

Adaptive Linear Filtering Design with Minimum Symbol Error Probability Criterion

Sheng Chen*

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

Abstract: Adaptive digital filtering has traditionally been developed based on the minimum mean square error (MMSE) criterion and has found ever-increasing applications in communications. This paper presents an alternative adaptive filtering design based on the minimum symbol error rate (MSER) criterion for communication applications. It is shown that the MSER filtering is smarter, as it exploits the non-Gaussian distribution of filter output effectively. Consequently, it provides significant performance gain in terms of smaller symbol error over the MMSE approach. Adopting Parzen window or kernel density estimation for a probability density function, a block-data gradient adaptive MSER algorithm is derived. A stochastic gradient adaptive MSER algorithm, referred to as the least symbol error rate, is further developed for sample-by-sample adaptive implementation of the MSER filtering. Two applications, involving single-user channel equalization and beamforming assisted receiver, are included to demonstrate the effectiveness and generality of the proposed adaptive MSER filtering approach.

Keywords: Adaptive filtering, mean square error, probability density function, non-Gaussian distribution, Parzen window estimate, symbol 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 minimum mean square error (MMSE) principle^[1,2]. One of the strengths of this design is that effective adaptive implementation can be achieved using the simple least mean square (LMS) algorithm, thus readily meet the real-time computational constraints of high-speed digital communication systems. For a communication system, however, it is the bit error rate (BER) or symbol error rate (SER), not the mean square error (MSE), that really matters. In digital communication applications, the probability density function (p.d.f.) of an adaptive filter output is generally a Gaussian mixture. This non-Gaussian nature can be exploited explicitly, leading to alternative approaches to the MMSE filtering. For single-user channel equalization application, adaptive minimum BER (MBER) linear equalizers and decision feedback equalizers (DFE) have been developed^[3~9]. Similar approaches have been adopted for linear multi-user detection in code-division multiple-access systems^[10~16]. The MBER beamforming using an antenna array for wireless communication has also been considered^[17~19]. These previous studies

have demonstrated that the MBER approach offers a potentially significant performance improvement and it provides a viable alternative to the traditional adaptive filtering based on the MMSE principle.

All these previous works on minimum error probability filtering are specifically designed for the simplest binary digital modulation scheme. Increasing demand for high-speed multimedia services over limited radio spectrum requires the employment of spectrum-efficient modulation schemes, such as multi-level quadrature amplitude modulation (M -QAM)^[20]. The main contribution of this paper is to extend the minimum error probability filtering to the general M -QAM case and to present a unified framework for adaptive MSER linear filtering. A complex-valued linear filtering model is given in the general communication setting, and the theoretical MSER filtering solution is derived. To effectively implement the MSER solution, a Parzen window p.d.f. estimation technique^[21~23] is adopted to approximate the non-Gaussian probability distribution of the filter output accurately, which naturally gives rise to a block-data gradient adaptive MSER algorithm. Sample-by-sample adaptive implementation of the MSER filtering is then considered, and a stochastic gradient adaptive MSER algorithm, referred to as the least symbol error rate (LSER), is derived which has a similar low computational complexity to the very simple LMS algorithm. Two applications, involving channel equalization and beamforming with

Manuscript received July 28, 2004; revised December 7, 2005.

*E-mail address: sqc@ecs.soton.ac.uk

antenna array, are used to illustrate the generality and effectiveness of the proposed adaptive MSER filtering. Simulation results obtained confirm the superior performance of the MSER filtering over the MMSE one.

2 System model

Consider the general tap-delay-line linear filter whose output is given by

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

where k denotes the sampling index, 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 complex-valued 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. In adaptive beamforming assisted receiver, $\mathbf{x}(k)$ consists of received signals at the antenna array. Generally, the underlying system that produces the filter input $\mathbf{x}(k)$ can be modelled 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 $L \times N$, and the transmitted information symbol vector $\mathbf{b}(k) = [b_0(k) \ b_1(k) \ \cdots \ b_{N-1}(k)]^T$. Typically, $b_i(k)$ and $b_q(k)$ are uncorrelated if $i \neq q$. For single-user applications, $\mathbf{b}(k)$ contains the current and previous $N - 1$ transmitted symbols and, for multi-user applications, $\mathbf{b}(k)$ consists of the transmitted symbols of different users. The modulation scheme is assumed to be M -QAM, and $b_i(k)$, $0 \leq i \leq N - 1$, takes values from the symbol set defined by

$$\mathcal{B} \triangleq \{b_{l,q} = u_l + ju_q, \ 1 \leq l, q \leq \sqrt{M}\} \quad (3)$$

where $u_l = 2l - \sqrt{M} - 1$ and $u_q = 2q - \sqrt{M} - 1$. The sampling rate is assumed to be equal to the symbol rate so that k also indicates the symbol index. The approach can be extended to the general case of sampling faster than the symbol rate, referred to as fractionally-spaced sampling^[24].

The purpose of the filter (1) is to enable an estimate $\hat{b}_d(k)$ of the ‘‘desired’’ symbol $b_d(k)$, the d -th element of $\mathbf{b}(k)$, where $0 \leq d \leq N - 1$. Consider the combined impulse response of the filter and the system given by

$$\mathbf{w}^H \mathbf{P} = \mathbf{w}^H [\mathbf{p}_0 \ \mathbf{p}_1 \ \cdots \ \mathbf{p}_{N-1}] = [c_0 \ c_1 \ \cdots \ c_{N-1}]. \quad (4)$$

The filter’s output can be expressed as

$$y(k) = c_d b_d(k) + \sum_{i \neq d} c_i b_i(k) + e(k) \quad (5)$$

where the first term is the desired signal and the second term is the residual interference. From (5), provided that c_d is real and positive, i.e. $c_d = c_{R_d} + jc_{I_d}$ satisfying $c_{R_d} > 0$ and

$$c_{I_d} = \text{Im}[\mathbf{w}^H \mathbf{p}_d] = 0 \quad (6)$$

the symbol decision $\hat{b}_d(k) = \hat{b}_{R_d}(k) + j\hat{b}_{I_d}(k)$ can be decoupled into the real and imaginary parts, given respectively by

$$\hat{b}_{R_d}(k) = \begin{cases} u_1, & \text{if } y_R(k) \leq c_{R_d}(u_1 + 1) \\ u_l, & \text{if } c_{R_d}(u_l - 1) < y_R(k) \leq c_{R_d}(u_l + 1) \\ & \text{for } 2 \leq l \leq \sqrt{M} - 1 \\ u_{\sqrt{M}}, & \text{if } y_R(k) > c_{R_d}(u_{\sqrt{M}} - 1) \end{cases} \quad (7)$$

$$\hat{b}_{I_d}(k) = \begin{cases} u_1, & \text{if } y_I(k) \leq c_{R_d}(u_1 + 1) \\ u_q, & \text{if } c_{R_d}(u_q - 1) < y_I(k) \leq c_{R_d}(u_q + 1) \\ & \text{for } 2 \leq q \leq \sqrt{M} - 1 \\ u_{\sqrt{M}}, & \text{if } y_I(k) > c_{R_d}(u_{\sqrt{M}} - 1) \end{cases} \quad (8)$$

where $y(k) = y_R(k) + jy_I(k)$. Fig. 1 depicts the decision boundaries associated with the decision $\hat{b}_d(k) = b_{l,q}$. In general, $\mathbf{w}^H \mathbf{p}_d$ is complex-valued and the rotating operation

$$\mathbf{w}^{\text{new}} = \frac{c_d^{\text{old}}}{|c_d^{\text{old}}|} \mathbf{w}^{\text{old}} \quad (9)$$

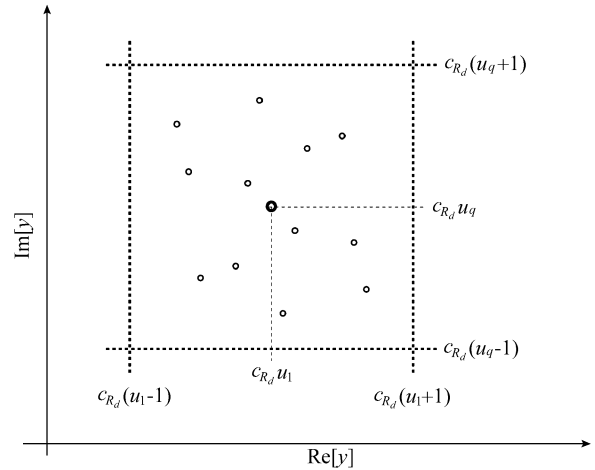


Fig. 1 Generic decision boundaries associated with point $c_d b_{l,q}$ assuming $c_{I_d} = 0$, and illustrations of symmetric distribution of $\mathcal{Y}_{l,q}$ around the mass center $c_d b_{l,q}$

can be used to make c_d real and positive. This rotation is a linear transformation and does not alter the SER of the underlying system. For the QAM modulation, the d -th column \mathbf{p}_d of the system matrix \mathbf{P} is required

at the receiver in order to make a decision regarding the transmitted symbol $b_d(k)$.

The classical design for the linear filter (1) is based on minimizing the MSE term of $E[|b_d(k) - y(k)|^2]$, which leads to the following MMSE solution

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

where $\sigma_b^2 = E[|b_i(k)|^2]$. The MMSE solution generally does not provide a minimum error probability, unless the conditional p.d.f. of $y(k)$ given $b_d(k) = b_{l,q}$ is Gaussian. However, the p.d.f. of $y(k)$ is obviously non-Gaussian, as can be seen from (5). Since the SER is the true performance indicator, it is desirable to consider the optimal MSER filter solution.

