Adaptive Minimum BER Interference Suppression for DS-CDMA using Averaging Methods in Multipath Fading Channels

Rodrigo C. de Lamare e Raimundo Sampaio-Neto

***Resumo*— Neste trabalho é investigado o uso de algoritmos de mínima taxa de erro combinados com métodos de "averaging"no projeto de receptores multiusuário lineares em sistemas DS-CDMA. Os algoritmos propostos buscam minimizar a taxa de erro de bits (BER) a partir do conjunto de treinamento conhecido pelo receptor utilizando estruturas de deteção multiusuário lineares. Uma análise comparativa de receptores lineares, usando algoritmos que minimizam o erro médio quadrático, algoritmos encontrados na literatura que minimizam a BER e os algoritmos propostos é realizada. Alguns experimentos em simulação mostram que as técnicas propostas são superiores à outros algoritmos analisados e podem operar com seqüências de treinamento mais curtas.**

***Palavras-Chave*— detecção multiusuário, CDMA, supressão de interferência, algoritmos adaptativos, métodos "averaging", mínima taxa de erro.**

***Abstract*— We investigate the use of adaptive minimum bit error rate (MBER) algorithms with averaging methods in the design of linear multiuser receivers (MUD) for DS-CDMA systems. The proposed algorithms minimise the bit error rate (BER) cost function from training data using linear multiuser detection structures. A comparative analysis of linear MUDs, employing minimum mean squared error (MMSE), previously reported MBER and the proposed MBER algorithms is carried out. Simulation experiments show that the MBER techniques with averaging outperform other analysed algorithms and can operate with shorter training sequences.**

***Keywords*— multiuser detection, CDMA, interference suppression, adaptive algorithms, averaging methods, minimum BER appoach.**

I. INTRODUCTION

Adaptive linear MUDs employing the MMSE criterion provide simple adaptive implementation and an attractive trade-off between performance, complexity and the need for side information [1]-[2]. However, it is well known that the mean squared error (MSE) cost function is not optimal in digital communications applications, and the most appropriate cost function is the BER [3],[4]. The approximate minimum bit error rate (AMBER) [3] and the least bit error rate (LBER) [4] are two of the most successful and suitable algorithms for adaptive implementation. However, these MBER algorithms usually require long training sequences to converge to lower bit error rates than those achieved by the techniques that employ the MSE cost function.

Este trabalho foi financiado pelo CNPq. Os autores estão com o CETUC/PUC-RIO, 22453-900, Rio de Janeiro - Brasil Fone: +55-21-31141701 Fax: +55-21-22945748. E-mails: delamare@infolink.com.br, raimundo@cetuc.puc-rio.br

In this work we investigate MBER based algorithms that can speed up the convergence of the receiver due to averaging methods and that require shorter training data. We perform a comparative analysis of linear MUDs using these algorithms in multipath fading channels. This paper is organised as follows. Section II briefly describes the DS-CDMA system model. Stochastic gradient (SG) algorithms are presented in Sections III. Section IV and V are dedicated to the Newton based and the averaged adaptive algorithms. Section VI shows and discusses the simulation results and Section VII gives the conclusions of this work.

II. DS-CDMA SYSTEM MODEL

Let us consider the downlink of a synchronous DS-CDMA system with K users, N chips per symbol and L_p propagation paths. The signal broadcasted by the base station intended for user k has a baseband representation given by:

$$x_k(t) = A_k \sum_{i=-\infty}^{\infty} b_k(i) s_k(t - iT) \tag{1}$$

where $b_k(i) \in \{\pm 1\}$ denotes the i-th symbol for user k, the real valued spreading waveform and the amplitude associated with user k are $s_k(t)$ and A_k, respectively. The spreading waveforms are expressed by $s_k(t) = \sum_{n=1}^{N} a_k(i)\phi(t - nT_c)$, where $a_k(i) \in \{\pm 1/\sqrt{N}\}$, $\phi(t)$ is the chip waveform, T_c is the chip duration and $N = T/T_c$ is the processing gain. Assuming that the receiver is synchronised with the main path and the users forming the composite signal experiment the same channel conditions in the downlink, the coherently demodulated composite received signal is

