Adaptive Minimum Symbol-Error-Rate CDMA Linear Multiuser Detector for Pulse-Amplitude Modulation

A.K. Samingan, S. Chen and L. Hanzo

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

Abstract— **An adaptive minimum symbol error rate (MSER) linear multiuser detector (MUD) is proposed for direct sequence code division multiple access (DS-CDMA) systems employing multilevel pulse-amplitude modulation (M-PAM) scheme. Based on a kernel density estimation for approximating the symbol error rate (SER) from training data, a least mean squares (LMS) style stochastic gradient algorithm called the least SER (LSER) is developed for training linear MUDs. Computer simulation is used to investigate the performance of this LSER MUD.**

I. Introduction

DS-CDMA [1] constitutes an attractive multiuser scheme that allows users to transmit at the same carrier frequency in an uncoordinated manner. However, this creates multiuser interference (MUI) which, if not controlled, can seriously degrade the quality of reception. A variety of MUDs have been proposed to combat MUI [2]–[9]. Multi bits per symbol modulation schemes can better utilize precise bandwidth. In this paper, we consider DS-CDMA systems which employ M-PAM modulation scheme. Within the class of linear MUDs, the MMSE detector [4],[9] is popular, as it often performs adequately and has simple adaptive implementation. However, it is well known that in general the MMSE solution is not the minimum bit error rate (MBER) solution.

For a binary modulation scheme, it has been demonstrated that a linear detector that minimizes the bit error rate (BER) can outperform the MMSE detector considerably at least in certain situations and adaptive LMS style MBER algorithms can be derived [10]–[12]. The stochastic gradient algorithm of [10] is referred to as the approximate MBER (AMBER), while the stochastic gradient algorithm of [11],[12] is called the least BER (LBER). The AMBER algorithm has simpler computational requirement but the LBER algorithm is known to have better performance in terms of convergence speed and steady state BER misadjustment [11],[12].

In this paper, we extend the LBER MUD to DS-CDMA systems with M-PAM symbol constellation. We first derive the SER of the linear detector for a known system. The MSER linear detector can be obtained by minimizing the SER. Using a kernel density approximation of the SER from training data, a stochastic gradient algorithm call the LSER is developed for training linear detector. Three versions of the LSER algorithm are presented, each in turn having simpler complexity. The simplest version has a complexity similar to that of the very simple AMBER algorithm. The AMBER algorithm can also be extended to the M-PAM case, and we refer to the resulting algorithm as the AMSER. In simulation, we compare the performance of the LSER linear detector with that of the AMSER one.

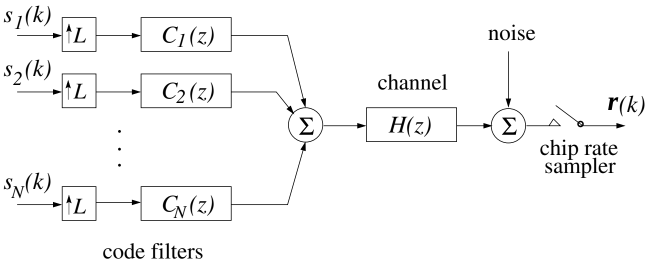

Fig. 1. Discrete-time model of down-link synchronous CDMA.

II. System model

The down-link synchronous DS-CDMA system with N users and L chips per symbol is depicted in Fig. 1, where $s_i(k)$ denotes the k-th symbol of user i, which is assumed to take the value from the symbol set

$$\mathcal{S} = \{s_l = 2l - M - 1,\ 1 \le l \le M\}\,, \tag{1}$$

the unit-length signature sequence for user i is $\bar{\mathbf{c}}_i = [\bar{c}_{i,1} \ \cdots \ \bar{c}_{i,L}]^T$, and the channel impulse response (CIR) is

$$H(z) = \sum_{i=0}^{n_h-1} h_i z^{-i}\,. \tag{2}$$

The received signal sampled at chip rate is given by:

$$\mathbf{r}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k) = \mathbf{P} \begin{bmatrix} \mathbf{s}(k) \\ \mathbf{s}(k-1) \\ \vdots \\ \mathbf{s}(k-D+1) \end{bmatrix} + \mathbf{n}(k)\,, \tag{3}$$

where $\mathbf{s}(k) = [s_1(k) \cdots s_N(k)]^T$ is the symbol vector of N users at k, D is the intersymbol interference span which de-