3 Minimum symbol error rate solution

Denote the $N_b = M^N$ number of possible sequences of $\mathbf{b}(k)$ as \mathbf{b}_i , $1 \leq i \leq N_b$. Then $\bar{\mathbf{x}}(k)$ can only take values from the finite signal set defined by

$$\mathcal{X} \triangleq \{\bar{\mathbf{x}}_i = \mathbf{P}\mathbf{b}_i, 1 \leq i \leq N_b\}. \quad (11)$$

This set can be partitioned into M subsets, depending on the specific value of $b_d(k)$, as follows:

$$\mathcal{X}_{l,q} \triangleq \{\bar{\mathbf{x}}_i \in \mathcal{X} : b_d(k) = b_{l,q}\}, 1 \leq l, q \leq \sqrt{M}. \quad (12)$$

Recall that the filter's output is given by

$$y(k) = \mathbf{w}^H \bar{\mathbf{x}}(k) + \mathbf{w}^H \mathbf{n}(k) = \bar{y}(k) + e(k) \quad (13)$$

where $e(k)$ is Gaussian distributed with zero mean and $E[|e(k)|^2] = 2\sigma_n^2 \mathbf{w}^H \mathbf{w}$. The noise-free part of the filter's output, namely $\bar{y}(k)$, only takes values from the scalar set

$$\mathcal{Y} \triangleq \{\bar{y}_i = \mathbf{w}^H \bar{\mathbf{x}}_i, 1 \leq i \leq N_b\} \quad (14)$$

and \mathcal{Y} can be divided into the M subsets conditioned on the value of $b_d(k)$:

$$\mathcal{Y}_{l,q} \triangleq \{\bar{y}_i \in \mathcal{Y} : b_d(k) = b_{l,q}\}, 1 \leq l, q \leq \sqrt{M}. \quad (15)$$

Lemma 1. The subsets $\mathcal{X}_{l,q}$, $1 \leq l, q \leq \sqrt{M}$, satisfy the shifting properties:

$$\mathcal{X}_{l+1,q} = \mathcal{X}_{l,q} + 2\mathbf{p}_d, 1 \leq l \leq \sqrt{M} - 1 \quad (16)$$

$$\mathcal{X}_{l,q+1} = \mathcal{X}_{l,q} + j2\mathbf{p}_d, 1 \leq q \leq \sqrt{M} - 1 \quad (17)$$

$$\mathcal{X}_{l+1,q+1} = \mathcal{X}_{l,q} + (2+j2)\mathbf{p}_d, 1 \leq l, q \leq \sqrt{M} - 1. \quad (18)$$

Proof. From the definitions of \mathbf{P} and \mathbf{b}_i , for each $\bar{\mathbf{x}}_i^{(l,q)} \in \mathcal{X}_{l,q}$, there exists a $\bar{\mathbf{x}}_i^{(l+1,q)} \in \mathcal{X}_{l+1,q}$, such that

$$\bar{\mathbf{x}}_i^{(l+1,q)} = \bar{\mathbf{x}}_i^{(l,q)} + (b_{l+1,q} - b_{l,q})\mathbf{p}_d = \bar{\mathbf{x}}_i^{(l,q)} + 2\mathbf{p}_d. \quad (19)$$

This verifies the shifting property (16). The proofs for (17) and (18) are similar. \square

Lemma 2. As a direct consequence of Lemma 1, the subsets $\mathcal{Y}_{l,q}$, $1 \leq l, q \leq \sqrt{M}$, satisfy the shifting properties:

$$\mathcal{Y}_{l+1,q} = \mathcal{Y}_{l,q} + 2c_d, 1 \leq l \leq \sqrt{M} - 1 \quad (20)$$

$$\mathcal{Y}_{l,q+1} = \mathcal{Y}_{l,q} + j2c_d, 1 \leq q \leq \sqrt{M} - 1 \quad (21)$$

$$\mathcal{Y}_{l+1,q+1} = \mathcal{Y}_{l,q} + (2+j2)c_d, 1 \leq l, q \leq \sqrt{M} - 1. \quad (22)$$

Lemma 3. The points of $\mathcal{Y}_{l,q}$ are distributed symmetrically around the symbol point $c_d b_{l,q}$.

Lemma 3 is obvious and is a direct consequence of symmetric distribution of the symbol constellation (3). This symmetric property is also illustrated in Fig. 1. Note that the distribution of $\mathcal{Y}_{l,q}$ is symmetric with respect to the two vertical decision boundaries $c_{R_d}(u_l \pm 1)$ and with respect to the two horizontal decision boundaries $c_{R_d}(u_q \pm 1)$.

For a linear filter to perform adequately it is implicitly assumed that the system is linearly separable. Linear separability is interpreted as follows: there exists \mathbf{w} satisfying condition (6) such that $\mathcal{Y}_{l,q}$ is completely separated from $\mathcal{Y}_{l+1,q}$ by the line $c_{R_d}(u_l + 1) + ju$, for $1 \leq l \leq \sqrt{M} - 1$ and $1 \leq q \leq \sqrt{M}$, and $\mathcal{Y}_{l,q}$ is completely separated from $\mathcal{Y}_{l,q+1}$ by the line $u + jc_{R_d}(u_q + 1)$, for $1 \leq l \leq \sqrt{M}$ and $1 \leq q \leq \sqrt{M} - 1$, where $u \in (-\infty, \infty)$ denotes a real-valued variable. Linear separability is not always guaranteed in practice. When the underlying system is linearly inseparable, a linear filter will have an irreducibly high SER floor and nonlinear filtering is required to achieve adequate performance^[25~31].

3.1 Symbol error rate expression

For the linear filter with weight vector \mathbf{w} satisfying (6), denote

$$P_E(\mathbf{w}) = \text{Prob}\{\hat{b}_d(k) \neq b_d(k)\} \quad (23)$$

$$P_{E_R}(\mathbf{w}) = \text{Prob}\{\hat{b}_{R_d}(k) \neq b_{R_d}(k)\} \quad (24)$$

$$P_{E_I}(\mathbf{w}) = \text{Prob}\{\hat{b}_{I_d}(k) \neq b_{I_d}(k)\}. \quad (25)$$

It is then easy to see that the SER is given by

$$P_E(\mathbf{w}) = P_{E_R}(\mathbf{w}) + P_{E_I}(\mathbf{w}) - P_{E_R}(\mathbf{w})P_{E_I}(\mathbf{w}). \quad (26)$$

The conditional p.d.f. of $y(k)$ given $b_d(k) = b_{l,q}$ is a Gaussian mixture defined by

$$p(y|b_{l,q}) = \frac{1}{N_{sb}2\pi\sigma_n^2 \mathbf{w}^H \mathbf{w}} \sum_{i=1}^{N_{sb}} e^{-\frac{|y - \bar{y}_i^{(l,q)}|^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \quad (27)$$

where $N_{sb} = N_b/M$ is the number of points in $\mathcal{Y}_{l,q}$, $\bar{y}_i^{(l,q)} = \bar{y}_{R_i}^{(l,q)} + j\bar{y}_{I_i}^{(l,q)} \in \mathcal{Y}_{l,q}$, and $y = y_R + jy_I$.

Note that c_d is real and the symbol decision is decoupled into (7) and (8). Taking into account the symmetric distribution of $\mathcal{Y}_{l,q}$, see Lemma 3, for $2 \leq l \leq \sqrt{M} - 1$, the conditional error probability of $\hat{b}_{R_d}(k) \neq u_l$ given $b_{R_d}(k) = u_l$ is

$$P_{E_R,l}(\mathbf{w}) = \frac{1}{N_{sb} 2\pi \sigma_n^2 \mathbf{w}^H \mathbf{w}} \sum_{i=1}^{N_{sb}} \int_{-\infty}^{\infty} e^{-\frac{(y_I - \bar{y}_{R_i}^{(l,q)})^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}} dy_I \times \left\{ \int_{-\infty}^{c_{R_d}(u_l-1)} + \int_{c_{R_d}(u_l+1)}^{\infty} \right\} e^{-\frac{(y_R - \bar{y}_{R_i}^{(l,q)})^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}} dy_R = \frac{2}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{R_i}^{(l,q)}(\mathbf{w})) \quad (28)$$

where

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} e^{-\frac{z^2}{2}} dz \quad (29)$$

$$g_{R_i}^{(l,q)}(\mathbf{w}) = \frac{\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l - 1)}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}}. \quad (30)$$

Further taking into account the shifting property, see Lemma 2 and noting that the error only occurs at one size in (28) if $l = 1$ or \sqrt{M} , it is straightforward to show that

$$P_{E_R}(\mathbf{w}) = \frac{\sqrt{M} - 1}{\sqrt{M}} \frac{2}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{R_i}^{(l,q)}(\mathbf{w})) = \gamma \frac{1}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{R_i}^{(l,q)}(\mathbf{w})) \quad (31)$$

where $\gamma = \frac{2\sqrt{M} - 2}{\sqrt{M}}$. It is seen that P_{E_R} can be evaluated using the real part of any single subset $\mathcal{Y}_{l,q}$.