$$r(t) = \sum_{k=1}^{K} \sum_{l=0}^{L_p-1} h_l(t) x_k(t - \tau_l) + n(t) \tag{2}$$

where $h_l(t)$ and τ_l are, respectively, the channel coefficient and the delay associated with the l-th path. Assuming that $\tau_l = lT_c$ and that the channel is constant during each symbol interval, the received signal $r(t)$ after filtering by a chip-pulse matched filter and sampled at chip rate yields the received vector

$$\mathbf{r}(i) = \mathbf{H} \begin{bmatrix} \mathbf{SA} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{SA} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{SA} \end{bmatrix} \begin{bmatrix} \mathbf{b(i)} \\ \mathbf{b(i-1)} \\ \vdots \\ \mathbf{b(i - L_s + 1)} \end{bmatrix} + \mathbf{n}(i) = $$

$$= \mathbf{s}(i) + \mathbf{n}(i) \tag{3}$$

where the Gaussian noise vector $\mathbf{n}(i) = [n_1(i) \ \dots \ n_N(i)]^T$ with $E[\mathbf{n}(k)\mathbf{n}^T(i)] = \sigma^2 \mathbf{I}$, the user bit vector is given by

$\mathbf{b}(i) = [b_1(i)\ \ldots\ b_K(i)]^T$, the user signature sequence matrix is described by $\mathbf{S} = [\mathbf{s}_1\ \ldots\ \mathbf{s}_K]$, where $\mathbf{s}_k = [a_{k,1} \ldots a_{k,N}]^T$, the diagonal user signal amplitude matrix is represented by $\mathbf{A} = diag\{A_1\ \ldots\ A_K\}$, and the $N\ \times (L_s \times N)$ matrix H is expressed by

$$\mathbf{H} = \begin{bmatrix} h_0(i) & h_1(i) & \ldots & h_{L_p-1}(i) & \ldots & 0 & \\ & & \ddots & & \ddots & \ddots & \\ & & h_0(i) & h_1(i) & & \ldots & h_{L_p-1}(i) \end{bmatrix} \quad (4)$$

The multiple access interference (MAI) is originated from the non-orthogonality between the received signature sequences. The intersymbol interference (ISI) span L_s depends on the length of the channel response, which is related to the length of the chip sequence. For $L_p = 1$, $L_s = 1$ (no ISI), for $1 < L_p \leq N, L_s = 2$, for $N < L_p \leq 2N, L_s = 3$ and so on. Consider a one-shot MUD, whose observation vector is $\mathbf{r}(i)$, the detected symbols for this MUD and user k are expressed by:

$$\hat{b}_k(i) = sgn(\mathbf{w_k}^T(i)\mathbf{r}(i)) = sgn(x_k(i)) \quad (5)$$

where $\mathbf{w}_k(i) = [w_1 \ldots w_N]^T$ is the receiver weight vector and $x_k(i)$ is the estimated symbol for user k and symbol i in a system with K users.

III. Stochastic Gradient (SG) Algorithms

A. The LMS algorithm

The adaptive solution for the linear MUD via the LMS algorithm [6] is based on the MMSE error criterion formed by the error signal $e_k(i) = d_k(i) - \mathbf{w}_k^T(i)\mathbf{r}(i)$, and is :

$$\mathbf{w}_k(i+1) = \mathbf{w}_k(i) + \mu e_k(i)\mathbf{r}(i) \quad (6)$$

where $d_k(i)$ is the desired signal for the k-th user taken from the training sequence and μ is the algorithm step size.

B. The AMBER algorithm

The AMBER [4] SG update equation for the linear MUD is expressed by:

$$\mathbf{w}_k(i+1) = \mathbf{w}_k(i) + \mu i_{d_k}(i) d_k(i)\mathbf{r}(i) \quad (7)$$

In practice, a modified error indicator function $i_{d_k}(i) = \frac{1}{2}(1 - sgn(d_k(i)x_k(i) - \tau))$ is employed, where the threshold τ is responsible for increasing the algorithm rate of convergence. This algorithm updates when an error is made and also when an error is almost made, becoming a smarter choice for updating the filter coefficients.