VTC'2003, Jeju Island, Korea
The support of the European Union and that of the EPSRC, UK is gratefull acknowledged.

pends on n_h and L, $\mathbf{n}(k) = [n_1(k) \cdots n_L(k)]^T$ is the Gaussian noise sample vector with $E[\mathbf{n}(k)\mathbf{n}^T(k)] = \sigma_n^2 \mathbf{I}$, and the $L \times DN$ system matrix $\mathbf{P}$ has the form

$$\mathbf{P} = \mathbf{H} \begin{bmatrix} \bar{\mathbf{C}}\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{C}}\mathbf{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{C}}\mathbf{A} \end{bmatrix} \tag{4}$$

with the $L \times DL$ CIR matrix $\mathbf{H}$ given by

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} & & & & \\ & h_0 & h_1 & \cdots & h_{n_h-1} & & & \\ & & \ddots & \ddots & \cdots & \ddots & & \\ & & & h_0 & h_1 & \cdots & h_{n_h-1} \end{bmatrix}, \tag{5}$$

the user code matrix given by $\bar{\mathbf{C}} = [\bar{\mathbf{c}}_1 \cdots \bar{\mathbf{c}}_N]$ and the diagonal user signal amplitude matrix $\mathbf{A} = \mathrm{diag}\{A_1 \cdots A_N\}$.

The linear detector for user i takes the form:

$$y(k) = \mathbf{w}^T \mathbf{r}(k) = \bar{y}(k) + e(k) \tag{6}$$

where $\mathbf{w} = [w_1 \cdots w_L]^T$ is the detector weight vector, $e(k)$ is Gaussian with zero mean and variance $\mathbf{w}^T\mathbf{w}\sigma_n^2$, and $\bar{y}(k) = \mathbf{w}^T \bar{\mathbf{r}}(k)$. Denote the $N_s = M^{DN}$ possible combinations of $[\mathbf{s}^T(k)\ \mathbf{s}^T(k-1) \cdots \mathbf{s}^T(k-D+1)]^T$ as $\mathbf{s}_j$, $1 \le j \le N_s$. The noise-free received signal vector $\bar{\mathbf{r}}(k)$ only takes value from the set $\mathcal{R} = \{\mathbf{r}_j = \mathbf{P}\mathbf{s}_j,\ 1 \le j \le N_s\}$, which can be divided into M subset depending on the value of $s_i(k)$:

$$\mathcal{R}_l = \{\mathbf{r}_j^{(l)} \in \mathcal{R} \mid s_i(k) = s_l\},\ 1 \le l \le M\,. \tag{7}$$

Similarly, $\bar{y}(k)$ can only take value from the set $\mathcal{Y} = \{y_j = \mathbf{w}^T\mathbf{P}\mathbf{s}_j,\ 1 \le j \le N_s\}$, which can be partitioned into M subset depending on the value of $s_i(k)$:

$$\mathcal{Y}_l = \{y_j^{(l)} \in \mathcal{Y} \mid s_i(k) = s_l\},\ 1 \le l \le M\,. \tag{8}$$

Let

$$\mathbf{g}^T = \mathbf{w}^T\mathbf{P} = \mathbf{w}^T[\mathbf{p}_1\ \mathbf{p}_2 \cdots \mathbf{p}_{DN}] = [g_1 \cdots g_{DN}]\,. \tag{9}$$

Then

$$\bar{y}(k) = \sum_{d=0}^{D-1} \sum_{j=1}^{N} g_{j+dN} s_j(k-d)\,. \tag{10}$$

Among the DN terms in (10), the term $g_i s_i(k)$ is the desired signal and the rest of the $DN-1$ terms are residual interferences. Thus, the decision thresholds are $0, \pm 2g_i, \cdots, \pm(M-2)g_i$, and the decision is made according to:

$$\hat{s}_i(k) = \begin{cases} s_1, & y(k) \le (s_1+1)g_i \\ s_l, & (s_l-1)g_i < y(k) \le (s_l+1)g_i, \\ & \quad \text{for } l = 2, \cdots, M-1, \\ s_M, & y(k) > (s_M-1)g_i. \end{cases} \tag{11}$$

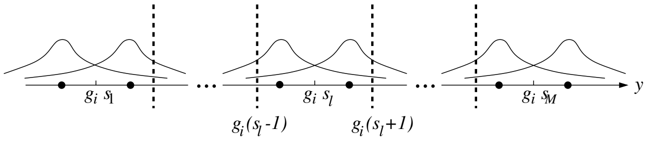

Fig. 2. Illustration of the symmetric distribution of $\mathcal{Y}_l$ around $g_i s_l$.