Similarly, P_{E_I} can be evaluated using the imaginary part of any single subset $\mathcal{Y}_{l,q}$ as

$$P_{E_I}(\mathbf{w}) = \gamma \frac{1}{N_{sb}} \sum_{i=1}^{N_{sb}} Q(g_{I_i}^{(l,q)}(\mathbf{w})) \quad (32)$$

with

$$g_{I_i}^{(l,q)}(\mathbf{w}) = \frac{\bar{y}_{I_i}^{(l,q)} - c_{R_d}(u_q - 1)}{\sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}}. \quad (33)$$

3.2 Minimum symbol-error-rate solution

The MSER solution \mathbf{w}_{MSER} in principle is obtained by minimizing $P_E(\mathbf{w})$ with respect to \mathbf{w} . However, there is no simple way of doing so due to the cross coupled term $P_{E_R}(\mathbf{w})P_{E_I}(\mathbf{w})$. Instead, the MSER solution is defined as the one that minimizes the upper bound of the SER given by

$$P_{E_B}(\mathbf{w}) = P_{E_R}(\mathbf{w}) + P_{E_I}(\mathbf{w}). \quad (34)$$

That is,

$$\mathbf{w}_{\text{MSER}} = \arg \min_{\mathbf{w}} P_{E_B}(\mathbf{w}). \quad (35)$$

The upper bound $P_{E_B}(\mathbf{w})$ is very tight, i.e. very close to the true SER $P_E(\mathbf{w})$. Unlike the closed-form MMSE solution (10), a numerical optimization has to be applied to obtain an MSER solution \mathbf{w}_{MSER} . The gradients of $P_{E_R}(\mathbf{w})$ and $P_{E_I}(\mathbf{w})$ with respect to \mathbf{w} can be shown to be respectively

$$\nabla P_{E_R}(\mathbf{w}) = \frac{\gamma}{2N_{sb} \sqrt{2\pi} \sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{i=1}^{N_{sb}} e^{-\frac{(\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l-1))^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \times \left(\frac{\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l - 1)}{\mathbf{w}^H \mathbf{w}} \mathbf{w} - \bar{\mathbf{x}}_i^{(l,q)} + (u_l - 1) \mathbf{p}_d \right) \quad (36)$$

$$\nabla P_{E_I}(\mathbf{w}) = \frac{\gamma}{2N_{sb} \sqrt{2\pi} \sigma_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{i=1}^{N_{sb}} e^{-\frac{(\bar{y}_{I_i}^{(l,q)} - c_{R_d}(u_q-1))^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \times \left(\frac{\bar{y}_{I_i}^{(l,q)} - c_{R_d}(u_q - 1)}{\mathbf{w}^H \mathbf{w}} \mathbf{w} + j \bar{\mathbf{x}}_i^{(l,q)} + (u_q - 1) \mathbf{p}_d \right). \quad (37)$$

With the gradient $\nabla P_{E_B}(\mathbf{w}) = \nabla P_{E_R}(\mathbf{w}) + \nabla P_{E_I}(\mathbf{w})$, the optimization problem (35) can be solved for iteratively using a gradient-based optimization algorithm. Since the SER is invariant to a positive scaling of \mathbf{w} , it is computationally advantageous to normalize \mathbf{w} to a unit-length using

$$\mathbf{w} := \mathbf{w} / \sqrt{\mathbf{w}^H \mathbf{w}} \quad (38)$$

after every iteration, so that the gradients (36) and (37) can be simplified to:

$$\nabla P_{E_R}(\mathbf{w}) = \frac{\gamma}{2N_{sb} \sqrt{2\pi} \sigma_n} \sum_{i=1}^{N_{sb}} e^{-\frac{(\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l-1))^2}{2\sigma_n^2}} \times \left(\left(\bar{y}_{R_i}^{(l,q)} - c_{R_d}(u_l - 1) \right) \mathbf{w} - \bar{\mathbf{x}}_i^{(l,q)} + (u_l - 1) \mathbf{p}_d \right) \quad (39)$$

$$\nabla P_{E_I}(\mathbf{w}) = \frac{\gamma}{2N_{sb} \sqrt{2\pi} \sigma_n} \sum_{i=1}^{N_{sb}} e^{-\frac{(\bar{y}_{I_i}^{(l,q)} - c_{R_d}(u_q-1))^2}{2\sigma_n^2}} \times \left(\left(\bar{y}_{I_i}^{(l,q)} - c_{R_d}(u_q - 1) \right) \mathbf{w} + j \bar{\mathbf{x}}_i^{(l,q)} + (u_q - 1) \mathbf{p}_d \right). \quad (40)$$

The rotating operation (9) should also be applied after each iteration to ensure a real and positive c_d . A further saving in computation is achieved by choosing the subset with $l, q = 1 + \sqrt{M}/2$, giving rise to $u_l - 1 = u_q - 1 = 0$. We point out that the simplified conjugate gradient algorithm^[15,32] provides an efficient means to find an MSER solution.

4 Adaptive MSER filtering

To derive an adaptive version of the MSER filtering, it is more convenient to explicitly write down the p.d.f. of $y(k)$

$$p(y) = \frac{1}{N_b 2\pi \sigma_n^2 \mathbf{w}^H \mathbf{w}} \sum_{l=1}^{\sqrt{M}} \sum_{q=1}^{\sqrt{M}} \sum_{i=1}^{N_{sb}} e^{-\frac{|y - \hat{y}_i^{(l,q)}|^2}{2\sigma_n^2 \mathbf{w}^H \mathbf{w}}} \quad (41)$$

and to express the two error probabilities alternatively as

$$P_{E_R}(\mathbf{w}) = \frac{\gamma}{N_b} \sum_{l=1}^{\sqrt{M}} \sum_{q=1}^{\sqrt{M}} \sum_{i=1}^{N_{sb}} Q(g_{R_i}^{(l,q)}(\mathbf{w})) \quad (42)$$

$$P_{E_I}(\mathbf{w}) = \frac{\gamma}{N_b} \sum_{l=1}^{\sqrt{M}} \sum_{q=1}^{\sqrt{M}} \sum_{i=1}^{N_{sb}} Q(g_{I_i}^{(l,q)}(\mathbf{w})). \quad (43)$$

In reality, the p.d.f. of $y(k)$ is unknown. Parzen window or kernel density estimate^[21–23] is a well known method for estimating a probability distribution. Parzen window method estimates a p.d.f. using a window or block of $y(k)$ by placing a symmetric unimodal kernel function on each $y(k)$. Kernel density estimation is capable of producing reliable p.d.f. estimates with short data records and is natural when dealing with Gaussian mixtures, such as the one given in (41). In our 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 MSER algorithm

Given a block of K training samples $\{\mathbf{x}(k), b_d(k)\}_{k=1}^K$, a kernel density estimate of the p.d.f. (41) is readily given by

$$\hat{p}(y) = \frac{1}{K 2\pi \rho_n^2 \mathbf{w}^H \mathbf{w}} \sum_{k=1}^K e^{-\frac{|y - y(k)|^2}{2\rho_n^2 \mathbf{w}^H \mathbf{w}}} \quad (44)$$

where the scaling parameter ρ_n is related to the standard deviation σ_n of the system noise. In [22], a lower bound of $\rho_n = \left(\frac{4}{3K}\right)^{\frac{1}{5}} \sigma_n$ is suggested. In practice, ρ_n can often be chosen from a large range of values.

From this estimated p.d.f. (44), the estimated upper-bound SER expression is given by

$$\begin{aligned} \hat{P}_{E_B}(\mathbf{w}) &= \hat{P}_{E_R}(\mathbf{w}) + \hat{P}_{E_I}(\mathbf{w}) = \\ &= \frac{\gamma}{K} \sum_{k=1}^K (Q(\hat{g}_{R_k}(\mathbf{w})) + Q(\hat{g}_{I_k}(\mathbf{w}))) \end{aligned} \quad (45)$$

with

$$\hat{g}_{R_k}(\mathbf{w}) = \frac{y_R(k) - \hat{c}_{R_d}(b_{R_d}(k) - 1)}{\rho_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (46)$$

$$\hat{g}_{I_k}(\mathbf{w}) = \frac{y_I(k) - \hat{c}_{R_d}(b_{I_d}(k) - 1)}{\rho_n \sqrt{\mathbf{w}^H \mathbf{w}}} \quad (47)$$

where $\hat{c}_{R_d} = \text{Re}[\mathbf{w}^H \hat{\mathbf{p}}_d]$ and $\hat{\mathbf{p}}_d$ an estimate of \mathbf{p}_d . The gradient of $\hat{P}_{E_B}(\mathbf{w})$ can readily be calculated with

$$\begin{aligned} \nabla \hat{P}_{E_R}(\mathbf{w}) &= \frac{\gamma}{2K \sqrt{2\pi} \rho_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{i=1}^K e^{-\frac{(y_R(k) - \hat{c}_{R_d}(b_{R_d}(k) - 1))^2}{2\rho_n^2 \mathbf{w}^H \mathbf{w}}} \\ &\times \left(\frac{y_R(k) - \hat{c}_{R_d}(b_{R_d}(k) - 1)}{\mathbf{w}^H \mathbf{w}} \mathbf{w} - \mathbf{x}(k) + (b_{R_d}(k) - 1) \hat{\mathbf{p}}_d \right) \end{aligned} \quad (48)$$

$$\begin{aligned} \nabla \hat{P}_{E_I}(\mathbf{w}) &= \frac{\gamma}{2K \sqrt{2\pi} \rho_n \sqrt{\mathbf{w}^H \mathbf{w}}} \sum_{i=1}^K e^{-\frac{(y_I(k) - \hat{c}_{R_d}(b_{I_d}(k) - 1))^2}{2\rho_n^2 \mathbf{w}^H \mathbf{w}}} \\ &\times \left(\frac{y_I(k) - \hat{c}_{R_d}(b_{I_d}(k) - 1)}{\mathbf{w}^H \mathbf{w}} \mathbf{w} + j\mathbf{x}(k) + (b_{I_d}(k) - 1) \hat{\mathbf{p}}_d \right) \end{aligned} \quad (49)$$

Upon substituting $\nabla P_{E_B}(\mathbf{w})$ by $\nabla \hat{P}_{E_B}(\mathbf{w}) = \nabla \hat{P}_{E_R}(\mathbf{w}) + \nabla \hat{P}_{E_I}(\mathbf{w})$ in the simplified conjugate gradient updating mechanism, a block-data based adaptive algorithm is obtained. The step size μ of the conjugate gradient algorithm and the scaling parameter ρ_n are two algorithmic parameters, which control the rate of convergence and determine the accuracy of the p.d.f. and hence SER estimate.