C. The LBER algorithm

The MUD BER depends on the distribution of the decision variable $x_k(i)$, which is a function of $\mathbf{w}_k$. The sign-adjusted decision variable for the linear receiver $x_{s_k}(i) = d_k(i)x_k(i)$ is drawn from a Gaussian mixture, described by:

$$x_{s_k}(i) = sgn(d_k(i))\left(\mathbf{w}_k^T\mathbf{s}(i) + \mathbf{w}_k^T\mathbf{n}(i)\right) \quad (8)$$

where the first term of (8) is the noise free sign-adjusted MUD output. The LBER algorithm for the linear MUD is:

$$\mathbf{w}_k(i+1) = \mathbf{w}_k(i) + \mu\frac{1}{\sqrt{2\pi}\rho}exp\left(\frac{-(x_k(i))^2}{2\rho^2}\right) sgn(d_k(i)) \times \left(\mathbf{I} - \mathbf{w}_k(i)\mathbf{w}_k^T(i)\right)\mathbf{r}(i) \quad (9)$$

where ρ the radius parameter, related to the noise standard deviation σ.

IV. Gradient-Newton Algorithms

Gradient-Newton (GN) algorithms [6] incorporate second-order statistics of input signals, increasing their convergence rate. They usually have a faster convergence rate than SG techniques, although they require a higher computational complexity. In practice, estimates of the autocorrelation matrix and the gradient vector are used to converge to the desired solution. In addition, to avoid the inversion of the autocorrelation matrix, the matrix inversion lemma is also employed. The update equation of Newton´s method is $\mathbf{w}_k(i+1) = \mathbf{w}_k(i) - \frac{1}{2}\mathbf{R}^{-1}g_{\mathbf{w_k}}(i)$, where $\mathbf{R}$ is the autocorrelation matrix of the observation vector $\mathbf{r}$ and $g_{\mathbf{w_k}}(i)$ is the gradient vector. In practice, only estimates of $\mathbf{R}$ and $g_{\mathbf{w_k}}(i)$ are available. These estimates can be applied to Newton´s formula to devise an update rule give by:

$$\mathbf{w}_k(i+1) = \mathbf{w}_k(i) - \delta\hat{\mathbf{R}}^{-1}(i)\hat{g}_{\mathbf{w_k}}(i) \quad (10)$$

The convergence factor δ is introduced to protect the algorithm from divergence, which is originated by the use of noisy estimates of $\mathbf{R}$ and $g_{\mathbf{w_k}}(i)$. To obtain an unbiased estimate of the observation matrix $\mathbf{R_u}$, we employ the following weighted sum:

$$\hat{\mathbf{R}}(i) = \alpha\mathbf{r}(i)\mathbf{r}^T(i) + (1-\alpha)\hat{\mathbf{R}}(i-1) \quad (11)$$

where α is a small factor chosen in the range $0 < \alpha \leq 0.1$ and $\mathbf{r}(i)$ is the observation vector. To avoid the required inversion of $\hat{\mathbf{R}}(i)$, we use the matrix inversion lemma:

$$[\mathbf{A}+\mathbf{BCD}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}]^{-1}\mathbf{DA}^{-1} \quad (12)$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are matrices with appropriate dimensions and $\mathbf{A}$ and $\mathbf{C}$ are non-singular. Choosing $A = (1-\alpha)\hat{\mathbf{R}}(i-1)$, $\mathbf{B} = \mathbf{D}^T = \mathbf{r}(i)$ and $\mathbf{C} = \alpha$, it can be shown that:

$$\hat{\mathbf{R}}^{-1}(i) = \frac{1}{1-\alpha}\left[\hat{\mathbf{R}}^{-1}(i-1) - \frac{\hat{\mathbf{R}}^{-1}(i-1)\mathbf{r}(i)\mathbf{r}^T(k)\hat{\mathbf{R}}^{-1}(i-1)}{\frac{1-\alpha}{\alpha} + \mathbf{r}^T(i)\hat{\mathbf{R}}^{-1}\mathbf{r}(i)}\right] \quad (13)$$

The resulting equation for the computation of $\hat{\mathbf{R}}^{-1}(i)$ is less complex to update ($O(PG^2)$) than its direct inversion ($O(PG^3)$).

