CONVERGENCE ANALYSIS OF CONSTRAINED JOINT ADAPTATION IN RECORDING CHANNELS

Lim Sze Chieh and George Mathew

Dept. of ECE, National University of Singapore, Singapore 117576
*Data Storage Institute, A*STAR, Singapore 117608*

Abstract: Although partial response (PR) equalization employing the linearly constrained least-mean-square (LCLMS) algorithm is widely used in recording channels, there is no literature on its convergence analysis. Existing analyses of the LMS algorithm assume that the input signals are jointly Gaussian, which is an invalid assumption for PR equalization with binary input. In this paper, we present a convergence analysis of the LCLMS algorithm, without the Gaussian assumption. An approximate expression is derived for the misadjustment. It is shown that the step-size range required to guarantee stability is larger for binary data compared to Gaussian data.

Keywords: Adaptation, adaptive equalization, constraints, LMS algorithm, mean-square error, partial response channels, recording channels, PR target.

1. INTRODUCTION

There has been considerable research effort to combine partial response (PR) equalization with the Viterbi detector to minimize noise and distortions in recording channels. Conventionally, PR equalization shapes the channel response into a predetermined PR target with integer-valued coefficients. The Viterbi algorithm is then applied to the equalizer output to estimate the stored data bits. In recent years, research has shown that substantial performance improvement can be attained by employing generalized PR (GPR) targets with non-integer-valued coefficients (Moon and Zeng, 1995). Due to time-varying or unknown channel characteristics, it is necessary to employ adaptive approaches to design the equalizer and PR target, with the linearly constrained least-mean-square (LCLMS) algorithm being the most widely used. Substantial research has been done on the LMS algorithm and it is well described in the literature.

The analyses (Farhang-Boroujeny, 2000; Feuer and Weinstein, 1985; Frost, 1972; Godara and Cantoni, 1986) assumed that the input data and desired signal are jointly Gaussian. Unfortunately, in PR equalization, the joint Gaussian assumption is invalid since the input data is binary, even though the channel noise is Gaussian. Although Claasen and Mecklenbräuker (1981) accommodate binary input, the filter input is restricted to a white process. In PR equalization, however, the equalizer input is correlated due to intersymbol interference. Further, because it is necessary to adapt both the equalizer and PR target, the analysis must consider two step-size parameters, whereas existing analyses consider only a single step-size parameter. Thus, existing analyses are inappropriate for studying the convergence of the LCLMS algorithm in PR equalization.

In this paper, a novel approach is first presented in Section 2 to study the convergence behavior of the LCLMS algorithm with a single step-size parameter for adapting both the equalizer and PR target, without making the Gaussian assumption. In Section 3, the necessary modifications are given to accommodate two step-size parameters. Simulation results, which corroborate the theoretical developments, are given in Section 4. The paper is concluded in Section 5.

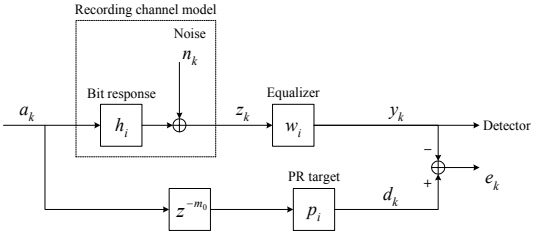

Fig. 1. Recording channel model with PR equalizer.

2. LCLMS ALGORITHM WITH BINARY DATA AND GAUSSIAN NOISE

2.1 System Model

Fig. 1 depicts the recording channel model used in this paper. The channel noise n_k is modeled as additive white Gaussian with variance σ_n^2 and the data bits a_k are chosen to be independent and identically distributed random variables taking values $\{-1, +1\}$ with equal probability. An appropriate value of the delay m_0 is assumed. The equalizer tap-weights, equalizer input, PR target tap-weights, and PR target input are defined, respectively, as the real-valued column vectors $\mathbf{w} = [w_0 \;\; w_1 \;\; \ldots \;\; w_{N_w-1}]^{\mathrm{T}}$, $\mathbf{z}_k = [z_k \;\; z_{k-1} \;\; \ldots \;\; z_{k-N_w+1}]^{\mathrm{T}}$, $\mathbf{p} = [p_0 \;\; p_1 \;\; \ldots \;\; p_{N_p-1}]^{\mathrm{T}}$, and $\mathbf{a}_{k-m_0} = [a_{k-m_0} \;\; a_{k-m_0-1} \;\; \ldots \;\; a_{k-m_0-N_p+1}]^{\mathrm{T}}$, where N_w is the equalizer length, N_p is the PR target length, and the superscript 'T' denotes the transpose operator.