4.2 Stochastic gradient adaptive MSER algorithm

In the Parzen window estimate (44), 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 width parameter should be chosen in such a way except that the dependency of the width parameter to \mathbf{w} in the true density (41) is noticed. However, the SER is invariant to $\mathbf{w}^H \mathbf{w}$. To fully take advantage of this fact, it is proposed to use a constant width ρ_n in density estimates. One 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 the resulting estimated SER has a much simpler form, which leads to considerable reduction in computational complexity. This is particularly relevant to the derivation of stochastic gradient updating mechanisms. Adopting this approach, an alternative Parzen window estimate of the true p.d.f. (41) is given by

$$\tilde{p}(y) = \frac{1}{K 2\pi \rho_n^2} \sum_{k=1}^K e^{-\frac{|y - y(k)|^2}{2\rho_n^2}} \quad (50)$$

and an approximation of the upper-bound SER is

$$\tilde{P}_{E_B}(\mathbf{w}) = \tilde{P}_{E_R}(\mathbf{w}) + \tilde{P}_{E_I}(\mathbf{w}) =$$

$$\frac{\gamma}{K} \sum_{k=1}^K (Q(\tilde{g}_{R_k}(\mathbf{w})) + Q(\tilde{g}_{I_k}(\mathbf{w}))) \quad (51)$$

with

$$\tilde{g}_{R_k}(\mathbf{w}) = \frac{y_R(k) - \hat{c}_{R_d}(b_{R_d}(k) - 1)}{\rho_n} \quad (52)$$

$$\tilde{g}_{I_k}(\mathbf{w}) = \frac{y_I(k) - \hat{c}_{R_d}(b_{I_d}(k) - 1)}{\rho_n}. \quad (53)$$

This approximation is valid, provided that the width ρ_n is chosen appropriately.

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

$$\tilde{p}(y, k) = \frac{1}{2\pi\rho_n^2} e^{-\frac{|y-y(k)|^2}{2\rho_n^2}} \quad (54)$$

and the corresponding one-sample SER “estimate” $\tilde{P}_{E_B}(\mathbf{w}, k)$. Using the instantaneous stochastic gradient of $\nabla \tilde{P}_{E_B}(\mathbf{w}, k) = \nabla \tilde{P}_{E_R}(\mathbf{w}, k) + \nabla \tilde{P}_{E_I}(\mathbf{w}, k)$ with

$$\nabla \tilde{P}_{E_R}(\mathbf{w}, k) = \frac{\gamma}{2\sqrt{2\pi}\rho_n} e^{-\frac{(y_R(k) - \hat{c}_{R_d}(k)(b_{R_d}(k) - 1))^2}{2\rho_n^2}} \times (-\mathbf{x}(k) + (b_{R_d}(k) - 1)\hat{\mathbf{p}}_d) \quad (55)$$

$$\nabla \tilde{P}_{E_I}(\mathbf{w}, k) = \frac{\gamma}{2\sqrt{2\pi}\rho_n} e^{-\frac{(y_I(k) - \hat{c}_{R_d}(k)(b_{I_d}(k) - 1))^2}{2\rho_n^2}} \times (j\mathbf{x}(k) + (b_{I_d}(k) - 1)\hat{\mathbf{p}}_d) \quad (56)$$

gives rise to a stochastic gradient adaptive algorithm, referred to as the LSER algorithm:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu \left(-\nabla \tilde{P}_{E_B}(\mathbf{w}(k), k) \right) \quad (57)$$

$$\hat{c}_d(k+1) = \mathbf{w}^H(k+1)\hat{\mathbf{p}}_d \quad (58)$$

$$\mathbf{w}(k+1) = \frac{\hat{c}_d(k+1)}{|\hat{c}_d(k+1)|} \mathbf{w}(k+1) \quad (59)$$

where the adaptive gain μ and the kernel width ρ_n should be set appropriately to ensure an adequate performance in terms of convergence rate and steady-state SER misadjustment. Note that there is no need to normalize the weight vector to a unit-length after each update.

It is interesting to see some analogy between the traditional adaptive filtering approach based on the MMSE criterion and the proposed adaptive MSER 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 MSER 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 MSER algorithm is formulated.

5 Application examples

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

5.1 Single-user channel equalization

In the communication system involving a dispersive channel, the received signal sample can be expressed as^[20]

$$r(k) = \sum_{i=0}^{n_a-1} a_i b(k-i) + n(k) \quad (60)$$

where n_a is the channel impulse response (CIR) length, a_i are complex-valued channel taps, $\{b(k)\}$ is the transmitted data symbol sequence, and $n(k)$ a complex-valued additive white Gaussian noise with $E[|n(k)|^2] = 2\sigma_n^2$. The system signal to noise ratio (SNR) is defined as $\text{SNR} = \mathbf{a}^H \mathbf{a} \sigma_b^2 / 2\sigma_n^2$, where $\mathbf{a} = [a_0 \ a_1 \ \dots \ a_{n_a-1}]^T$ is the channel tap vector and $E[|b(k)|^2] = \sigma_b^2$. A decision feedback equalizer (DFE) is employed at the receiver, which takes the form

$$y(k) = \mathbf{w}^H \mathbf{r}(k) + \mathbf{f}^H \hat{\mathbf{b}}_{fb}(k) \quad (61)$$

where $\mathbf{r}(k) = [r(k) \ r(k-1) \ \dots \ r(k-L+1)]^T$ is the observation vector, $\hat{\mathbf{b}}_{fb}(k) = [\hat{b}(k-d-1) \ \dots \ \hat{b}(k-d-n_{fb})]^T$ is the past detected symbol vector, \mathbf{w} and $\mathbf{f} = [f_1 \ \dots \ f_{n_{fb}}]^T$ are the feedforward and feedback filter coefficient vectors with orders L and n_{fb} respectively. At symbol instance k , the DFE provides an estimate $\hat{b}(k-d)$ of the transmitted symbol $b(k-d)$, where d is called the decision delay. The DFE structure parameters will be chosen as

$$L = n_a, \ n_{fb} = n_a - 1, \ d = n_a - 1 \quad (62)$$

as this choice is sufficient to guarantee a desired linear separability, see Lemma 4.

The DFE (61) can be translated into the linear filter (1) under the assumption that the past decisions are correct (also see [6]). The received signal vector can be expressed as

$$\mathbf{r}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k) = \mathbf{A}\mathbf{b}(k) + \mathbf{n}(k) \quad (63)$$

where the $L \times (L + n_a - 1)$ CIR matrix \mathbf{A} has the form

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & \dots & a_{n_a-1} & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_{n_a-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \dots & \ddots & 0 \\ 0 & \dots & 0 & a_0 & a_1 & \dots & a_{n_a-1} \end{bmatrix} =$$

$$[\mathbf{P} \mid \mathbf{P}_{fb}] \quad (64)$$

with \mathbf{P} having a dimension of $L \times L$ and \mathbf{P}_{fb} having a dimension of $L \times (n_a - 1)$, and

$$\begin{aligned} \mathbf{b}(k) = & [b(k) \ b(k-1) \cdots b(k-d) \mid \\ & b(k-d-1) \cdots b(k-L-n_a+2)]^T = \\ & [\mathbf{b}_{ff}^T(k) \mid \mathbf{b}_{fb}^T(k)]^T. \end{aligned} \quad (65)$$

Noticing (62), the last column of \mathbf{P} is $\mathbf{p}_d = [a_{n_a-1} \cdots a_1 \ a_0]^T$. Under the assumption that the past decisions are correct, $\hat{\mathbf{b}}_{fb}(k) = \mathbf{b}_{fb}(k)$ and $\mathbf{r}(k)$ can be expressed as

$$\mathbf{r}(k) = \mathbf{P}\mathbf{b}_{ff}(k) + \mathbf{P}_{fb}\hat{\mathbf{b}}_{fb}(k) + \mathbf{n}(k). \quad (66)$$

Thus, the decision feedback translates the original observation space $\mathbf{r}(k)$ into a new space $\mathbf{x}(k)$ defined by

$$\begin{aligned} \mathbf{x}(k) \triangleq & \mathbf{r}(k) - \mathbf{P}_{fb}\hat{\mathbf{b}}_{fb}(k) = \mathbf{P}\mathbf{b}_{ff}(k) + \mathbf{n}(k) = \\ & \bar{\mathbf{x}}(k) + \mathbf{n}(k). \end{aligned} \quad (67)$$

In this translated observation space, the DFE (61) becomes

$$y(k) = \mathbf{w}^H \mathbf{x}(k) = \bar{y}(k) + e(k). \quad (68)$$

The feedback filter coefficients do not disappear. They have been set to their optimal values $\mathbf{f}_{\text{opt}} = -\mathbf{P}_{fb}^H \mathbf{w}_{\text{MMSE}}$, where \mathbf{w}_{MMSE} denotes the MMSE solution for \mathbf{w} . There is no need to explicitly estimate \mathbf{f}_{opt} , because the elements of $\mathbf{x}(k)$ can be computed recursively according to [6] as

$$\begin{cases} x(k-i) = z^{-1}x(k-i+1) - a_{n_a-i}\hat{b}(k-d-1), \\ \quad \quad \quad i = L-1, \dots, 2, 1 \\ x(k) = r(k) \end{cases} \quad (69)$$

where z^{-1} denotes the unit delay operator. Thus, in an adaptive implementation, one needs to estimate the CIR using for example the LMS algorithm, rather than estimate \mathbf{f} . Note that the CIR is needed anyway in order to make the symbol decision according to (7) and (8).

The equalizer (68) and its observation model (67) have identical forms to the linear filter (1) and its observation model (2), with $N = L = n_a$. Furthermore, the DFE has the following desired property of linear separability.

