Adaptive Minimum Bit-Error-Rate Filtering

S. Chen

Department of Electronics and Computer Science
University of Southampton, Highfield
Southampton SO17 1BJ, U.K.
Tel./Fax: +44 (0)23 8059 6660/4508
E-mail: sqc@ecs.soton.ac.uk.

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 provides significant performance gain in terms of smaller bit error error (BER) over the MMSE approach. Adopting the classical Parzen window or kernel density estimation for a probability density function (p.d.f.), 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.

Keywords — Adaptive filtering, mean square error, probability density function, non-Gaussian distribution, Parzen window estimate, bit error rate, stochastic gradient algorithm.

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 p.d.f. 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 equalization application, adaptive MBER linear equalizer and decision feedback equalizer 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 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 p.d.f. 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 equalization and beamforming with 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

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

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 equalization, 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

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

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 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

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

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

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

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

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

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

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

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

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

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

$$\mathcal{Y}_R \triangleq \{\bar{y}_{R,q} = \Re[\bar{y}_q], \ \ 1 \le q \le N_b\} \tag{8}$$

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

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

The classical Wiener solution for the linear filter (1)

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

where $\mathbf{p}_d$ denotes the d-th column of $\mathbf{P}$, is generally not the optimal MBER solution. For $\mathbf{w}_{\mathrm{MMSE}}$ to be an MBER solution, the conditional p.d.f. 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 $\mathbf{w}_{\mathrm{ZF}}$ on the other hand does achieve a Gaussian conditional p.d.f., since the combined impulse response of the ZF filter and the system matrix $\mathbf{c}_{\mathrm{ZF}} = \mathbf{w}_{\mathrm{ZF}}^H \mathbf{P}$ has all zero elements except the d-th element c_d. That is, the ZF filter output is

$$y(k) = c_d b_d(k) + e(k). \tag{11}$$

However, the ZF filter suffers from a problem of serious noise enhancement 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 p.d.f. of $y_R(k)$ is a mixed sum of Gaussian distributions:

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

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

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

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

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

$\bar{y}_{R,q}^{(+)} \in \mathcal{Y}_R^{(+)}$ 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 $\mathcal{Y}_R^{(-)}$.

The MBER filtering solution is then defined as

$$\mathbf{w}_{\mathrm{MBER}} = \arg\min_{\mathbf{w}} P_E(\mathbf{w}). \tag{15}$$

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

$$\nabla P_E(\mathbf{w}) = \frac{1}{2N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^H\mathbf{w}}}\sum_{q=1}^{N_{sb}}\exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2\mathbf{w}^H\mathbf{w}}\right)\mathrm{sgn}(b_{q,d})\left(\frac{\bar{y}_{R,q}^{(+)}\mathbf{w}}{\mathbf{w}^H\mathbf{w}} - \bar{\mathbf{x}}_q^{(+)}\right). \tag{16}$$

With the gradient, the optimization problem (15) can be solved for iteratively using a gradient-based optimization algorithm. It is also computationally advantageous to normalize $\mathbf{w}$ to a unit-length after every iteration, so that the gradient can be simplified as

$$\nabla P_E(\mathbf{w}) = \frac{1}{2N_{sb}\sqrt{2\pi}\sigma_n}\sum_{q=1}^{N_{sb}}\exp\left(-\frac{\left(\bar{y}_{R,q}^{(+)}\right)^2}{2\sigma_n^2}\right)\mathrm{sgn}(b_{q,d})\left(\bar{y}_{R,q}^{(+)}\mathbf{w} - \bar{\mathbf{x}}_q^{(+)}\right). \tag{17}$$

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

Initialization. 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*.

$$\mathbf{w}(\iota+1) = \mathbf{w}(\iota) + \mu\mathbf{d}(\iota)$$

$$\mathbf{w}(\iota+1) = \frac{\mathbf{w}(\iota+1)}{\|\mathbf{w}(\iota+1)\|}$$

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

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

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

As the BER surface $P_E(\mathbf{w})$ is highly nonlinear, it is advisable to periodically re-set the search direction $\mathbf{d}$ to the negative gradient in the above conjugate gradient algorithm.

4 Adaptive Minimum Bit Error Rate Filtering

In reality, the p.d.f. of $y_R(k)$ is unknown,. The key to adaptive implementation of the MBER filtering solution is an effective estimate of the p.d.f. (12). Parzen window or kernel density estimate [20]–[22] is a best known method for estimating a probability distribution.

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 p.d.f. (12) is given by:

$$\hat{p}(y_R) = \frac{1}{K\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}} \sum_{k=1}^{K} \exp\left(-\frac{(y_R - y_R(k))^2}{2\rho_n^2 \mathbf{w}^H\mathbf{w}}\right) \tag{18}$$

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

$$\hat{P}_E(\mathbf{w}) = \frac{1}{K}\sum_{k=1}^{K} Q\left(\hat{g}_k(\mathbf{w})\right) \tag{19}$$

with

$$\hat{g}_k(\mathbf{w}) = \frac{\mathrm{sgn}(b_d(k)) y_R(k)}{\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}}. \tag{20}$$

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

$$\nabla\hat{P}_E(\mathbf{w}) = \frac{1}{2K\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}} \sum_{k=1}^{K} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2\mathbf{w}^H\mathbf{w}}\right) \mathrm{sgn}(b_d(k)) \left(\frac{y_R(k)\mathbf{w}}{\mathbf{w}^H\mathbf{w}} - \mathbf{x}(k)\right). \tag{21}$$

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 μ and the kernel width ρ_n are the two algorithm parameters.

4.2 Stochastic gradient adaptive MBER algorithm

An alternative Parzen window estimate to the true p.d.f. (12) is given by

$$\tilde{p}(y_R) = \frac{1}{K\sqrt{2\pi}\rho_n} \sum_{k=1}^{K} \exp\left(-\frac{(y_R - y_R(k))^2}{2\rho_n^2}\right) \tag{22}$$

and an approximation of the BER is

$$\tilde{P}_E(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^{K} Q\left(\tilde{g}_k(\mathbf{w})\right) \tag{23}$$

with

$$\tilde{g}_k(\mathbf{w}) = \frac{\mathrm{sgn}(b_d(k)) y_R(k)}{\rho_n}. \tag{24}$$

This approximation is valid provided that the width ρ_n is chosen appropriately. An advantage of using a constant width ρ_n, rather than a variable one $\rho_n\sqrt{\mathbf{w}^H\mathbf{w}}$, in the density estimate is that the gradient of $\tilde{P}_E(\mathbf{w})$ has a much simpler form, which leads to considerable reduction in computational complexity. Adopting this approach and considering sample-by-sample adaptation leads to the following stochastic gradient adaptive MBER algorithm:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \frac{\mathrm{sgn}(b_d(k))}{2\sqrt{2\pi}\rho_n} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2}\right) \mathbf{x}(k). \tag{25}$$

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

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\left(b_d(k) - y(k)\right)\mathbf{x}^*(k). \tag{26}$$

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 p.d.f. 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.