A. LMS-Newton algorithm

The LMS-Newton [6] algorithm employs the error signal $e_i(k) = d_k(i) - \mathbf{w}_k(i)\mathbf{r}(i)$, which corresponds to the MMSE solution. Thus, the estimate of the gradient $\hat{g}_{\mathbf{w_k}}(i)$ is replaced by $e(i)\mathbf{r}(i)$ to yield the expression of the LMS-Newton algorithm for the linear MUD as given by:

$$\mathbf{w}_k(i+1) = \mathbf{w}_k(i) + \delta\hat{\mathbf{R}}^{-1}(i)e_k(i)\mathbf{r}(i) \quad (14)$$

where $d_i(k)$ is the desired signal taken from the training sequence, $\mathbf{r}(i)$ is the observation vector for the MUD and μ is the algorithm step size.

B. Gradient-Newton-AMBER algorithm

An approach similar to LMS-Newton can be used to devise a Gradient-Newton based algorithm that minimises a given objective function $g(\mathbf{w}_k(i))$. We chose the objective function $g(\mathbf{w}_k(i))$ used in the AMBER algorithm [4] as an approximation to an MBER function. Then, we use the observation matrix $\hat{\mathbf{R}}^{-1}(k)$ to speed up the convergence rate of the algorithm and obtain the Gradient-Newton AMBER update equation for the MUD:

$$\mathbf{w}_k(i+1) = \mathbf{w}_k(i) + \delta\hat{\mathbf{R}}^{-1}(i) i_{d_k}(i) d_k(i)\mathbf{r}(i) \qquad (15)$$

C. Gradient-Newton-LBER algorithm

An algorithm similar to the LMS-Newton can be devised employing an approach analogous to the LBER algorithm [5]. Using Newton´s update rule, the Gradient-Newton-LBER algorithm for the linear receiver is given by:

$$\mathbf{w}_k(i+1) = \mathbf{w}_k(i) + \delta\frac{1}{K\sqrt{2\pi}\rho} exp\left(\frac{-(x_k(i))^2}{2\rho^2}\right) sgn(d_k(i)) \times\hat{\mathbf{R}}^{-1}(i)\left(\mathbf{I} - \mathbf{w}_k(i)\mathbf{w}_k^T(i)\right)\mathbf{r}(i) \qquad (16)$$

Note that the LMS-Newton, GN-AMBER and GN-LBER algorithms only differ from their SG counterparts by the addition of the inverse correlation matrix $\hat{\mathbf{R}}^{-1}(i)$ in (13).

V. Averaging Methods

In this section, we describe the proposed averaged stochastic gradient (SG) and Gradient-Newton (GN) algorithms that adjust the parameters of the MUDs based on the minimisation of the MSE and the BER cost functions. We have chosen these techniques because the error surfaces of the MBER cost functions exhibit local minima and with these approaches one can control the rate of convergence by carefulluy tuning the step size. The proposed algorithms are based on the concept of accelerating the convergence by averaging [7], which allows the use of larger step sizes. Averaged versions of the algorithms can be developed by introducing the following recursion:

$$\bar{\mathbf{w}}_k(i+1) = (1-\beta)\bar{\mathbf{w}}_k(i) + \beta\mathbf{w}_k(i+1) \qquad (17)$$

where β is the averaging factor. If we consider the mean weight vector and assume that for slow adaptation $\mathbf{r}(i)\mathbf{r}^T(i)$ behaves like it ensemble average $\mathbf{R}(i) = E[\mathbf{r}(i)\mathbf{r}^T(i)]$, i. e. ergodicity, and $\mathbf{r}(i)$ is assumed to be independent from the previous $\mathbf{w}_k(i-1)$. From the results shown in Ljung [7], the use of (14) with an SG algorithm and for large i, the algorithm behaves like $\bar{\mathbf{w}}_k(i+1) = (1-\beta)\bar{\mathbf{w}}_k(i) + \beta E[\mathbf{r}(i)\mathbf{r}^T(i)]^{-1}\mathbf{z}_w(i)$, where $z_w(i)$ is the gradient vector. Note that the equation is independent of μ and depends only on β. Let $\mathbf{w}_k^*$ denote the optimum weight vector, which is the Wiener filter if we consider MMSE estimation and a filter that achieves MBER for the AMBER and the LBER approaches. $\varepsilon(i)$ denotes the weight error vector between $\mathbf{w}_k(i)$ and $\mathbf{w}_k^*$. Thus, it follows