III. The MSER linear detector

It is straightforward to see that $\mathcal{Y}_{l+1} = \mathcal{Y}_l + 2g_i$, for $l = 1, \cdots, M-1$. We will assume that $\mathcal{Y}_{l+1}$ and $\mathcal{Y}_l$ are linearly separable. This assumption is necessary for a linear detector to work. We further point out that the distribution of $\mathcal{Y}_l$ is symmetric around the symbol point $g_i s_l$, as illustrated in Fig. 2. These properties allow us to consider just one subset $\mathcal{Y}_l$ in the derivation of the SER for the linear detector (6). The conditional probability density function (p.d.f.) of $y(k)$ given $s_i(k) = s_l$ is

$$p_y(y_s|s_l) = \frac{1}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \sum_{j=1}^{N_{sb}} \exp\left(\frac{-(y_s - y_j^{(l)})^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{12}$$

where $N_{sb} = N_s/M$ is the number of points in $\mathcal{Y}_l$. Following [13], it can be shown that the SER of the user i linear detector with weight vector $\mathbf{w}$ is

$$P_E(\mathbf{w}) = \gamma \frac{1}{N_{sb}} \sum_{j=1}^{N_{sb}} Q\left(f_{l,j}(\mathbf{w})\right) \tag{13}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(\frac{-y^2}{2}\right) dy\,, \tag{14}$$

$$f_{l,j}(\mathbf{w}) = \frac{y_j^{(l)} - g_i(s_l-1)}{\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\,, \tag{15}$$

and $\gamma = (2M-2)/M$. The gradient of $P_E(\mathbf{w})$ with respect to $\mathbf{w}$ is

$$\nabla P_E(\mathbf{w}) = \frac{\gamma}{N_{sb}\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}} \times \sum_{j=1}^{N_{sb}} \exp\left(\frac{-(y_j^{(l)} - g_i(s_l-1))^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \times \left(\frac{(y_j^{(l)} - g_i(s_l-1))}{\mathbf{w}^T\mathbf{w}}\mathbf{w} - \mathbf{r}_j^{(l)} + \mathbf{p}_i(s_l-1)\right). \tag{16}$$

A steepest-descent or simplified conjugate gradient algorithm [12], [13] can be used to minimize the SER expression (13) to arrive at the MSER solution. It is computationally advantageous to normalize $\mathbf{w}$ to a unit-length after each iteration as $\mathbf{w} = \mathbf{w}/\sqrt{\mathbf{w}^T\mathbf{w}}$. Computational requirements can further be simplified by considering the subset $\mathcal{Y}_l$ with $l = 1 + M/2$, which results in $s_l - 1 = 0$.

IV. Adaptive MSER linear detector

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

$$p_y(y_s) = \frac{1}{N_s\sqrt{2\pi}\sigma_n\sqrt{\mathbf{w}^T\mathbf{w}}}\sum_{l=1}^{M}\sum_{j=1}^{N_{sb}}\exp\left(\frac{-(y_s - y_j^{(l)})^2}{2\sigma_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{17}$$

and express the SER alternatively as

$$P_E(\mathbf{w}) = \gamma\frac{1}{N_s}\sum_{l=1}^{M}\sum_{j=1}^{N_{sb}} Q\left(f_{l,j}(\mathbf{w})\right) . \tag{18}$$