A trivial solution that minimizes the mean-square error (MSE) $\mathrm{E}[e_k^2]$ is $\mathbf{w} = \mathbf{0}$ and $\mathbf{p} = \mathbf{0}$, where $\mathrm{E}[\cdot]$ denotes statistical expectation. This, of course, is useless and undesirable since it corresponds to no transmission through the channel. Thus, the objective is to minimize the MSE, subject to $\mathbf{Bp} = \mathbf{b}$, where $\mathbf{B}$ is a $N_c \times N_p$ matrix, $\mathbf{b}$ is a $N_c \times 1$ vector, and N_c is the number of linearly independent constraint(s).

In the subsequent derivations, it is assumed that

- $\{a_k\}$ and $\{n_k\}$ are zero-mean and mutually independent stationary processes;
- At time k, the tap-weight vectors, $\mathbf{w}_k$ and $\mathbf{p}_k$, are independent of the input vectors, $\mathbf{z}_k$ and $\mathbf{a}_{k-m_0}$ (henceforth called the independence assumption).

2.2 LCLMS Algorithm

The LCLMS algorithm is given by (Frost, 1972)

$$\mathbf{w}_{k+1} = \mathbf{w}_k + 2\mu_w e_k \mathbf{z}_k ,$$
$$\mathbf{p}_{k+1} = \mathbf{A}(\mathbf{p}_k - 2\mu_p e_k \mathbf{a}_{k-m_0}) + \mathbf{B}^{\#}\mathbf{b} \qquad (1)$$

where μ_w is the equalizer step-size parameter, μ_p is the PR target step-size parameter, $\mathbf{B}^{\#} = \mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{-1}$, $\mathbf{A} = \mathbf{I} - \mathbf{B}^{\#}\mathbf{B}$, and $e_k = \mathbf{a}_{k-m_0}^{\mathrm{T}}\mathbf{p}_k - \mathbf{z}_k^{\mathrm{T}}\mathbf{w}_k$. Also, it can be easily seen that

$$e_k = \mathbf{a}_{k-m_0}^{\mathrm{T}}\mathbf{g}_k - \mathbf{z}_k^{\mathrm{T}}\mathbf{f}_k + e_{opt,k} \qquad (2)$$

where $\mathbf{g}_k = \mathbf{p}_k - \mathbf{p}_{opt}$, $\mathbf{f}_k = \mathbf{w}_k - \mathbf{w}_{opt}$, $\mathbf{p}_{opt}$ is the optimum PR target tap-weight vector, $\mathbf{w}_{opt}$ is the optimum equalizer tap-weight vector, and $e_{opt,k} = \mathbf{a}_{k-m_0}^{\mathrm{T}}\mathbf{p}_{opt} - \mathbf{z}_k^{\mathrm{T}}\mathbf{w}_{opt}$ is the optimum error. Substituting (2) into (1), and using $\mathbf{Bp}_{opt} = \mathbf{b}$, yields

$$\mathbf{f}_{k+1} = \mathbf{f}_k - 2\mu_w \mathbf{z}_k \mathbf{z}_k^{\mathrm{T}}\mathbf{f}_k + 2\mu_w \mathbf{z}_k \mathbf{a}_{k-m_0}^{\mathrm{T}}\mathbf{g}_k + 2\mu_w e_{opt,k}\mathbf{z}_k ,$$
$$\mathbf{g}_{k+1} = \mathbf{A}(\mathbf{g}_k - 2\mu_p \mathbf{a}_{k-m_0}\mathbf{a}_{k-m_0}^{\mathrm{T}}\mathbf{g}_k + 2\mu_p \mathbf{a}_{k-m_0}\mathbf{z}_k^{\mathrm{T}}\mathbf{f}_k - 2\mu_p e_{opt,k}\mathbf{a}_{k-m_0}) . \qquad (3)$$