Lemma 4. With the choice of DFE structure (62), $\mathcal{Y}_{l,q}$, $1 \leq l, q \leq \sqrt{M}$, are linearly separable.

Proof. Choose the weight vector $\mathbf{w} = [0 \ 0 \ \cdots \ 0 \ \frac{1}{a_0}^T]^T$. It is obvious that condition (6) is satisfied, as $\mathbf{w}^H \mathbf{p}_d = 1$. For $l = 1, \dots, \sqrt{M} - 1$ and $1 \leq q \leq \sqrt{M}$, it is easy to see that

$$\mathbf{w}^H \bar{\mathbf{x}}_i^{(l,q)} = b_{l,q}, \quad \forall \bar{\mathbf{x}}_i^{(l,q)} \in \mathcal{X}_{l,q} \quad (70)$$

$$\mathbf{w}^H \bar{\mathbf{x}}_i^{(l+1,q)} = b_{l+1,q}, \quad \forall \bar{\mathbf{x}}_i^{(l+1,q)} \in \mathcal{X}_{l+1,q}. \quad (71)$$

That is, $\mathcal{Y}_{l,q}$ is completely separated from $\mathcal{Y}_{l+1,q}$ by the line $(u_l + 1) + ju$. Similarly, for $1 \leq l \leq \sqrt{M}$ and $q = 1, \dots, \sqrt{M} - 1$, $\mathcal{Y}_{l,q}$ is completely separated from $\mathcal{Y}_{l,q+1}$ by the line $u + j(u_q + 1)$. \square

In the simulation, the 16-QAM symbols were transmitted through the two-tap channel $\mathbf{a}^T = [0.2 - j0.2 \ -1.0 + j1.0]$. The DFE structure parameters were set according to (62) as $L = 2$, $d = 1$ and $n_{fb} = 1$. Fig. 2 compares the SER performance of the MMSE DFE with

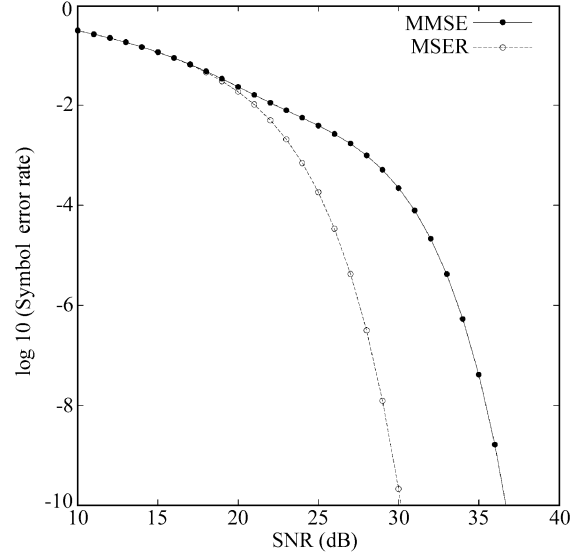


Fig. 2 Symbol error rate performance comparison of the MMSE and MSER DFEs

that of the MSER one, where it can be seen that the MSER solution offers more than 5 dB improvement in SNR at the SER level of 10^{-4} over the MMSE solution. The conditional p.d.f. $p(y|b_{l,q})$, the two marginal conditional p.d.f.'s $p(y_R|b_{l,q})$ and $p(y_I|b_{l,q})$, the subset $\mathcal{Y}_{l,q}$ and its real and imaginary parts of the MMSE DFE output, given $b(k-d) = 1 + j$ and with an SNR = 27 dB, are compared with those of the MSER DFE in Fig. 3. For this example, the total number of signal points is $N_b = 16^2 = 256$ and therefore $\mathcal{Y}_{l,q}$ contains $N_{sb} = 16$ points. In Fig. 3, the equalizer weight vector \mathbf{w} has been normalized to a unit length, so that the SER is determined by the minimum distance from the subset $\mathcal{Y}_{l,q}$ to the corresponding decision boundaries. It can be seen from Fig. 3 that this minimum distance is about 0.4 for the MMSE DFE and about 0.6 for the MSER DFE. This explains why the MMSE DFE has a higher SER than the MSER DFE. The conditional density distribution of $y(k)$ for the MMSE DFE looks more like a Gaussian one, while the non-Gaussian nature of

this density distribution is evident for the MSER DFE.

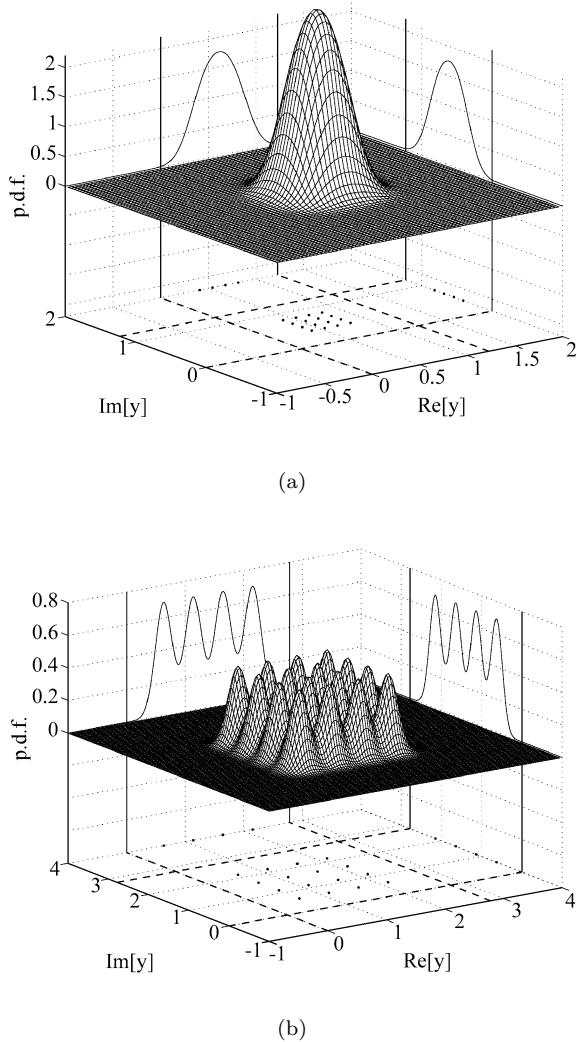
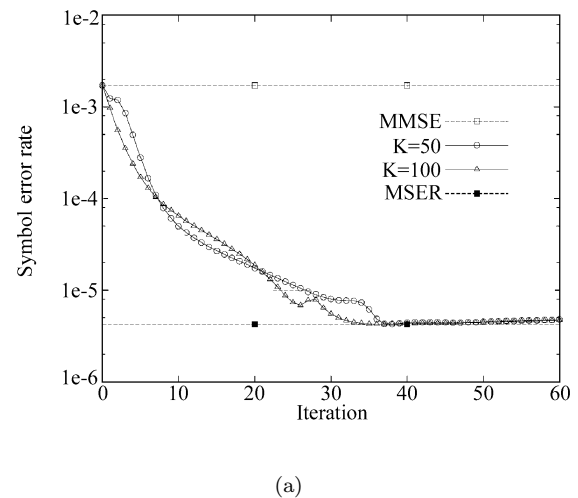


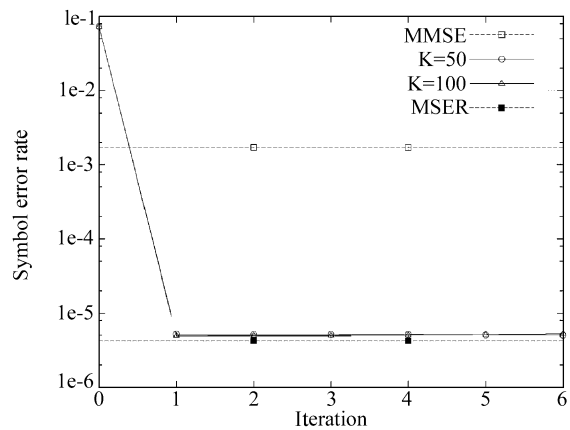
Fig. 3 Conditional probability density functions $p(y|1+j)$ (surfaces), marginal conditional probability density functions $p(y_R|1+j)$ and $p(y_I|1+j)$ (curves), signal subsets $\mathcal{Y}_{3,3}$ (dots) and their real and imaginary parts (dots) for: (a) the MMSE DFE, and (b) the MSER DFE, given SNR = 27 dB. The equalizer weight vector has been normalized to a unit length

The performance of the block-data based adaptive MSER algorithm with the simplified conjugate gradient updating mechanism was next studied. A perfect estimate $\hat{\mathbf{p}}_d$ was assumed and the step size μ and the kernel width ρ_n were determined empirically to provide the best performance in terms of convergence speed and estimation accuracy. Fig. 4 illustrates the convergence rate of the algorithm under SNR = 27 dB and given two different initial weight vector conditions, where two block sizes $K = 50$ and 100 were used. From Fig. 4, it can be seen that the convergence speed of this block-data adaptive MSER algorithm is rapid. As the SER estimate $\hat{P}_{E_B}(\mathbf{w})$ is a complicated nonlinear function

of \mathbf{w} , the initial condition affects convergence speed. For this example, with $\mathbf{w}(0)$ chosen arbitrarily to be $[0.0 + j0.0 \ 0.1 - j0.1]^T$, it took only one iteration to converge, compared this with about 40 iterations that was needed with $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$. The performance of the stochastic gradient adaptive MSER algorithm was then investigated. Fig. 5 shows the learning curves of the LSER algorithm averaged over 100 runs, given SNR = 27 dB and two different initial weight vector conditions. From Fig. 5, it can be seen that this LSER algorithm has a fast convergence rate. There are two learning curves in both Fig. 5 (a) and (b), corresponding to training and decision directed (DD) adaptation in which $b(k-d)$ is substituted by the equalizer's estimate $\hat{b}(k-d)$. It is seen that even with the DD adaptation, the LSER algorithm performs well.