5 Application Examples

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

5.1 Single-user channel equalization

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

$$x(k) = \sum_{i=0}^{n_a-1} a_i b(k-i) + n(k) \tag{27}$$

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 equalizer at the receiver is a linear filter in the form of (1), where $\mathbf{x}(k) = [x(k)\ x(k-1) \cdots x(k-L+1)]^T$ with L known as the equalizer order. The $L \times M$ system matrix $\mathbf{P}$ in (2) has the form

$$\mathbf{P} = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & a_{n_a-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & a_0 & a_1 & \cdots & a_{n_a-1} \end{bmatrix} \tag{28}$$

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

In the simulation study, the following three-tap channel

$$\mathbf{a}^T = [-0.5 + j0.4\ \ 0.7 + j0.6\ \ 0.4 + j0.3] \tag{29}$$

and a five-tap ($L = 5$) equalizer were used. For this example, it was found that the optimal equalizer delay is $d = 3$. Fig. 1 compares the BER performance of the MMSE equalizer with that of the MBER one. Table 1 lists the MMSE and MBER solutions, $\mathbf{w}_{\mathrm{MMSE}}$ and $\mathbf{w}_{\mathrm{MBER}}$, together with the associated BERs given SNR$=12$ dB. From Table 1, it can be seen that $\mathbf{w}_{\mathrm{MMSE}}$ and $\mathbf{w}_{\mathrm{MBER}}$ are very different. The conditional p.d.f. of the real part of the MMSE equalizer output, given $b(k-d) = +1$ and with a SNR$=12$ dB, is compared with that of the MBER equalizer in Fig. 2. For this example, the subset $\mathcal{Y}_R^{(+)}$ contains $N_{sb} = 64$ points. In Fig. 2 the equalizer weight vector has been normalized to a unit length, so that the BER is main 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 equalizer is smaller than that for the MBER equalizer. Also, the density distribution of $y_R(k)$

for the MMSE equalizer is broader than that for the MBER equalizer. This means that the noise $e(k)$ at the MMSE equalizer output has a larger variance. These two factors explain why the MMSE equalizer has a higher BER than the MBER equalizer.

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. Fig. 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. Fig. 4 shows the learning curves of the algorithm 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. It can also be seen that, once the BER is below certain level (typically 0.01), decision-directed adaptation can be applied with the estimate $\hat{b}(k-d)$ substituting $b(k-d)$ as the desired response.

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 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 narrow-band which does not induce intersymbol 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

$$\mathbf{P} = [A_0\mathbf{s}_0 \; A_1\mathbf{s}_1 \cdots A_{M-1}\mathbf{s}_{M-1}], \tag{30}$$

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

$$\mathbf{s}_i = [\exp(j\omega t_0(\theta_i)) \;\; \exp(j\omega t_1(\theta_i)) \cdots \exp(j\omega t_{L-1}(\theta_i))]^T \tag{31}$$

with $t_l(\theta_i)$ being the relative time delay at array element l for source i and θ_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 $1 \le i \le M-1$. The beamformer at receiver is a linear filter in the form of (1) with $d = 0$ in the decision rule (3).

The simulation example used consisted six signal sources and a three-element antenna array. Fig. 5 shows the locations of the desired source and the interfering sources graphically. Fig. 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 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. Figs. 7 and 8 depict the conditional p.d.f.s of the MMSE and MBER beamformers given $b_0(k) = +1$ together with the associated subsets $\mathcal{Y}_R^{(+)}$, under the two same conditions as given in Tables 2 and 3. Again, in these two figures, the beamformer weight vector had been normalized 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 of $\frac{4}{32} = 0.125$.

Performance of the block-data gradient adaptive MBER algorithm portrayed in Section 4.1 was next tested. Fig. 9 illustrates the convergence rates of the algorithm given SNR$= 14$ dB and $\text{SIR}_i = 0$ dB for $1 \le i \le 5$, 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 $\mathbf{w}_{\text{MMSE}}$ is typically not a good initial choice for the adaptive MBER algorithm in terms of convergence rate. Performance of the stochastic gradient adaptive MBER algorithm described in Section 4.2 was also investigated. Fig. 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 equalization and nonlinear multiuser detection [27]–[30]. If the linear restriction is removed, it can readily be shown that the true optimal filtering solution is the maximum *a posteriori* probability or Bayesian one, which is formulated as:

$$y_B(k) = \frac{1}{N_s \left(2\pi\sigma_n^2\right)^L} \sum_{q=1}^{N_s} \text{sgn}(b_{q,d}) \exp\left(\frac{-\|\mathbf{x}(k) - \bar{\mathbf{x}}_q\|^2}{2\sigma_n^2}\right) \tag{32}$$

where $\bar{\mathbf{x}}_q \in \mathcal{X}$ and $\text{sgn}(b_{q,d})$ acts as class label. Note that $y_B(k)$ is real-valued due to its p.d.f. interpretation, but $\mathbf{x}(k)$ and $\bar{\mathbf{x}}_q$ are complex-valued. Because the number of vector states N_s 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:

$$y(k) = f(\mathbf{x}(k); \mathbf{w}) \tag{33}$$

where the nonlinear map f is generally complex-valued and is realized 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

$$\left.\begin{array}{l} y(k) = f(\mathbf{x}(k); \mathbf{w}(k-1)) \\ \mathbf{w}(k) = \mathbf{w}(k-1) + \mu(b_d(k) - y(k))\frac{\partial f^*(\mathbf{x}(k);\mathbf{w}(k-1))}{\partial \mathbf{w}} \end{array}\right\} \tag{34}$$

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

By linearizing the nonlinear filter (33) around $\bar{\mathbf{x}}(k)$, it can be approximated as

$$y(k) \approx \bar{y}(k) + e(k) \tag{35}$$

where

$$\bar{y}(k) = f(\bar{\mathbf{x}}(k); \mathbf{w}) \tag{36}$$

can only take the value from the set

$$\tilde{\mathcal{Y}} \triangleq \{\tilde{y}_q = f(\bar{\mathbf{x}}_q; \mathbf{w}),\ 1 \le q \le N_b\} \tag{37}$$

and

$$e(k) = \left[\frac{\partial f(\bar{\mathbf{x}}(k); \mathbf{w})}{\partial \mathbf{x}}\right]^H \mathbf{n}(k) \tag{38}$$

is Gaussian with zero mean and variance

$$\tilde{\rho}_n^2(\mathbf{w}) = \frac{2\sigma_n^2}{N_b}\sum_{q=1}^{N_b}\left[\frac{\partial f(\bar{\mathbf{x}}_q; \mathbf{w})}{\partial \mathbf{x}}\right]^H \frac{\partial f(\bar{\mathbf{x}}_q; \mathbf{w})}{\partial \mathbf{x}}. \tag{39}$$

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

$$p(y_R) \approx \frac{1}{N_b\sqrt{2\pi\tilde{\rho}_n^2(\mathbf{w})}}\sum_{q=1}^{N_b}\exp\left(-\frac{(y_R - \tilde{y}_{R,q})^2}{2\tilde{\rho}_n^2(\mathbf{w})}\right) \tag{40}$$

and the BER of the nonlinear filter is approximately

$$P_E(\mathbf{w}) \approx \frac{1}{N_b}\sum_{q=1}^{N_b} Q(g_q(\mathbf{w})) \tag{41}$$

with

$$g_q(\mathbf{w}) = \frac{\mathrm{sgn}(b_{q,d})\tilde{y}_{R,q}}{\tilde{\rho}_n(\mathbf{w})} = \frac{\mathrm{sgn}(b_{q,d})f_R(\bar{\mathbf{x}}_q; \mathbf{w})}{\tilde{\rho}_n(\mathbf{w})}. \tag{42}$$

Using the kernel density estimate in the form of (22) with a constant ρ_n^2 to approximate the density (40) naturally leads to a block-data based gradient adaptive near MBER algorithm for training the nonlinear filter (33). This can be further simplified to give rise to a stochastic gradient adaptive near MBER algorithm in the form:

$$\left.\begin{array}{l} y(k) = f(\mathbf{x}(k); \mathbf{w}(k-1)) \\ \mathbf{w}(k) = \mathbf{w}(k-1) + \mu \frac{\mathrm{sgn}(b_d(k))}{2\sqrt{2\pi}\rho_n} \exp\left(-\frac{y_R^2(k)}{2\rho_n^2}\right) \frac{\partial f_R(\mathbf{x}(k);\mathbf{w}(k-1))}{\partial \mathbf{w}} \end{array}\right\} \quad (43)$$

for a sample-by-sample adaptation. The derivative $\frac{\partial f_R}{\partial \mathbf{w}}$ depends on the particular nonlinear map employed. Previous studies [31]–[34] have applied this adaptive near MBER nonlinear filtering approach to equalization 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 p.d.f. 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.