$$\varepsilon(i) = (1-\beta)\bar{\mathbf{w}}_k(i) - (1-\beta)\bar{\mathbf{w}}_k^* - \beta\bar{\mathbf{w}}_k^* + \beta\mathbf{R}(i)^{-1}\mathbf{r}(i)\mathbf{r}^T(i)$$

$$\varepsilon(i) = (1-\beta)\varepsilon(i) - \beta\bar{\mathbf{w}}_k^* + \beta\mathbf{R}(i)^{-1}\mathbf{r}(i)\mathbf{r}^T(i) \qquad (18)$$

Taking the expectation on both sides of (18), it was shown in [8] that $\mathbf{w}_k(i)$ converges to $\mathbf{w}_k^*$, i.e.,$E[\mathbf{w}_k(i)] \longrightarrow \mathbf{w}_k^*$, as $i \to \infty$, $\varepsilon(i) \to 0$ and it does not depend on the eigenvalues of $\mathbf{R}$. For the GN techniques the simulations verify that the averaging procedure can also improve convergence.

VI. Simulations

The performance of the MUDs with the adaptive algorithms was evaluated in a DS-CDMA system that employs Gold sequences of length $N = 15$ and whose carrier frequency is 1900 MHz. The sequence of channel coefficients $h_l(i) = p_l|\alpha_l(i)|$ $(l = 0,1,2)$, where $\alpha_l(i)$, $l = 0,1,2$, is a complex Gaussian random sequence obtained by passing complex white Gaussian noise through a filter with approximate transfer function $\gamma/\sqrt{1-(f/f_d)^2}$ where γ is a normalization constant, $f_d = v/\lambda$ is the maximum Doppler shift, λ is the wavelength of the carrier frequency, and v is the speed of the mobile. The channel parameters are $p_0 = 1$, $p_1 = 0.5$ and $p_2 = 0.3$. In all situations, the MUDs operate with Gold sequences of length $PG = 15$, process 200 symbols in training mode (TR) and then switch to the decision-directed (DD) mode.

The convergence performance of the algorithms for a system with $N = 4$ users are shown in Figs. 1 and 2, where the MUDs process 200 symbols in TR and 800 symbols in DD, averaged over 100 independent experiments. We have chosen a training sequence with 200 symbols because simulation experiments indicated it was the minimum necessary length so that MBER algorithms could clearly outperform MSE ones. The criterion used to select the averaging factor β was that it could approximate the inverse of the covariance without making the receiver unstable. Thus, we suggest a value around $\beta = 0.15$ because larger values can make the receiver unstable, whereas smaller values result in inferior performance.

The average BER performance versus E_b/N_0 is shown in Figs. 3 and 4 for a system with a varying number of users where each MUD processes 10^3 symbols averaged over 100 independent experiments. The results show that the averaged algorithms are capable of accelerating the convergence of the algorithms, since they allow the receivers to use larger step sizes without the risk of losing track of the channel, saving transmitting power and increasing system´s capacity.

For the stochastic gradient algorithms, averaging methods considerably increase the convergence performance of the receivers, whereas for the Newton-type algorithms the gains in performance are less significant. The BER and convergence results for the stochastic gradient algorithms, depicted in Figs. 1 and 3, show that the AMBER-AV technique achieves the best performance, followed by the LBER-AV, the AMBER, the LMS-AV , the LBER and the LMS methods. For the Newton-type algorithms, the AMBER-AV technique achieves the best convergence and BER performance, followed by the LBER-AV, the AMBER, the LBER, the LMS-AV and the LMS approaches, as shown in Figs. 2 and 4.