As the p.d.f. of $y(k)$ is unknown, a kernel density estimate [14] can be constructed from data samples. Given a block of K training samples $\{\mathbf{r}(k), s_i(k)\}$, a kernel density estimate of the p.d.f. (17) is given by

$$\hat{p}_y(y_s) = \frac{1}{K\sqrt{2\pi}\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}}\sum_{k=1}^{K}\exp\left(\frac{-(y_s - y(k))^2}{2\rho_n^2\mathbf{w}^T\mathbf{w}}\right) \tag{19}$$

where the radius parameter ρ_n is related to σ_n. from this estimated p.d.f., the estimated SER is given by

$$\hat{P}_E(\mathbf{w}) = \gamma\frac{1}{K}\sum_{k=1}^{K} Q(\hat{f}_k(\mathbf{w})) \tag{20}$$

where

$$\hat{f}_k(\mathbf{w}) = \frac{y(k) - \hat{g}_i(s_i(k) - 1)}{\rho_n\sqrt{\mathbf{w}^T\mathbf{w}}} , \tag{21}$$

$\hat{g}_i = \mathbf{w}^T\hat{\mathbf{p}}_i$, and $\hat{\mathbf{p}}_i$ an estimate of $\mathbf{p}_i$. With $\nabla\hat{P}_E(\mathbf{w})$, the block-data adaptive steepest-descent or conjugate gradient algorithm can be derived.

Using a single-sample estimate of $p_y(y_s)$, a sample-by-sample stochastic gradient adaptive algorithm can be derived, which is referred to as the LSER. Three versions of the LSER algorithm are presented, each involving a different level of approximation.

Version 1. Notice that $\mathbf{p}_i$ is the convolution of $\mathbf{H}$ with the code vector $\mathbf{c}_i = \bar{\mathbf{c}}_i A_i$. Assume that a separate channel estimate $\hat{\mathbf{h}}(k) = [\hat{h}_0(k)\ \hat{h}_1(k)\cdots\hat{h}_{n_h-1}(k)]^T$ is provided, which can then be used to calculate $\hat{g}_i(k) = \mathbf{w}^T(k)\hat{\mathbf{p}}_i(k)$. Using a re-scaling after each update to ensure $\mathbf{w}^T(k)\mathbf{w}(k) = 1$, the stochastic gradient at k is

$$\nabla\hat{P}_E(\mathbf{w}(k),k) = \frac{\gamma}{\sqrt{2\pi}\rho_n}\times$$
$$\exp\left(\frac{-(y(k) - \hat{g}_i(k)(s_i(k)-1))^2}{2\rho_n^2}\right)\{(y(k) - \hat{g}_i(k)\times$$
$$(s_i(k)-1))\mathbf{w}(k) - \mathbf{r}(k) + (s_i(k)-1)\hat{\mathbf{p}}_i(k)\} . \tag{22}$$

The weight updating equation is then given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\left(-\nabla\hat{P}_E(\mathbf{w}(k),k)\right) \tag{23}$$

where μ is an adaptive step size.

Version 2. At k, use a moving average update for $\hat{g}_i(k)$

$$\hat{g}_i(k) = (1-\lambda)\hat{g}_i(k-1) + \lambda\frac{y(k)}{s_i(k)} \tag{24}$$

where $0 < \lambda < 1$ is a step size. Using a re-scaling after each update to ensure $\mathbf{w}^T(k)\mathbf{w}(k) = 1$ and further assuming that g_i does not depend on $\mathbf{w}$ leads to a simplified stochastic gradient

$$\nabla\hat{P}_E(\mathbf{w}(k),k) = \frac{\gamma}{\sqrt{2\pi}\rho_n}\times$$
$$\exp\left(\frac{-(y(k) - \hat{g}_i(k)(s_i(k)-1))^2}{2\rho_n^2}\right)\{(y(k) - \hat{g}_i(k)\times$$
$$(s_i(k)-1))\mathbf{w}(k) - \mathbf{r}(k)\} . \tag{25}$$

The weight updating equation has the same form as (23).