Acknowledgment

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

References

[1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[2] S. Haykin, *Adaptive Filter Theory*. 3rd edition, Upper Saddle River, NJ: Prentice-Hall, 1996.

[3] E. Shamash and K. Yao, "On the structure and performance of a linear decision feedback equalizer based on the minimum error probability criterion," in *Proc. ICC'74*, 1974, pp.25F1–25F5.

[4] S. Chen, E.S. Chng, B. Mulgrew and G. Gibson, "Minimum-BER linear-combiner DFE," in *Proc. ICC'96* (Dallas, Texas), 1996, Vol.2, pp.1173–1177.

[5] C.C. Yeh and J.R. Barry, "Approximate minimum bit-error rate equalization for binary signaling," in *Proc. ICC'97* (Montreal, Canada), 1997, Vol.2, pp.1095-1099.

[6] S. Chen, B. Mulgrew, E.S. Chng and G. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *IEE Proc. Communications*, Vol.145, No.5, pp.316–322, 1998.

[7] S. Chen and B. Mulgrew, "The minimum-SER linear-combiner decision feedback equalizer," *IEE Proc. Communications*, Vol.146, No.6, pp.347–353, 1999.

[8] B. Mulgrew and S. Chen, "Stochastic gradient minimum-BER decision feedback equalisers," in *Proc. IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control* (Lake Louise, Alberta, Canada), Oct.1-4, 2000, pp.93–98.

[9] C.C. Yeh and J.R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Trans. Communications*, Vol.48, No.7, pp.1226–1235, 2000.

[10] B. Mulgrew and S. Chen, "Adaptive minimum-BER decision feedback equalisers for binary signalling," *Signal Processing*, Vol.81, No.7, pp.1479–1489, 2001.

[11] N.B. Mandayam and B. Aazhang, "Gradient estimation for sensitivity analysis and adaptive multiuser interference rejection in code-division multi-access systems," *IEEE Trans. Communications*, Vol.45, No.7, pp.848–858, 1997.

[12] C.C. Yeh, R.R. Lopes and J.R. Barry, "Approximate minimum bit-error rate multiuser detection," in *Proc. Globecom'98* (Sydney, Australia), Nov. 1998, pp.3590–3595.

[13] X.F. Wang, W.S. Lu and A. Antoniou, "Constrained minimum-BER multiuser detection," in *Proc. ICASSP* (Phoenix, USA), May 14-18, 1999, Vol.5, pp.2603–2606.

[14] I.N. Psaromiligkos, S.N. Batalama and D.A. Pados, "On adaptive minimum probability of error linear filter receivers for DS-CDMA channels," *IEEE Trans. Communications*, Vol.47, No.7, pp.1092–1102, 1999.

[15] S. Chen, A.K. Samingan, B. Mulgrew and L. Hanzo, "Adaptive minimum-BER linear multiuser detection," in *Proc. ICASSP* (Salt Lake City, Utah, USA), May 7-11, 2001, Vol.4, pp.2253–2256.

[16] S. Chen, A.K. Samingan, B. Mulgrew and L. Hanzo, "Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels," *IEEE Trans. Signal Processing*, Vol.49, No.6, pp.1240–1247, 2001.

[17] S. Chen, N.N. Ahmad and L. Hanzo, "Minimum bit error rate beamforming," submitted to *Electronics Letters*, 2002.

[18] S. Chen, N.N. Ahmad and L. Hanzo, "Smart beamforming for wireless communications: a novel minimum bit error rate approach," submitted to *2nd IMA Int. Conf. Mathematics in Communications* (Lancaster, U.K.), Dec. 16-18, 2002.

[19] S. Chen, N.N. Ahmad and L. Hanzo, "Adaptive minimum bit error rate beamforming," submitted to *IEEE Trans. Wireless Communications*, 2002.

[20] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, Vol.33, pp.1066–1076, 1962.

[21] B.W. Silverman, *Density Estimation*. London: Chapman Hall, 1996.

[22] A.W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press, 1997.

[23] S. Chen, B. Mulgrew and L. Hanzo, "Stochastic least-symbol-error-rate adaptive equalization for pulse-amplitude modulation," to be presented at *ICASSP 2002* (Orlando, Florida, USA), May 13-17, 2002.

[24] S. Chen, "Minimum symbol-error-rate equalisation," to be presented at *IEE Non-linear and Non-Gaussian Signal Processing Workshop* (Peebles, Hydro, Scotland), July 8-9, 2002.

[25] S. Chen, B. Mulgrew, A.K. Samingan and L. Hanzo, "Stochastic gradient adaptive minimum symbol-error-rate equalization for multi-level pulse-amplitude modulation," submitted to *IEE Proc. Communications*, 2002.

[26] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley, 1993.

[27] S. Chen, B. Mulgrew and P.M. Grant, "A clustering technique for digital communications channel equalisation using radial basis function networks," *IEEE Trans. Neural Networks*, Vol.4, No.4, pp.570–579, 1993.

[28] S. Chen, B. Mulgrew and S. McLaughlin, "Adaptive Bayesian equaliser with decision feedback," *IEEE Trans. Signal Processing*, Vol.41, No.9, pp.2918–2927, 1993.

[29] S. Chen, S. McLaughlin, B. Mulgrew and P.M. Grant, "Bayesian decision feedback equaliser for overcoming co-channel interference," *IEE Proc. Communications*, Vol.143, No.4, pp.219–225, 1996.

[30] S. Chen, A.K. Samingan and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels," *IEEE Trans. Neural Networks*, Vol.12, No.3, pp.604-611, 2001.

[31] S. Chen, B. Mulgrew and L. Hanzo, "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, pp.223–232.

[32] S. Chen, B. Mulgrew and L. Hanzo, "Least bit error rate adaptive nonlinear equalizers for binary signalling," *IEE Proc. Communications*, accepted, 2002.

[33] S. Chen, A.K. Samingan and L. Hanzo, "Adaptive near minimum error rate training for neural networks with application to multiuser detection in CDMA communication systems," submitted to *IEEE Trans. Neural Networks*, 2002.

[34] S. Chen, "Least bit error rate adaptive multiuser detection," submitted as a chapter contribution to: L.P. Wang, eds. *Soft Computing for Communications*, Physica Verlag, 2002.

Table 1: Weight vectors and BERs of the MMSE and MBER equalizers 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	4.15×10^{-5}	2.49×10^{-7}

Table 2: Weight vectors and BERs of the MMSE and MBER beamformers given SNR= 14 dB and SIR$_i = 0$ dB for $1 \le i \le 5$.