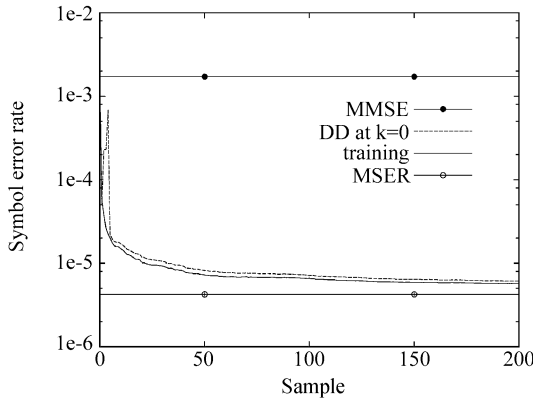


(a)

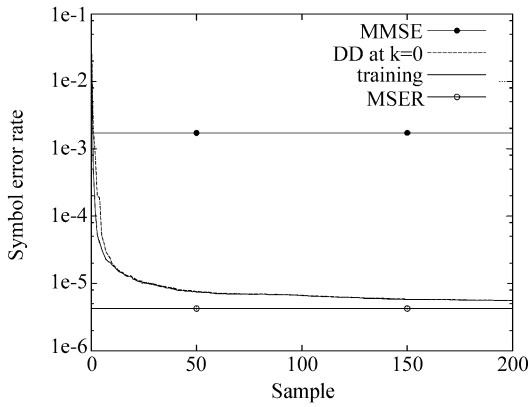


(b)

Fig. 4 Convergence rate of the block-data gradient adaptive MSER algorithm for the equalization example given SNR = 27 dB: (a) $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$, $\mu = 0.8$ and $\rho_n^2 = \sigma_n^2 \approx 0.02$, and (b) $\mathbf{w}(0) = [0.0 + j0.0 \ 0.1 - j0.1]^T$, $\mu = 0.8$ and $\rho_n^2 = 2\sigma_n^2 \approx 0.04$



(a)



(b)

Fig. 5 Learning curves of the stochastic gradient adaptive MSER algorithm averaged over 100 runs for the equalization example given $\text{SNR} = 27$ dB: (a) $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$, $\mu = 0.9$ and $\rho_n^2 = 50\sigma_n^2 \approx 1.04$, and (b) $\mathbf{w}(0) = [0.0 + j0.0 \ 0.1 - j0.1]^T$, $\mu = 0.9$ and $\rho_n^2 = 100\sigma_n^2 \approx 2.08$, where DD denotes decision directed adaptation with $\hat{b}(k-d)$ substituting $b(k-d)$

5.2 Adaptive beamforming assisted receiver

The ever-increasing demand for mobile communication capacity has motivated the employment of space division multiple access for improving the achievable spectral efficiency. A particular approach that has shown real promise in achieving substantial capacity enhancements is the use of adaptive beamforming with antenna arrays. Adaptive beamforming is capable of separating signals transmitted on the same carrier frequency, provided that they are separated sufficiently in the spatial domain. Consider the system that supports N 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 $L \times N$ system matrix \mathbf{P} is defined by

$$\mathbf{P} = [A_0 \mathbf{s}_0 \ A_1 \mathbf{s}_1 \ \cdots \ A_{N-1} \mathbf{s}_{N-1}] \quad (72)$$

with A_i denoting the channel coefficient for user i and the steering vector for source i

$$\mathbf{s}_i = [\exp(j\omega t_0(\theta_i)) \ \cdots \ \exp(j\omega t_{L-1}(\theta_i))]^T \quad (73)$$

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 $\mathbf{b}(k) = [b_0(k) \ b_1(k) \ \cdots \ b_{N-1}(k)]^T$. 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 SNR is defined as $\text{SNR} = |A_0|^2 \sigma_b^2 / 2\sigma_n^2$ and the desired signal to interference ratio (SIR) with respect to interfering user i is defined as $\text{SIR}_i = A_0^2 / A_i^2$ for $1 \leq i \leq N-1$. The beamformer at receiver is a linear filter in the form of (1) with $d = 0$ in the decision rule (7) and (8).

The simulation example consisted of three 16-QAM signal sources and a two-element antenna array. Fig. 6 shows the locations of the desired source and the interfering sources. The minimum spatial separation was the difference in angles of arrival between the desired user 0 and the interferer 2, which was $\theta \leq 70^\circ$. Fig. 7 compares the SER of the MMSE solution with that of the MSER one, given $\theta = 60^\circ$ and under two different conditions: (a) the desired user and the two interfering sources had equal power, and (b) the desired user and

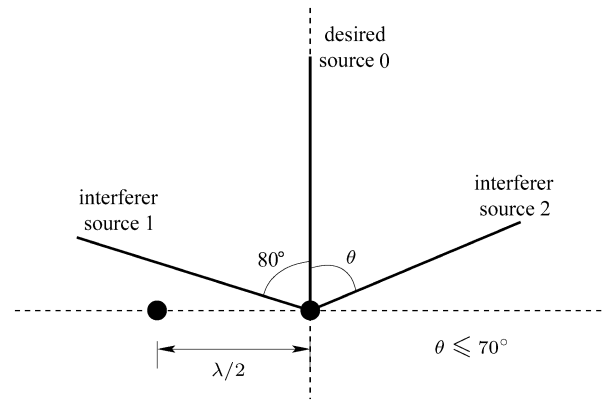


Fig. 6 Locations of the desired source and the interfering sources with respect to the two-element linear antenna array having $\lambda/2$ element spacing, where λ is the wavelength

the interfering source 1 had equal power, but the interfering source 2 had 6 dB higher power than the desired user. It can be seen that when the 2nd interfering user's power was increased by 6 dB, the MMSE beamformer's performance degraded considerably while the performance of the MSER beamformer was only affected slightly. Thus, the MSER beamformer is robust to the near-far effect. Figs. 8 and 9 depict the conditional p.d.f. $p(y|1+j)$, the two marginal conditional p.d.f.'s $p(y_R|1+j)$ and $p(y_I|1+j)$, the subset $\mathcal{Y}_{3,3}$ and its real and imaginary parts for the MMSE and MSER beamformers, given $\theta = 60^\circ$, SNR = 27 dB, and two SIR conditions, respectively. The total number of signal points was $N_b = 16^3 = 4096$ and the subset $\mathcal{Y}_{3,3}$ contained $N_{sb} = 256$ points. It is seen from Fig. 8 that in the equal user power case, the minimum distance from $\mathcal{Y}_{3,3}$ to its corresponding decision boundaries for the MSER solution was slightly larger than that for the MMSE solution. This explains the slightly better SER performance of the MSER beamformer over the MMSE one as shown in Fig. 7 with $SIR_1 = SIR_2 = 0$ dB. When facing a strong interference signal, Fig. 9 (a) shows that the minimum separation between $\mathcal{Y}_{l,q}$ for the MMSE solution was reduced dramatically, thus causing a significant performance degradation as seen in Fig. 7 for the case of $SIR_2 = -6$ dB. The reason of near-far robustness for the MSER solution can be seen clearly from Fig. 9 (b), which shows an almost unchanged minimum separation between $\mathcal{Y}_{l,q}$ compared with the equal user power case.

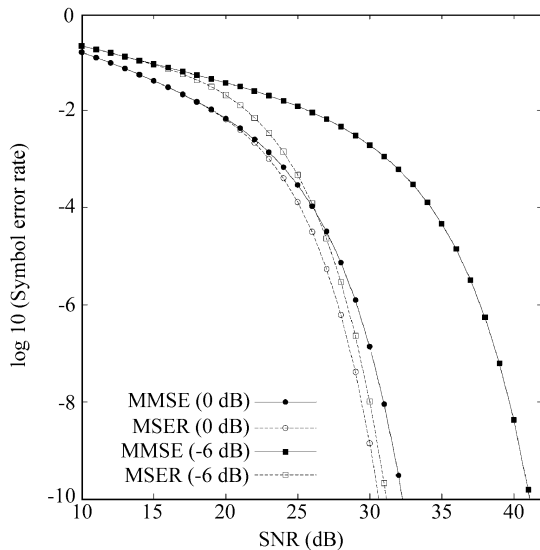
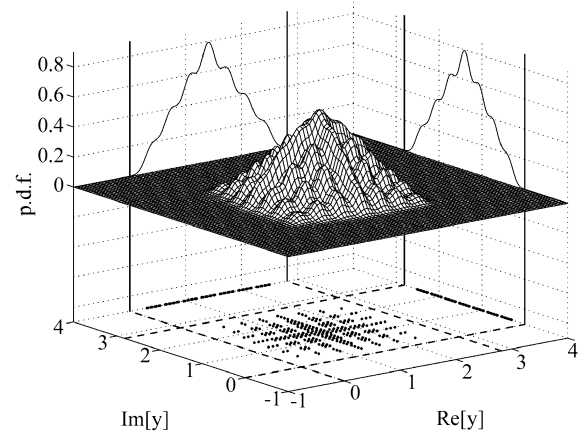
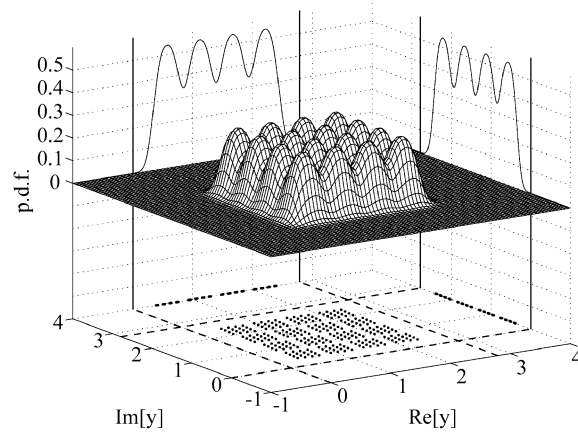