With $\mu_w = \mu_p = \mu$, (3) can be expressed as

$$\mathbf{v}_{k+1} = \mathbf{\Psi}[(\mathbf{I} - 2\mu \mathbf{y}_k \mathbf{y}_k^{\mathrm{T}})\mathbf{v}_k + 2\mu e_{opt,k}\mathbf{y}_k] \qquad (4)$$

where $\mathbf{v}_k = \begin{bmatrix} \mathbf{f}_k \\ \mathbf{g}_k \end{bmatrix}$, $\mathbf{\Psi} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix}$, and $\mathbf{y}_k = \begin{bmatrix} \mathbf{z}_k \\ -\mathbf{a}_{k-m_0} \end{bmatrix}$.

The matrix $\mathbf{A}$ performs orthogonal projection onto a subspace orthogonal to the row space of the constraint matrix $\mathbf{B}$ (Cowen, 1997). The matrix $\mathbf{\Psi}$ is also a projection matrix. A projection matrix is both symmetric and idempotent. These two properties are crucial in subsequent derivations. Let us investigate the characteristics of $\mathbf{\Psi}$. The constraints can be expressed as $\mathbf{\Theta e} = \mathbf{b}$, where $\mathbf{\Theta} = [\mathbf{0} \;\; \mathbf{B}]$ and $\mathbf{e}^{\mathrm{T}} = [\mathbf{w}^{\mathrm{T}} \;\; \mathbf{p}^{\mathrm{T}}]$. Since $\mathbf{\Psi}$ performs orthogonal projection onto a subspace $S^{\perp}$ orthogonal to the subspace S spanned by the rows of $\mathbf{\Theta}$, there exist N_c mutually orthogonal eigenvectors in S with eigenvalue 0. Similarly, there also exist $(N - N_c)$ mutually orthogonal eigenvectors in $S^{\perp}$ with eigenvalue 1, where $N = N_w + N_p$.

From (1), it can be easily deduced that $\mathrm{E}[e_{opt,k}\mathbf{z}_k] = \mathbf{0}$ and $\mathrm{E}[e_{opt,k}\mathbf{A}\mathbf{a}_{k-m_0}] = \mathbf{0}$. Consequently, the principle of orthogonality can be written as

$$\mathrm{E}[e_{opt,k}\mathbf{\Psi}\mathbf{y}_k] = \mathbf{0} . \qquad (5)$$

Note that $\mathbf{\Psi}$ is singular if there is at least one linear constraint. Thus, in general, it cannot be concluded that $\mathrm{E}[e_{opt,k}\mathbf{y}_k] = \mathbf{0}$.

2.3 Average Tap-Weight Behavior

With the tap-weights initialized such that they satisfy the constraint(s), it can be inferred that $\mathbf{v}_0 = \mathbf{\Psi}\mathbf{v}_0$. In addition, since the tap-weights lie on the constraint hyperplane at every adaptation step, it is observed that $\mathbf{v}_k = \mathbf{\Psi}\mathbf{v}_k$, for any positive integer k. Since $\mathbf{\Psi}$ is idempotent, it is evident that

$$\mathbf{v}_k = \mathbf{\Psi}^n \mathbf{v}_k , \qquad n \in Z^+, k = 0, 1, 2, \ldots \qquad (6)$$

As a result, with $\mathbf{x}_k = \mathbf{\Psi}\mathbf{y}_k$, (4) can be expressed as

$$\mathbf{v}_{k+1} = (\mathbf{I} - 2\mu \mathbf{x}_k \mathbf{x}_k^{\mathrm{T}})\mathbf{v}_k + 2\mu e_{opt,k}\mathbf{x}_k . \qquad (7)$$