In terms of computational complexity, the averaging methods require $2N$ multiplications and $N+1$ additions beyond the SG and the GN algorithms.

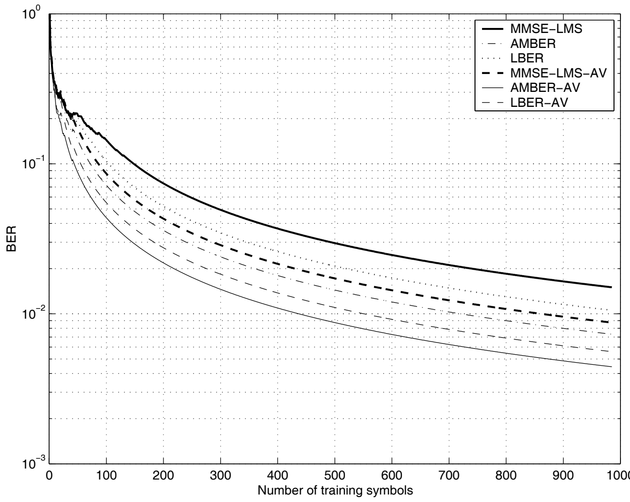

Fig. 1. Convergence of the stochastic gradient algorithms with $K = 4$ users, the mobile moves at 80 km/h in a scenario where the same power is transmitted to all users and the desired user works at $E_b/N_0 = 12$ dB. Parameters: $\mu_0 = 0.001$ for standard algorithms and $\mu_0 = 0.0025$ for the averaged algorithms, $\rho = 8\sigma^2$, $\beta = 0.15$ and $\tau = 0.15$.

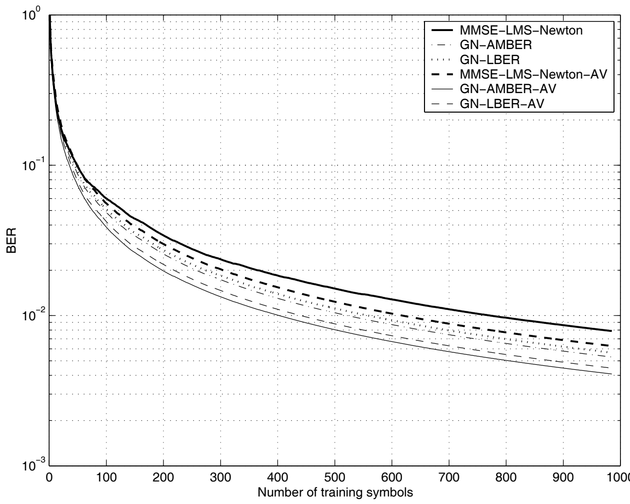

Fig. 2. Convergence of the Gradient-Newton algorithms with $K = 4$ users, the mobile moves at 80 km/h in a scenario where the same power is transmitted to all users and the desired user works at $E_b/N_0 = 12$ dB. Parameters: $\alpha = 0.01$, $\delta = 0.0005$ for standard algorithms and $\delta = 0.001$ for the averaged algorithms, $\rho = 8\sigma^2$, $\beta = 0.15$ and $\tau = 0.15$.

VII. Conclusions

Averaged MBER algorithms for MUDs have been proposed and evaluated in multipath Rayleigh fading channels. The averaging concept has accelerated the convergence of the algorithms and yielded a performance superior to previously reported MBER and MMSE algorithms.

Referências

[1] S. Verdu, *Multiuser Detection*, Cambridge, 1998.

[2] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA,"*IEEE Transactions on Communications, vol. 42, no. 12, December 1994.*

[3] C. Yeh, R. R Lopes and J. R. Barry, "Approximate Minimum Bit-Error Rate Multiuser Detection", GLOBECOM´98, Sydney, Australia, November 1998, pp. 3590-3595.