Fig. 7 Symbol error rate performance comparison of the MMSE and MSER beamformers, given $\theta = 60^\circ$, $SIR_1 = 0$ dB, and two SIR_2 (0 dB and -6 dB)

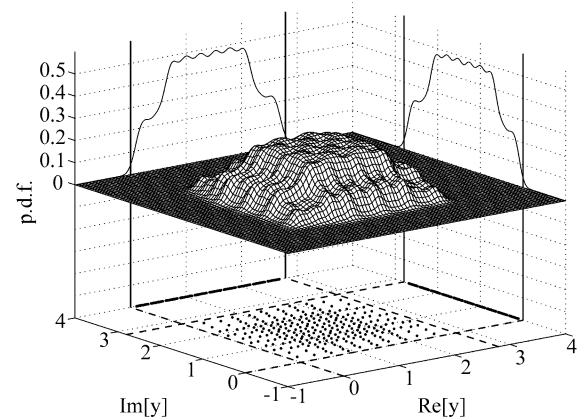


(a)



(b)

Fig. 8 Conditional probability density functions $p(y|1+j)$ (surfaces), marginal conditional probability density functions $p(y_R|1+j)$ and $p(y_I|1+j)$ (curves), signal subsets $\mathcal{Y}_{3,3}$ (dots) and their real and imaginary parts (dots) for: (a) the MMSE beamformer, and (b) the MSER beamformer, given $\theta = 60^\circ$, SNR = 27 dB and $SIR_1 = SIR_2 = 0$ dB. The beamformer weight vector has been normalized to a unit length



(a)

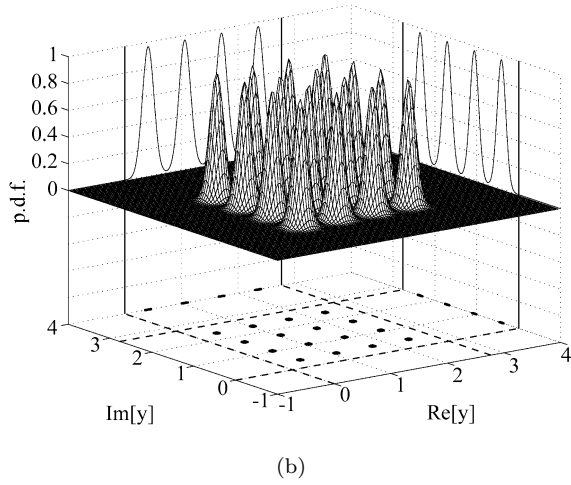
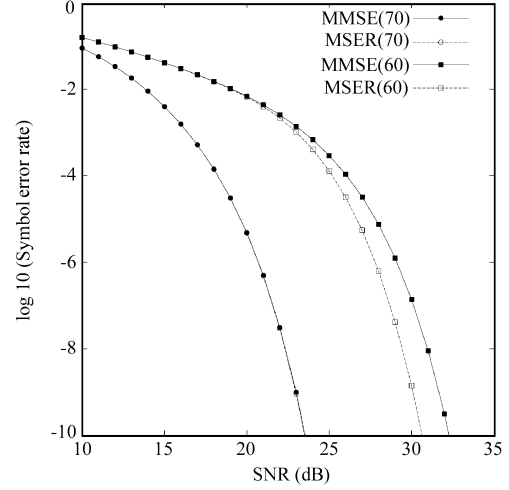


Fig. 9 Conditional probability density functions $p(y|+1+j)$ (surfaces), marginal conditional probability density functions $p(y_R|+1+j)$ and $p(y_I|+1+j)$ (curves), signal subsets $\mathcal{Y}_{3,3}$ (dots) and their real and imaginary parts (dots) for: (a) the MMSE beamformer, and (b) the MSER beamformer, given $\theta = 60^\circ$, SNR = 27 dB, $SIR_1 = 0$ dB and $SIR_2 = -6$ dB. The beamformer weight vector has been normalized to a unit length

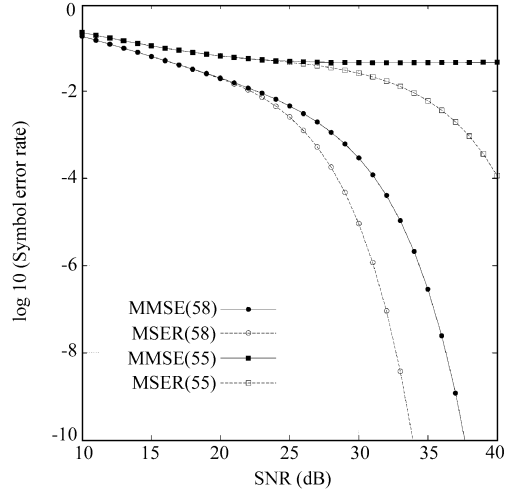
The performance of the two beamformers was also investigated under the equal user power condition. The varying minimum spatial separation θ , and the results are depicted in Fig. 10. For $\theta = 70^\circ$, the performances of the two beamformers were indistinguishable. When θ was reduced to 60° and 58° , the MSER beamformer achieved above 1 dB and 3 dB improvements in SNR, respectively, at the SER level of 10^{-4} over the MMSE solution. For $\theta = 55^\circ$, the MMSE beamformer could not achieve linear separability and exhibited an irreducible SER floor, while the MSER beamformer could still manage to achieve a linear separability. When the minimum spatial separation was below 55° , the system was inherently linearly inseparable.

Performance of the block-data gradient adaptive MSER algorithm was next tested. Again a perfect estimate $\hat{\mathbf{p}}_d$ was assumed and the step size μ and the kernel width ρ_n were found empirically to provide the best performance in terms of convergence speed and estimation accuracy. Fig. 11 illustrates the convergence rates of the algorithm given SNR = 27 dB, $SIR_1 = 0$ dB and $SIR_2 = -6$ dB, and with the two different initial weight vectors. It can be seen that this block-data adaptive MSER algorithm converges rapidly. Performance of the stochastic gradient adaptive MSER algorithm was also investigated. Fig. 12 shows the learning curves of the algorithm averaged over 100 runs, under the same conditions of Fig. 11, where DD denotes decision-directed adaptation with $\hat{b}_0(k)$ substituting $b_0(k)$ as the desired response. It can be seen that once the SER is below certain level (0.01 for this example), DD adaptation

can be applied.

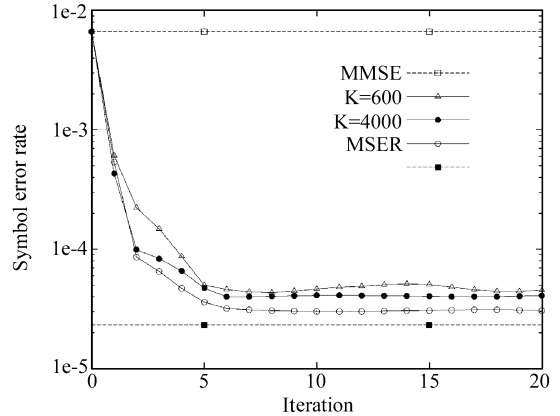


(a)



(b)

Fig. 10 Symbol error rate performance comparison of the MMSE and MSER beamformers given $SIR_1 = SIR_2 = 0$ dB: (a) $\theta = 70^\circ$ and 60° , and (b) $\theta = 58^\circ$ and 55°



(a)

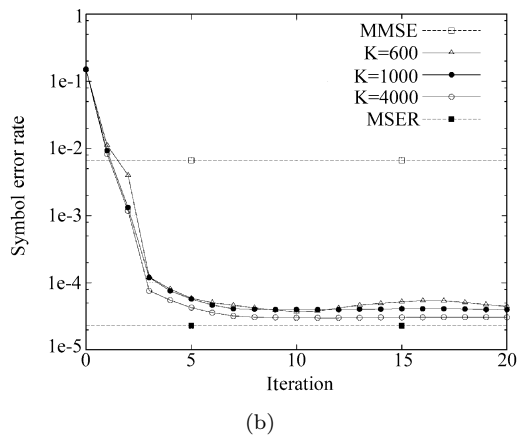


Fig. 11 Convergence rate of the block-data gradient adaptive MSER algorithm for the beamforming example given SNR = 27 dB, SIR₁ = 0 dB and SIR₂ = -6 dB:

- (a) $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$, $\mu = 0.05$ and $\rho_n^2 = \sigma_n^2 \approx 0.01$,
and (b) $\mathbf{w}(0) = [0.7 - j0.1 \ 0.6 + j0.1]^T$,
 $\mu = 0.05$ and $\rho_n^2 = \sigma_n^2 \approx 0.01$

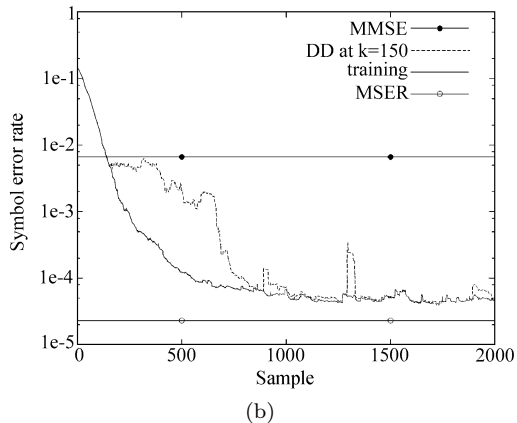
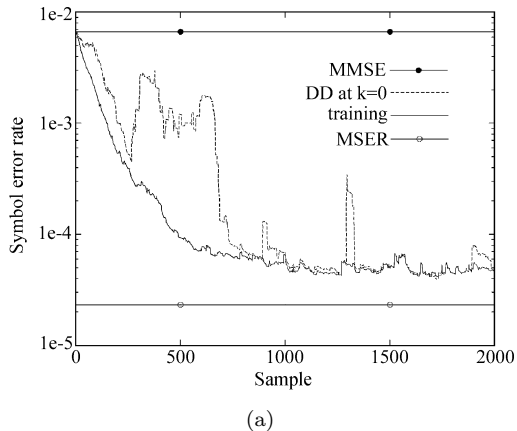