	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	1.00×10^{-2}	5.40×10^{-8}

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	1.25×10^{-1}	5.60×10^{-8}

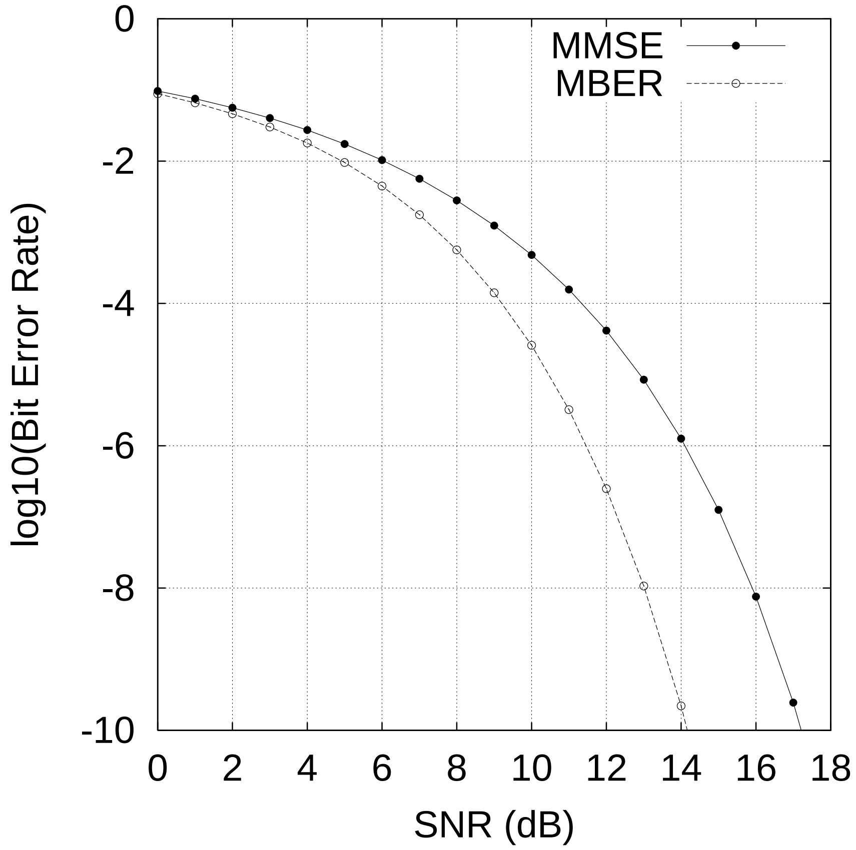
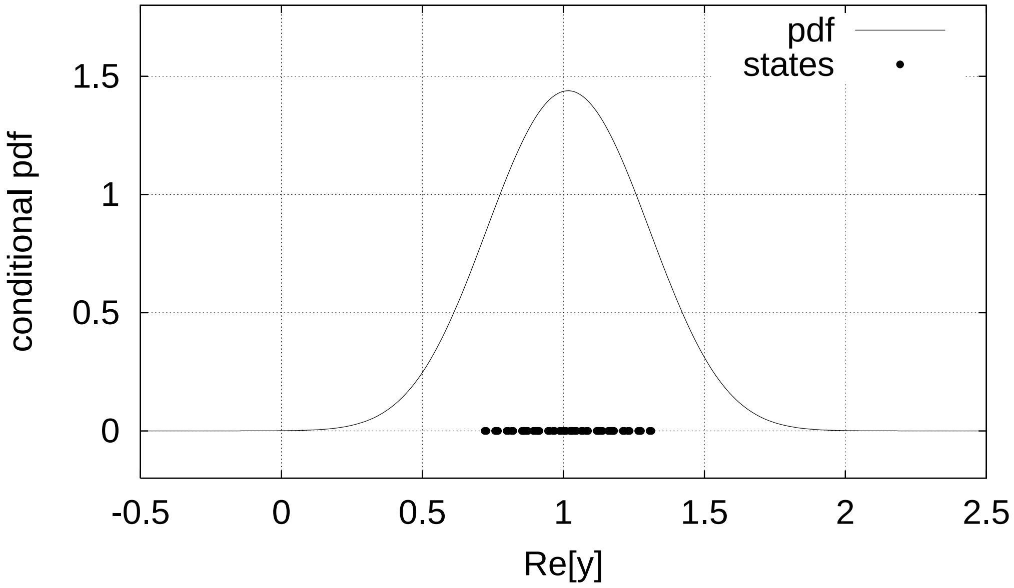
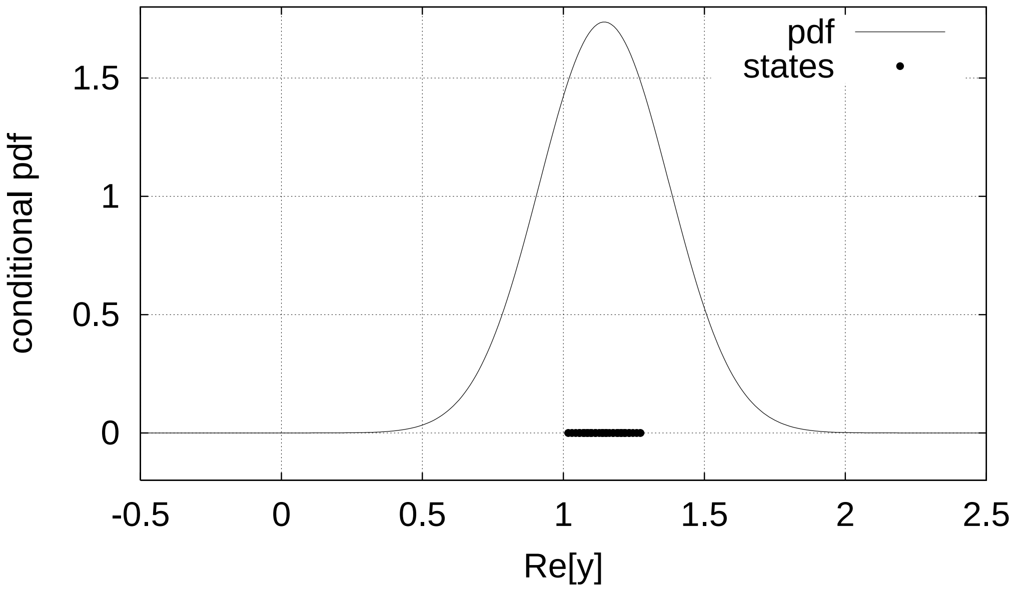


Figure 1: Bit error rate performance comparison of the MMSE and MBER equalizers.