Version 3. In this version of the LSER, further simplifications are made. Specifically, a one-sample kernel density estimation of

$$\hat{p}_y(y_s,k) = \frac{1}{\sqrt{2\pi}\rho_n}\exp\left(\frac{-(y_s - y(k))^2}{2\rho_n^2}\right) \tag{26}$$

is used, and ρ_n^2 is assumed to be independent of $\mathbf{w}$. The simplified stochastic gradient at k is thus given by

$$\nabla\hat{P}_E(\mathbf{w}(k),k) =$$
$$\frac{\gamma}{\sqrt{2\pi}\rho_n}\exp\left(\frac{-(y(k) - \hat{g}_i(k)(s_i(k)-1))^2}{2\rho_n^2}\right)\mathbf{r}(k) . \tag{27}$$

The move average (24) is used to update $\hat{g}_i(k)$. Notice that, for this version, a weight normalization is not necessary. The weight updating equation is given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) -$$
$$\frac{\mu\gamma}{\sqrt{2\pi}\rho_n}\exp\left(\frac{-(y(k) - \hat{g}_i(k)(s_i(k)-1))^2}{2\rho_n^2}\right)\mathbf{r}(k) . \tag{28}$$

For a comparison purpose, we also extend the AMBER algorithm [10] to the M-PAM case. The resulting AMSER algorithm has the form

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu I_k(\tau)\mathrm{sgn}(\hat{e}(k))\mathbf{r}(k) \tag{29}$$

where the error function $\hat{e}(k)$ is given by

$$\hat{e}(k) = s_i(k) - \frac{y(k)}{\hat{g}_i(k)} , \tag{30}$$

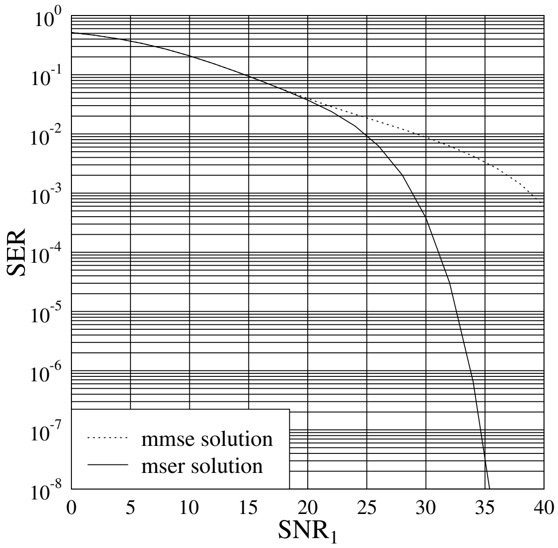

Fig. 3. Symbol error rate comparison of the MMSE and MSER linear detectors for user 1 of Example 1. $\mathrm{SNR}_1 = \mathrm{SNR}_2$.

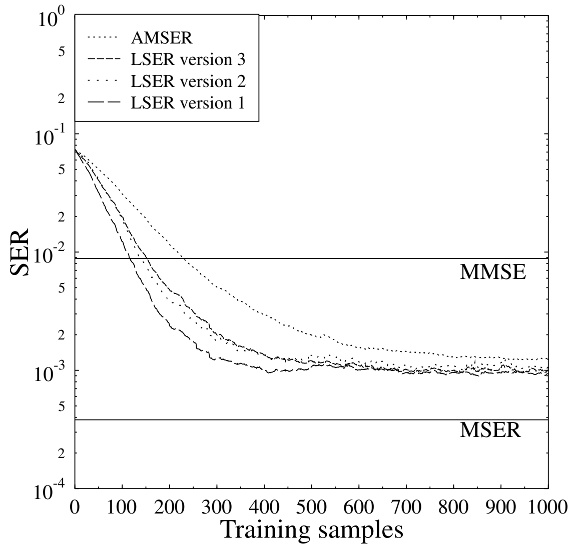

Fig. 4. Learning curves of the four stochastic gradient adaptive MSER algorithms for user 1 of Example 1, given $\mathrm{SNR}_1 = \mathrm{SNR}_2 = 30$ dB.

$\hat{g}_i(k)$ is updated using the moving average (24), and the indication function $I_k(\tau)$ is defined by:

$$I_k(\tau) = \begin{cases} 1, & \frac{y(k)}{\hat{g}_i(k)} < s_i(k) - 1 + \tau, \; s_i(k) \neq -M + 1 \,, \\ 1 & \frac{y(k)}{\hat{g}_i(k)} > s_i(k) + 1 - \tau, \; s_i(k) \neq M - 1 \,, \\ 0, & \text{otherwise} \,. \end{cases} \tag{31}$$

The adaptive gain μ and the nonnegative threshold τ are the two algorithm parameters. In terms of computational requirements, the versions 1 and 2 of the LSER are more complex than the AMSER, while the version 3 of the LSER has a similar complexity to the AMSER.