Fig. 12 Learning curves of the stochastic gradient adaptive MSER algorithm averaged over 100 runs for the beamforming example given SNR = 27 dB, SIR₁ = 0 dB and SIR₂ = -6 dB: (a) $\mathbf{w}(0) = \mathbf{w}_{\text{MMSE}}$, $\mu = 0.0005$ and $\rho_n^2 = \sigma_n^2 \approx 0.01$, and (b) $\mathbf{w}(0) = [0.7 - j0.1 \ 0.6 + j0.1]^T$, $\mu = 0.0005$ and $\rho_n^2 = \sigma_n^2 \approx 0.01$, where DD denotes decision directed adaptation with $\hat{b}_0(k)$ substituting $b_0(k)$

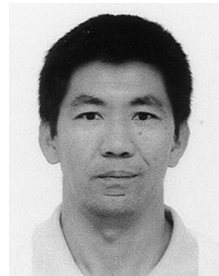
6 Conclusions

An adaptive linear filtering technique based on the novel MSER principle has been proposed for applications to communication systems with complex-valued filters and M -QAM signalling. It has been demonstrated that the MSER filtering is capable of achieving significant performance gains in terms of reduced SER over the traditional MMSE filtering. This is due to the fact that the MSER filtering can exploit effectively the non-Gaussian nature of the underlying system density distribution. Adaptive implementation of the proposed MSER filtering has been developed based on the Parzen window estimation for the p.d.f. of the filter's output. A block-data based simplified conjugate gradient adaptive MSER algorithm has been shown to converge rapidly and requires a reasonably small data block size to accurately approximate the theoretical MSER solution. An LMS-style stochastic gradient adaptive MSER algorithm, referred to as the LSER, has been shown to perform well, and the algorithm has similar computational requirements to the low-complexity LMS algorithm.

References

- [1] B. Widrow, S. D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1985.
- [2] S. Haykin. *Adaptive Filter Theory*. 3rd edition, Prentice-Hall, Upper Saddle River, NJ, 1996.
- [3] E. Shamash, K. Yao. On the Structure and Performance of a Linear Decision Feedback Equalizer Based on the Minimum Error Probability Criterion. In *Proceedings of International Communications Conference*, pp. 25F1–25F5, 1974.
- [4] S. Chen, E. S. Chng, B. Mulgrew, G. Gibson. Minimum-BER Linear-combiner DFE. In *Proceedings of International Communications Conference*, Dallas, Texas, vol. 2, pp. 1173–1177, 1996.
- [5] C. C. Yeh, J. R. Barry. Approximate Minimum Bit-Error Rate Equalization for Binary Signaling. In *Proceedings of International Communications Conference*, Montreal, Canada, vol. 2, pp. 1095–1099, 1997.
- [6] S. Chen, B. Mulgrew, E. S. Chng, G. Gibson. Space Translation Properties and the Minimum-BER Linear-combiner DFE. *IEEE Proceedings of Communications*, vol. 145, no. 5, pp. 316–322, 1998.
- [7] B. Mulgrew, S. Chen. Stochastic Gradient Minimum-BER Decision Feedback Equalisers. In *Proceedings of IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control*, Lake Louise, Alberta, Canada, Oct. 1-4, pp. 93–98, 2000.
- [8] C. C. Yeh, J. R. Barry. Adaptive Minimum Bit-Error Rate Equalization for Binary Signaling. *IEEE Transactions Communications*, vol. 48, no. 7, pp. 1226–1235, 2000.
- [9] B. Mulgrew, S. Chen. Adaptive Minimum-BER Decision Feedback Equalisers for Binary Signalling. *Signal Processing*, vol. 81, no. 7, pp. 1479–1489, 2001.
- [10] N. B. Mandayam, B. Aazhang. Gradient Estimation for Sensitivity Analysis and Adaptive Multiuser Interference Rejection in Code-division Multi-access Systems. *IEEE Transactions on Communications*, vol. 45, no. 7, pp. 848–858, 1997.

- [11] C. C. Yeh, R. R. Lopes, J. R. Barry. Approximate Minimum Bit-Error Rate Multiuser Detection. In *Proceedings of Globecom'98*, Sydney, Australia, pp. 3590–3595, 1998.
- [12] X. F. Wang, W. S. Lu, A. Antoniou. Constrained Minimum-BER Multiuser Detection. In *Proceedings of International Conference on Acoustics Speech and Signal Processing*, Phoenix, USA, vol. 5, pp. 2603–2606, May 14–18, 1999.
- [13] I. N. Psaromiligkos, S. N. Batalama, D. A. Pados. On Adaptive Minimum Probability of Error Linear Filter Receivers for DS-CDMA Channels. *IEEE Transactions Communications*, vol. 47, no. 7, pp. 1092–1102, 1999.
- [14] S. Chen, A. K. Samangan, B. Mulgrew, L. Hanzo. Adaptive Minimum-BER Linear Multiuser Detection. In *Proceedings of International Conference on Acoustics Speech and Signal Processing*. Salt Lake City, Utah, USA, vol.4, pp. 2253–2256, May 7–11, 2001.
- [15] S. Chen, A. K. Samangan, B. Mulgrew, L. Hanzo. Adaptive Minimum-BER Linear Multiuser Detection for DS-CDMA Signals in Multipath Channels. *IEEE Transactions on Signal Processing*, vol. 49, no. 6, 1240–1247, 2001.
- [16] R. C. de Lamare, R. Sampaio-Neto. Adaptive MBER Decision Feedback Multiuser Receivers in Frequency Selective Fading Channels. *IEEE Communications Letters*, vol. 7, no. 2, pp. 73–75, 2003.
- [17] S. Chen, N. N. Ahmad, L. Hanzo. Smart Beamforming for Wireless Communications: A Novel Minimum Bit-error Rate Approach. In *Proceedings of 2nd IMA International Conference Mathematics in Communications*, Lancaster, U.K., Dec. 16–18, 2002.
- [18] S. Chen, L. Hanzo, N. N. Ahmad. Adaptive Minimum Bit-error Rate Beamforming Assisted Receiver for Wireless Communications. In *Proceedings of International Conference on Acoustics Speech and Signal Processing 2003*, Hong Kong, China, vol. IV, pp. 640–643, April 6–10, 2003.
- [19] S. Chen, N. N. Ahmad, L. Hanzo. Adaptive Minimum Bit-Error Rate Beamforming. *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 341–348, 2005.
- [20] J. G. Proakis. *Digital Communications*. 3rd ed., McGraw-Hill, New York, 1995.
- [21] E. Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, vol. 33, pp. 1066–1076, 1962.
- [22] B. W. Silverman. *Density Estimation*. Chapman Hall, London, 1996.
- [23] A. W. Bowman, A. Azzalini. *Applied Smoothing Techniques for Data Analysis*. Oxford University Press, Oxford, UK, 1997.
- [24] R. Johnson, Jr. P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, R. A. Casas. Blind Equalization Using the Constant Modulus Criterion: A Review. *Proceedings of IEEE*, vol. 86, no. 10, pp. 1927–1950, 1998.
- [25] S. Chen, B. Mulgrew, P. M. Grant. A Clustering Technique for Digital Communications Channel Equalisation Using Radial Basis Function Networks. *IEEE Transactions on Neural Networks*, vol. 4, no. 4, pp. 570–579, 1993.
- [26] S. Chen, B. Mulgrew, S. McLaughlin. Adaptive Bayesian Equaliser with Decision Feedback. *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2918–2927, 1993.
- [27] S. Chen, S. McLaughlin, B. Mulgrew, P. M. Grant. Bayesian Decision Feedback Equaliser for Overcoming Co-Channel Interference. *IEE Proceedings of Communications*, vol. 143, no. 4, pp. 219–225, 1996.
- [28] S. Chen, A. K. Samangan, L. Hanzo. Support Vector Machine Multiuser Receiver for DS-CDMA Signals in Multipath Channels. *IEEE Transactions on Neural Networks*, vol. 12, no. 3, pp. 604–611, 2001.
- [29] S. Chen, B. Mulgrew, L. Hanzo. Adaptive Least Error Rate Algorithm for Neural Network Classifier. In *Proceedings of 2001 IEEE Workshop Neural Networks for Signal Processing*, Falmouth, MA, USA, pp. 223–232, Sept. 10–12, 2001.
- [30] S. Chen, B. Mulgrew, L. Hanzo. Least Bit-Error Rate Adaptive Nonlinear Equalizers for Binary Signalling. *IEE Proceedings of Communications*, vol. 150, no. 1, pp. 29–36, 2003.
- [31] S. Chen, L. Hanzo, A. Wolfgang. Nonlinear Multi-Antenna Detection Methods. *EURASIP Journal of Applied Signal Processing*, vol. 2004, no. 9, pp. 1225–1237, 2004.
- [32] M. S. Bazaraa, H. D. Sherali, C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley, New York, 1993.



Sheng Chen obtained a B.Eng degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, and a Ph.D. degree in control engineering from the City University at London in 1986. He joined the School of Electronics and Computer Science at the University of Southampton in September 1999. He previously held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth.

He has published over 240 research papers. His recent research interests include adaptive nonlinear signal processing, wireless communications, modeling and identification of nonlinear systems, neural networks and machine learning, finite-precision digital controller design, evolutionary computation methods and optimization.

Dr Chen is a Senior Member of IEEE.