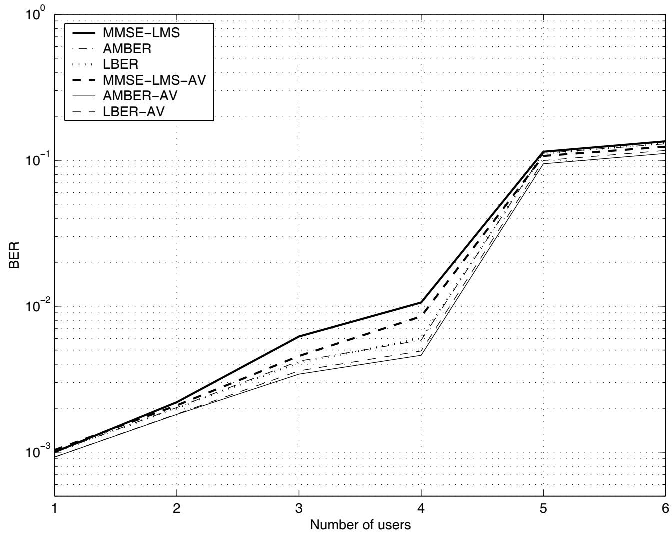

Fig. 3. BER performance of the stochastic gradient algorithms with a varying number of users, in a scenario where the mobile moves at 80 km/h and the same power is transmitted to all users and the desired user works at $E_b/N_0 = 10$ dB. Parameters: $\mu_0 = 0.001$ for standard algorithms and $\mu_0 = 0.0025$ for the averaged algorithms, $\rho = 4\sigma^2$, $\beta = 0.15$ and $\tau = 0.15$.

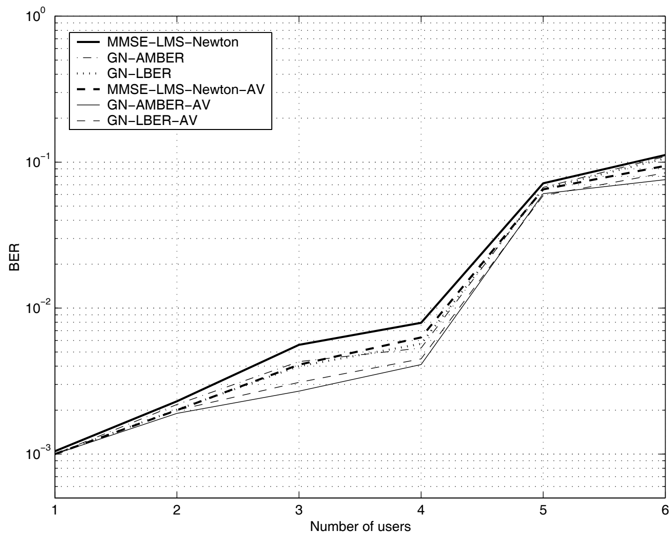

Fig. 4. BER performance of the Gradient-Newton algorithms with a varying number of users, in a scenario where the mobile moves at 80 km/h and the same power is transmitted to all users and the desired user works at $E_b/N_0 = 10$ dB. Parameters: $\alpha = 0.01$, $\delta = 0.0005$ for standard algorithms and $\delta = 0.001$ for the averaged algorithms, $\rho = 4\sigma^2$, $\beta = 0.15$ and $\tau = 0.15$.

[4] S. Chen, A K. Samingan, B. Mulgrew and L. Hanzo, "Adaptive Minimum BER Linear Multiuser Detection,"IEEE Proc. ICASSP, Salt Lake City, USA, May 2001, pp. 2253-2256.

[5] R. C. de Lamare and R. Sampaio-Neto, "Adaptive Multiuser Receivers for DS-CDMA Using Minimum BER Gradient-Newton Algorithms,"IEEE Proc. PIMRC'2002, Lisbon, Portugal, September 2002.

[6] P. S. R. Diniz, Adaptive Filtering: Algorithms and Practical Implementations, Kluwer Academic Publishers, Boston, MA, 1997.

[7] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging,"SIAM J. Contr. Optim., vol. 30, no. 4, pp. 838-855, July 1992.

[8] L. Ljung, "Aspects on accelerated convergence in stochastic approximation schemes,"in Proc. 33rd IEEE Conf. Decision and Control, Lake Buena Vista, FL, 1994, pp. 1649-1652.