V. Simulation results

Example 1. This was a two-user system employing 4-PAM modulation with 4 chips per symbol. The two code sequences were $(+1, +1, -1, -1)$ and $(+1, -1, -1, +1)$, respectively, and the CIR at chip rate was

$$H(z) = 0.8 + 0.2z^{-3} \,. \tag{32}$$

The two users had equal signal power, that is, the user 1 signal to noise ratio SNR_1 was equal to SNR_2 of user 2. Fig. 3 compares the SER performance of the MMSE linear detector with that of the MSER one for user 1, given a range of SNR_1.

Convergence performance of the LSER and AMSER algorithms were investigated given $\mathrm{SNR}_1 = 30$ dB. For the version 1 of the LSER, the receiver was assumed to know the CIR $\mathbf{h}$, and the two algorithm parameters were chosen to be $\mu = 0.001$ and $\rho_n = 1.5\sigma_n$. For the versions 2 and 3, the same $\mu = 0.001$ and $\rho_n = 1.5\sigma_n$ were used with $\lambda = 0.02$ for moving average updating of $\hat{g}_i(k)$. Only the versions 1 and 2 employed a weight normalization after each updating. The AMSER had a same $\lambda = 0.02$ with the two algorithm parameters given by $\mu = 0.001$ and $\tau = 0.4$. Fig. 4 depicts the convergence performance of these four stochastic gradient adaptive algorithms, where the results were averaged on 100 runs. It can be seen that for this example the very simplified version 3 of the LSER algorithm was only slightly inferior to the full LSER algorithm of version 1, and it had better performance than the AMSER algorithm in terms of convergence rate and steady-state SER misadjustment.

Example 2. This was a 4-user system employing 4-PAM modulation with 8 chips per symbol. The four code sequences were $(+1, +1, +1, +1, -1, -1, -1, -1)$, $(+1, -1, +1, -1, -1, +1, -1, +1)$, $(+1, +1, -1, -1, -1, -1, +1, +1)$ and $(+1, -1, -1, +1, -1, +1, +1, -1)$, respectively, and the CIR was

$$H(z) = 0.9 + 0.2z^{-7} \,. \tag{33}$$

The 4 users had equal signal power. Fig. 5 shows the SERs of the MMSE and MSER linear detectors for user 1. Notice that in this case the MMSE detector did not result in linearly separable subsets $\mathcal{Y}_l$, $1 \leq l \leq 4$, and its SER curve exhibited a floor. However, the MSER detector produced a linearly separable result and hence had better performance.

With $\mathrm{SNR}_1 = 36$ dB, learning curves of the LSER and AMSER algorithms for user 1 are depicted in Fig. 6, where the results were averaged over 10 runs. For the version 1 of the LSER, again the receiver was assumed to know the CIR $\mathbf{h}$, and the two algorithm parameters were set to $\mu = 0.0001$ and $\rho_n = 5\sigma_n$. For the versions 2 and 3, the

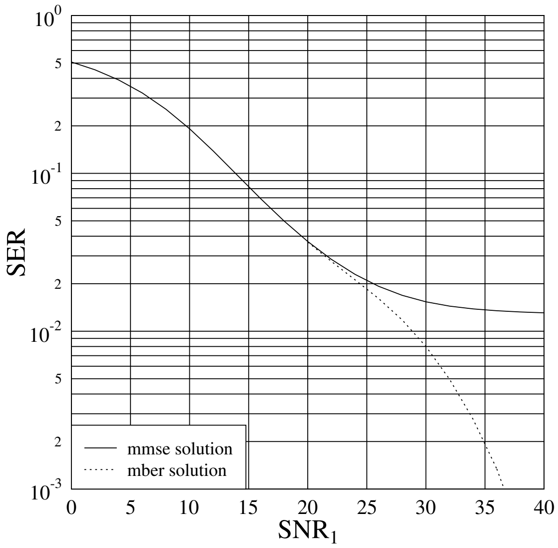

Fig. 5. Symbol error rate comparison of the MMSE and MSER linear detectors for user 1 of Example 2. SNR_i, $1 \le i \le 4$, are identical.

