier. In: *Proc. 2001 IEEE Workshop Neural Networks for Signal Processing* (Falmouth, MA, USA), Sept.10-12, 2001, 223–232
18. Bazaraa, M.S., Sherali, H.D., Shetty, C.M. (1993) *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley
19. Parzen, E. (1962) On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, **33**, 1066–1076
20. Silverman, B.W. (1996) *Density Estimation*. London: Chapman Hall
21. Bowman, A.W., Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press
22. Chen, S., Samingan, A.K., Hanzo, L. (2001) Support vector machine multiuser receiver for DS-CDMA signals in multipath channels. *IEEE Trans. Neural Networks*, **12**, 604-611