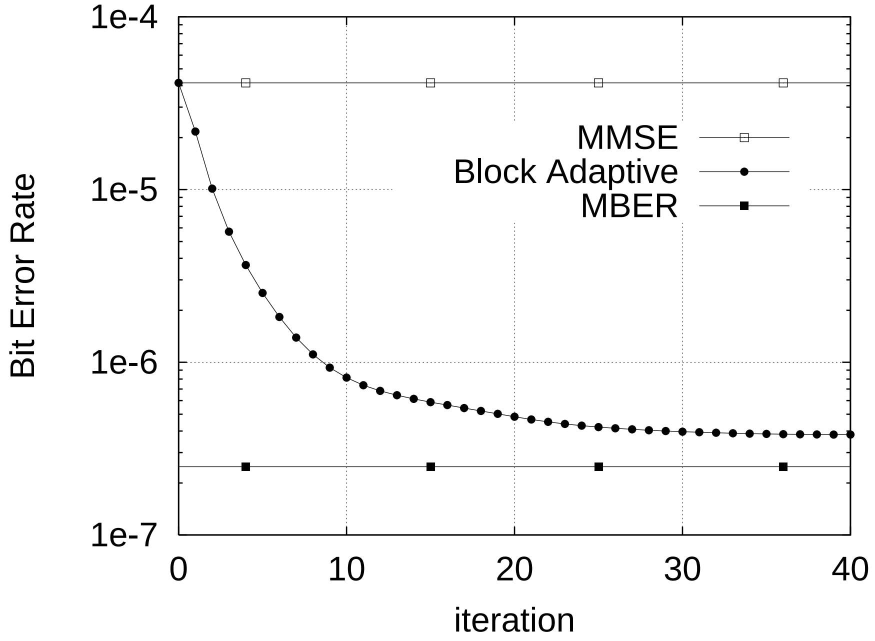


(a) MMSE

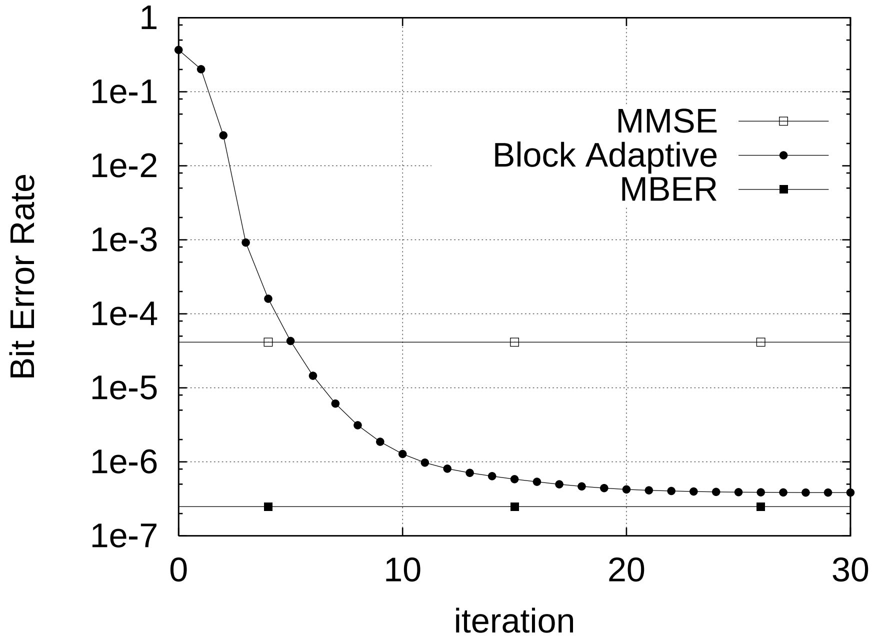


(b) MBER

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

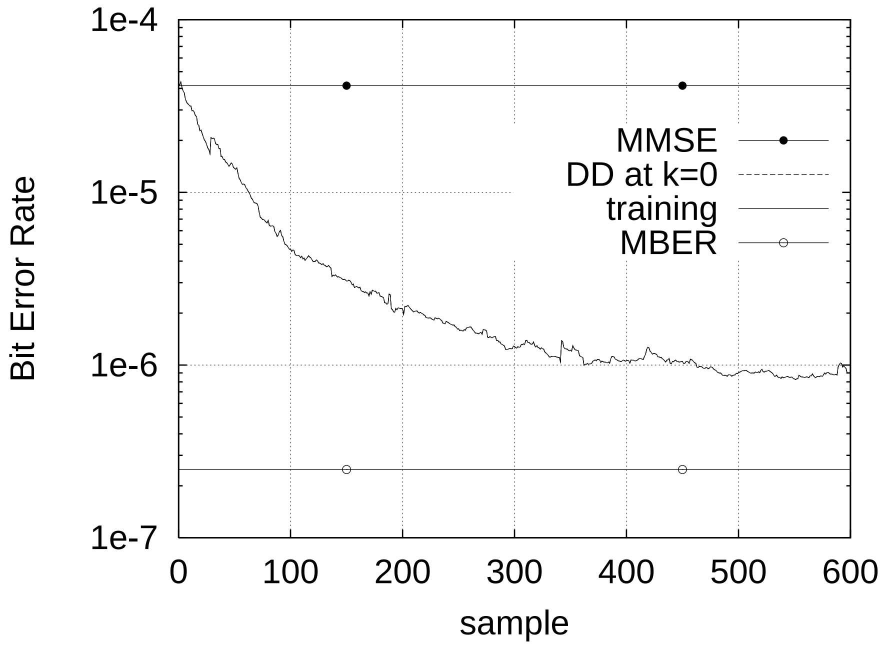


(a) $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$

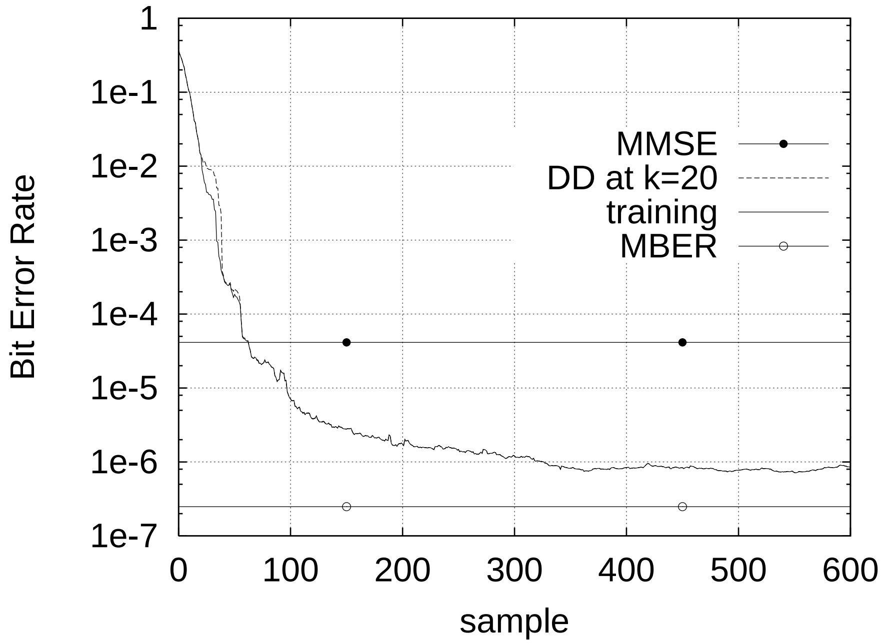


(b) $\mathbf{w}(0) = [0.1 + j0.0 \;\; 0.1 + j0.0 \;\; 0.1 + j0.0 \;\; 0.1 + j0.0 \;\; 0.1 + j0.0]^T$

Figure 3: Convergence rate of the block-data gradient adaptive MBER algorithm for the equalization 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) $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, step size $\mu = 0.1$ and square width $\rho_n^2 = 3\sigma_n^2 \approx 0.14$.



(b) $\mathbf{w}(0) = [0.1 + j0.0 \;\; 0.1 + j0.0 \;\; 0.1 + j0.0 \;\; 0.1 + j0.0 \;\; 0.1 + j0.0]^T$, step size $\mu = 0.3$ and square width $\rho_n^2 = 4\sigma_n^2 \approx 0.19$.