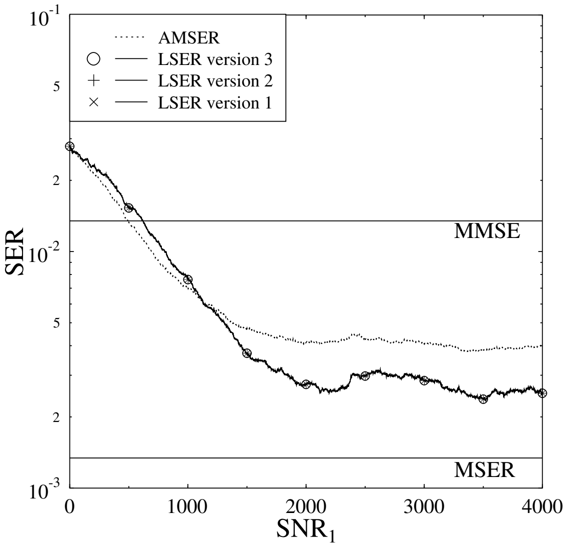

Fig. 6. Learning curves of the four stochastic gradient adaptive MSER algorithms for user 1 of Example 2, given $\text{SNR}_i = 36$ dB, $1 \le i \le 4$.

same $\mu = 0.0001$ and $\rho_n = 5\sigma_n$ together with $\lambda = 0.002$ were used. Only the versions 1 and 2 employed a weight normalization after each updating. The AMSER had $\lambda = 0.002$, $\mu = 0.0001$ and $\tau = 0.6$. It can be seen that the three versions of the LSER algorithm had similar convergence performance, which were better than that of the AMSER algorithm.

VI. Conclusions

A stochastic gradient MSER adaptive algorithm has been derived for DS-CDMA systems with M-PAM modulation scheme. The algorithm has been motivated from the kernel density estimation of the SER as a smooth function of training data, and three versions of this LSER algorithm have been presented. The most simplified version of the algorithm is computationally very simple, and a desired feature of this stochastic gradient algorithm is that the amount of the weight updating is a continuous and decreasing function of a soft distance from the decision boundary. Simulation results indicate that this adaptive LSER linear detector outperforms an existing adaptive MSER detector called the AMSER.

References

[1] R. Prasad, *CDMA for Wireless Personal Communications*. Artech House, Inc., 1996.

[2] S. Verdú, “Minimum probability of error for asynchronous Gaussian multiple-access channels,” *IEEE Trans. Information Theory*, Vol.IT-32, No.1, pp.85–96, 1986.

[3] Z. Xie, R.T. Short and C.K. Rushforth, “A family of suboptimum detectors for coherent multiuser communications,” *IEEE J. Selected Areas in Communications*, Vol.8, No.4, pp.683–690, 1990.

[4] U. Madhow and M.L. Honig, “MMSE interference suppression for direct-sequence spread-spectrum CDMA,” *IEEE Trans. Communications*, Vol.42, No.12, pp.3178–3188, 1994.

[5] S.L. Miller, “An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection,” *IEEE Trans. Communications*, Vol.43, No.2/3/4, pp. 1746–1755, 1995.

[6] S. Moshavi, “Multi-user detection for DS-CDMA communications,” *IEEE Communications Magazine*, Vol.34, No.10, pp.124–136, 1996.

[7] H.V. Poor and S. Verdú, “Probability of error in MMSE multiuser detection,” *IEEE Trans. Information Theory*, Vol.43, No.3, pp.858–871, 1997.

[8] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[9] G. Woodward and B.S. Vucetic, “Adaptive detection for DS-CDMA,” *Proc. IEEE*, Vol.86, No.7, pp.1413–1434, 1998.

[10] 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.

[11] 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.

[12] 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.

[13] S. Chen, B. Mulgrew, A.K. Samingan and L. Hanzo, “Stochastic gradient adaptive minimum symbol-error-rate equalization for pulse-amplitude modulation,” submitted to *IEEE Trans. Signal Processing*, 2001.

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