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

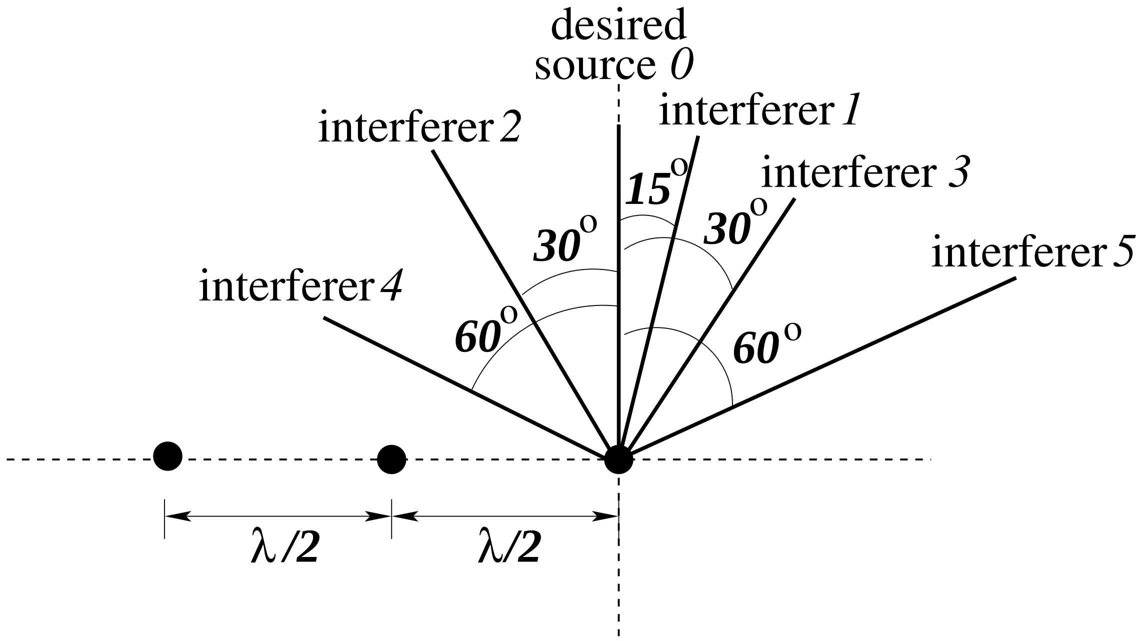
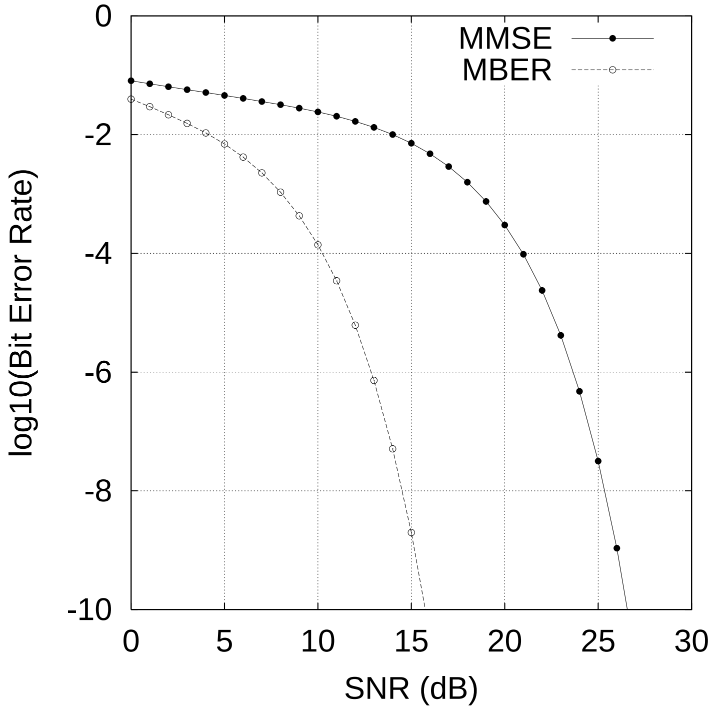
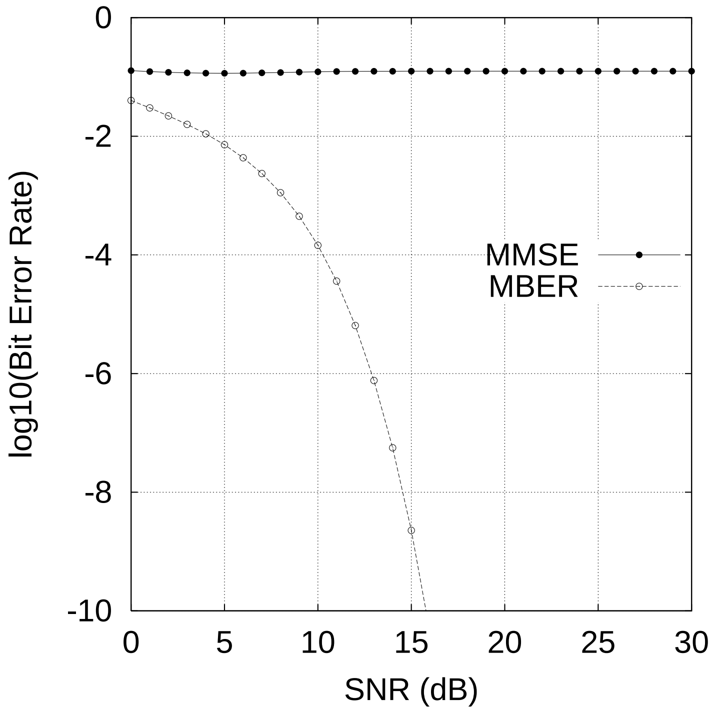


Figure 5: Locations of the desired source and the interfering sources with respect to the three-element linear antenna array having $\lambda/2$ element spacing, where λ is the wavelength.

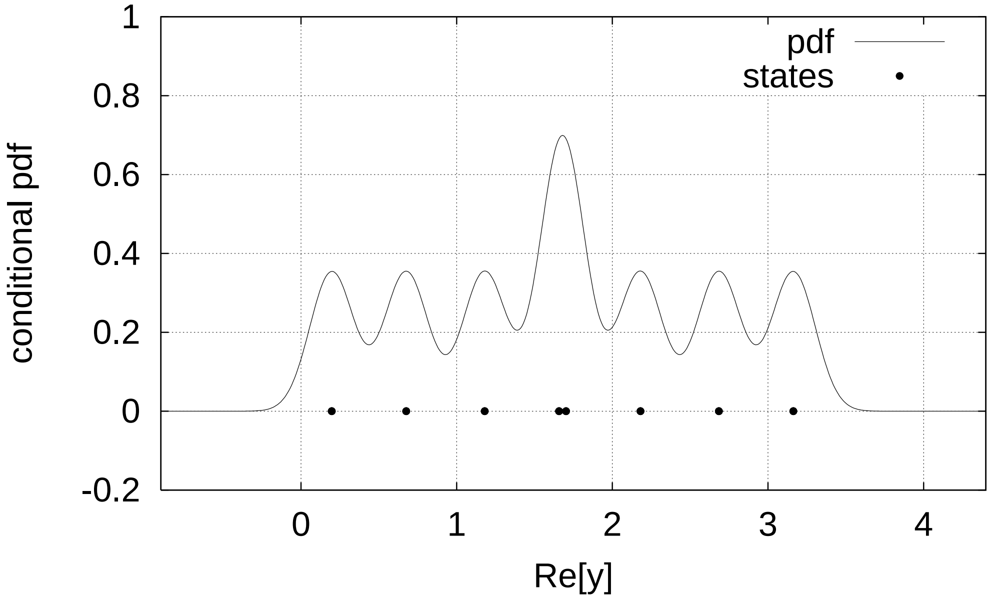


(a) $\mathrm{SIR}_i = 0$ dB for $1 \leq i \leq 5$.

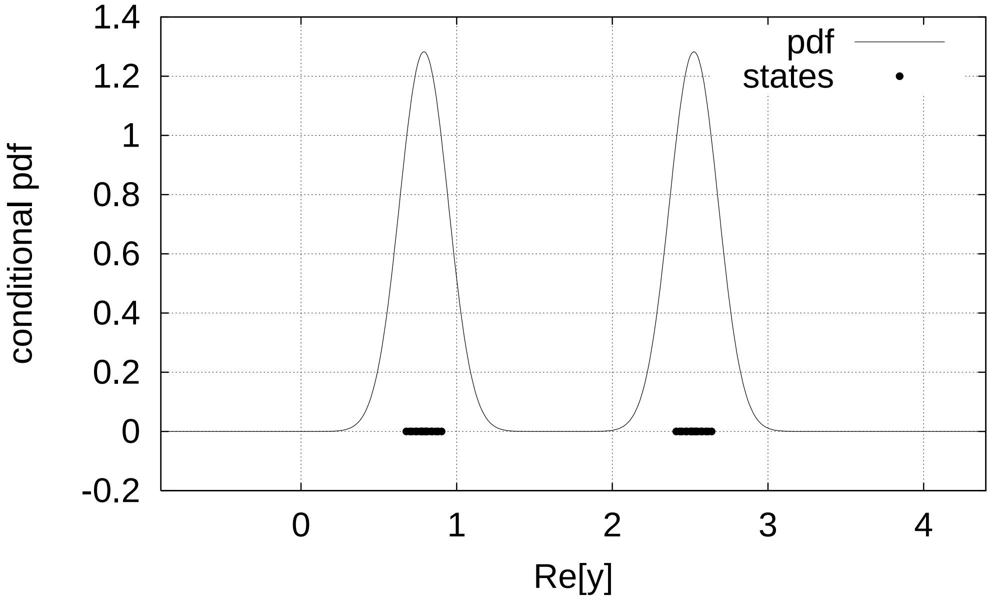


(b) $\mathrm{SIR}_i = 0$ dB for $i = 1, 3, 4, 5$, and $\mathrm{SIR}_2 = -6$ dB.

Figure 6: Bit error rate performance comparison of the MMSE and MBER beamformers.

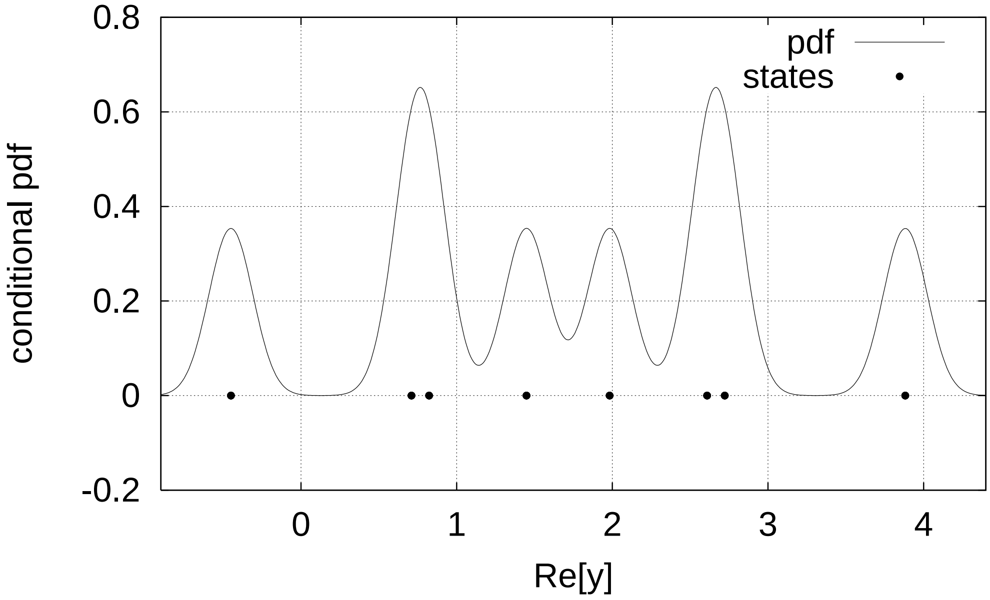


(a) MMSE

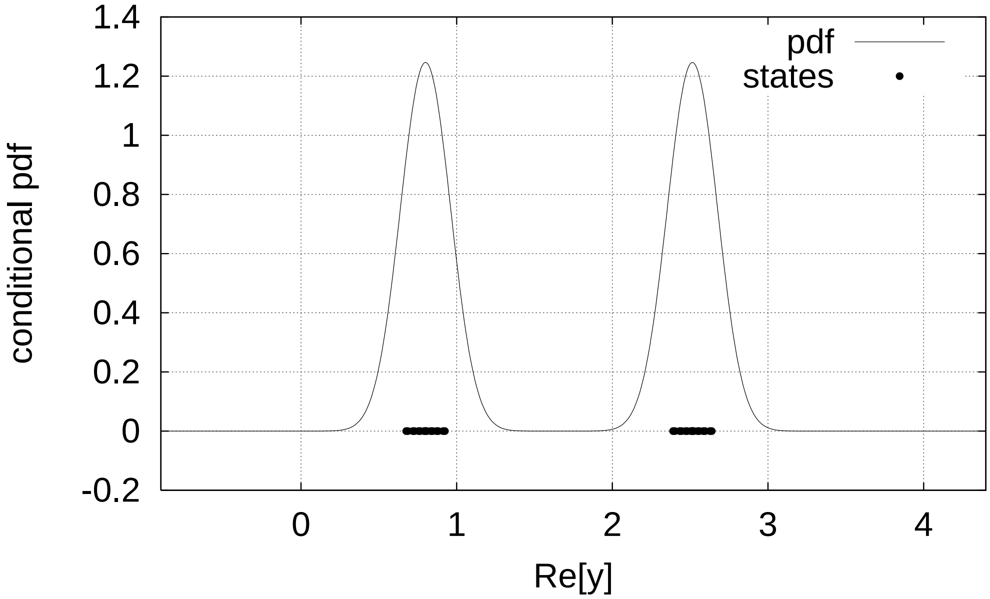


(b) MBER

Figure 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 $1 \leq i \leq 5$.

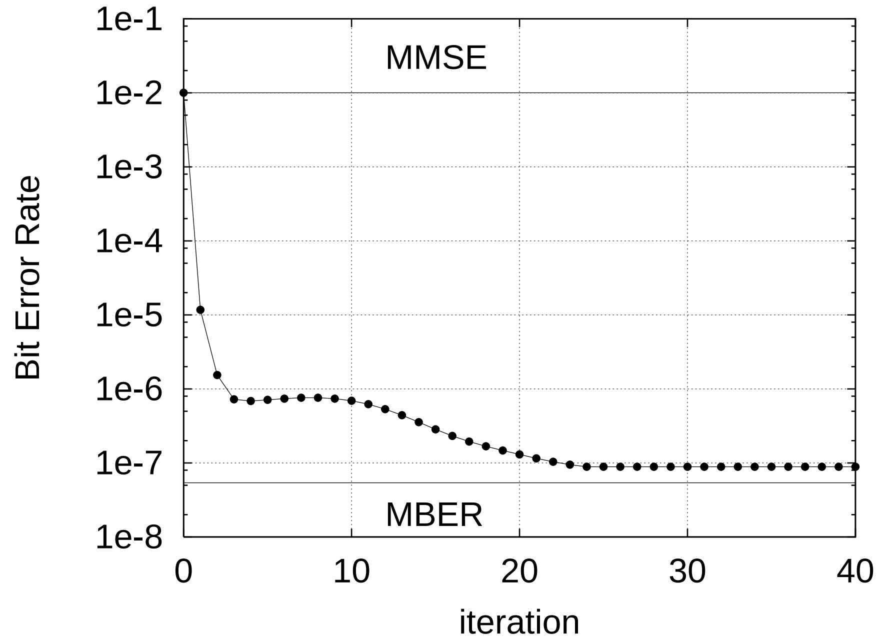


(a) MMSE

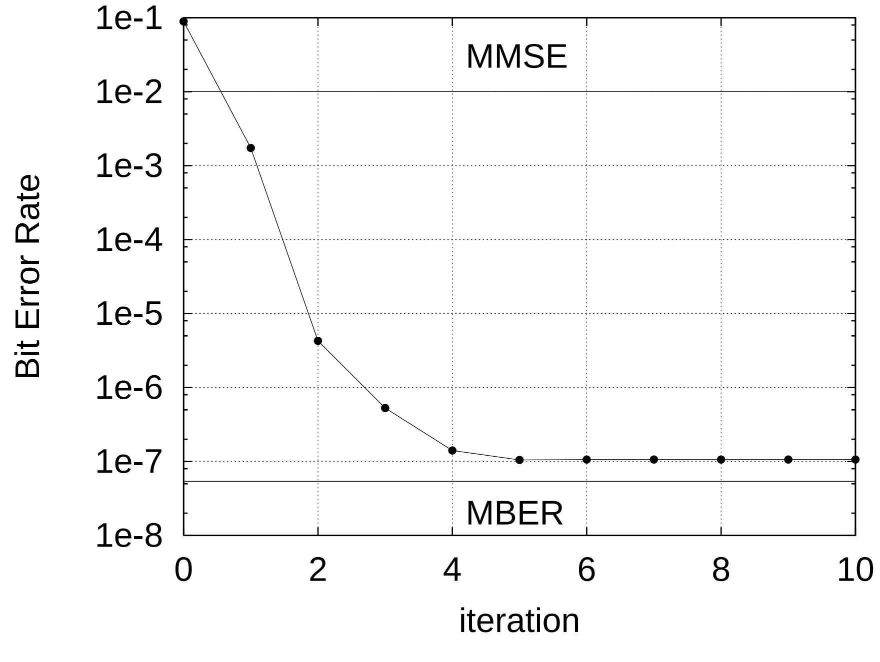


(b) MBER

Figure 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.

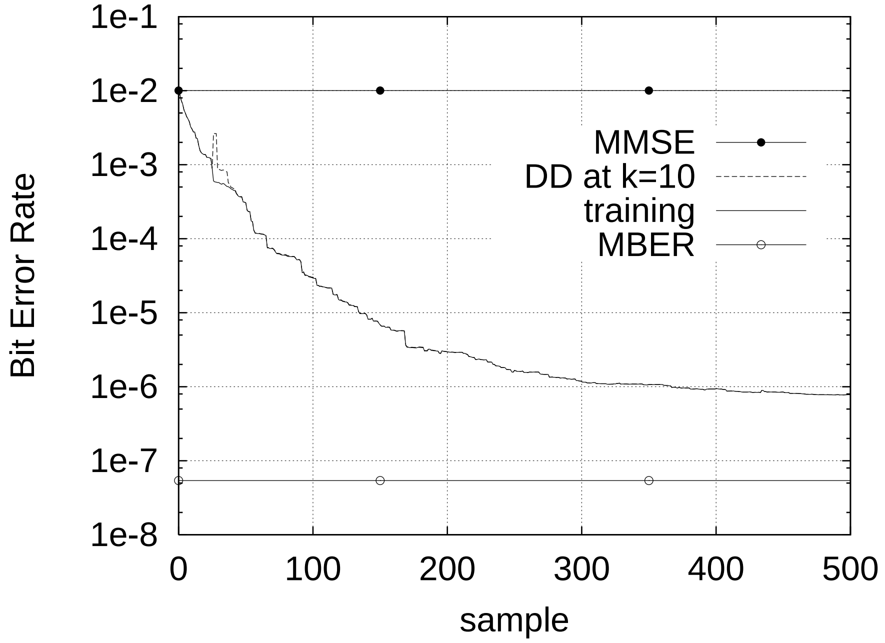


(a) $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$

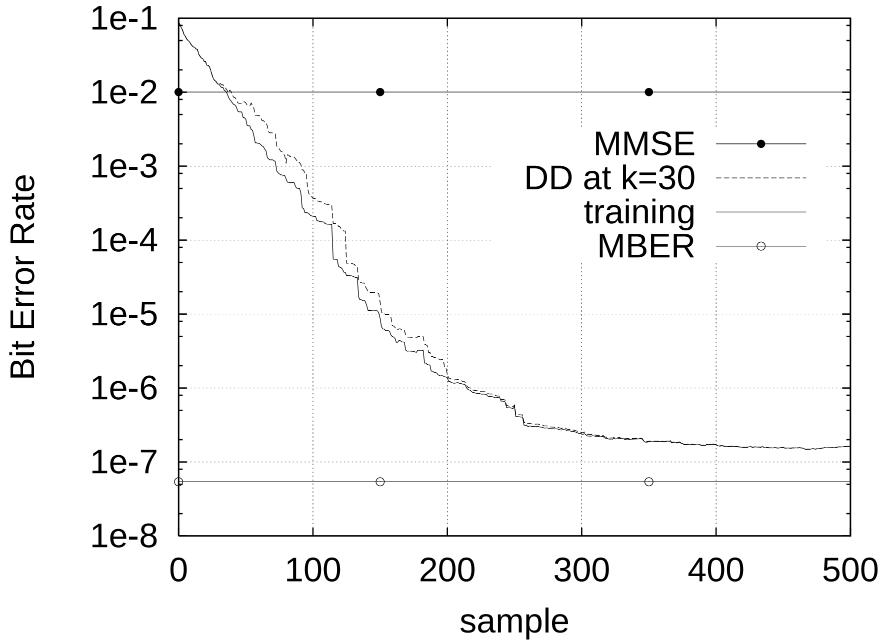


(b) $\mathbf{w}(0) = [0.1 + j0.0 \;\; 0.1 + j0.0 \;\; 0.1 + j0.0]^T$

Figure 9: Convergence rate of the 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 square width $\rho_n^2 = 4\sigma_n^2 \approx 0.08$.



(a) $\mathbf{w}(0) = \mathbf{w}_{\mathrm{MMSE}}$, 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 $\rho_n^2 = 4\sigma_n^2 \approx 0.08$.

Figure 10: Learning curves of the stochastic gradient adaptive MBER algorithm averaged over 30 runs for the beamforming example given SNR$= 14$ dB and SIR$_i = 0$ dB for $1 \le i \le 5$, where DD denotes decision directed adaptation with $\hat{b}_0(k)$ substituting $b_0(k)$.