Under the independence assumption and the principle of orthogonality in (5), it follows from (7) that

$$\mathrm{E}[\mathbf{v}_{k+1}] = (\mathbf{I} - 2\mu \mathbf{R}_x)\mathrm{E}[\mathbf{v}_k] \qquad (8)$$

with $\mathbf{R}_x = \mathrm{E}[\mathbf{x}_k \mathbf{x}_k^{\mathrm{T}}] = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{T}}$, where $\mathbf{\Lambda}$ is a diagonal matrix consisting of the eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ of $\mathbf{R}_x$ and the columns of $\mathbf{Q}$ contain the corresponding orthonormal eigenvectors. With $\mathbf{v}'_k = \mathbf{Q}^{\mathrm{T}}\mathbf{v}_k$, (8) can

be rewritten as the vector recursive equation $\mathrm{E}[\mathbf{v}'_{k+1}] = (\mathbf{I} - 2\mu\mathbf{\Lambda})\mathrm{E}[\mathbf{v}'_k]$, which can be separated into the scalar recursive equations

$$\mathrm{E}[v'_{k+1,i}] = (1 - 2\mu\lambda_i)\mathrm{E}[v'_{k,i}] \tag{9}$$

where $v'_{k,i}$ is the i-th element of the vector $\mathbf{v}'_k$.

Note that $\mathbf{R}_x = \mathbf{\Psi}\mathbf{R}_y\mathbf{\Psi}$, where $\mathbf{R}_y = \mathrm{E}[\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]$ is positive definite. Thus, the eigenvectors of $\mathbf{\Psi}$ with eigenvalue 0 are also eigenvectors of $\mathbf{R}_x$ with eigenvalue 0. These eigenvectors lie in the subspace S and the projection of $\mathbf{v}_k$ in the directions of these eigenvectors is zero for all k. Since $\mathbf{v}_k$ has already converged in these directions, it does not pose any problem even if $\lambda_i = 0$ in these directions. From (9), in order for $\mathbf{v}_k$ to converge in the remaining directions given by the orthonormal eigenvectors which lie in the subspace $S^{\perp}$, it is imperative that the remaining eigenvalues are nonzero. Let $\mathbf{q} \neq \mathbf{0}$ be one such eigenvector. Then, $\mathbf{\Psi}\mathbf{q} = \mathbf{q}$ and

$$\mathbf{q}^{\mathrm{T}}\mathbf{R}_x\mathbf{q} = \mathbf{q}^{\mathrm{T}}\mathbf{\Psi}\mathbf{R}_y\mathbf{\Psi}\mathbf{q} = \mathbf{q}^{\mathrm{T}}\mathbf{R}_y\mathbf{q} > 0 \tag{10}$$

since $\mathbf{R}_y$ is positive definite. Further, since $\mathbf{R}_x$ is at least positive semidefinite (psd), (10) implies that the eigenvalue corresponding to $\mathbf{q}$ must be positive. From (9), it can be seen that for $\mathrm{E}[v'_{k,i}]$ to converge to zero for all i, it is necessary that $|1 - 2\mu\lambda_i| < 1$ for those positive eigenvalues λ_i. Thus, it is concluded that

$$0 < \mu < 1/\lambda_{\max} \tag{11}$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{R}_x$.

2.4 MSE Behavior

The convergence of the LCLMS algorithm requires the convergence of the mean of $\mathbf{v}_k$ to zero and the convergence of the variance of the elements of $\mathbf{v}_k$ to some limited values (Farhang-Boroujeny, 2000). In this section, the latter is investigated, without using the Gaussian assumption.

From (2) and (6), $e_k = e_{opt,k} - \mathbf{v}_k^{\mathrm{T}}\mathbf{y}_k = e_{opt,k} - \mathbf{v}_k^{\mathrm{T}}\mathbf{x}_k$. As a result, it follows that

$$\xi_k = \mathrm{E}[e_k^2] = \xi_{\min} + \mathrm{tr}(\mathbf{D}_k\mathbf{R}_x) \tag{12}$$

where $\xi_{\min} = \mathrm{E}[e_{opt,k}^2]$, $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, and $\mathbf{D}_k = \mathrm{E}[\mathbf{v}_k\mathbf{v}_k^{\mathrm{T}}]$. Further, excess MSE is defined as $\xi_{\mathrm{ex}} = \mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x)$.

From (7), using the independence assumption and the principle of orthogonality (5), we get

$$\begin{aligned}\mathrm{E}[\mathbf{v}_{k+1}^{\mathrm{T}}\mathbf{v}_{k+1}] &= \mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{v}_k] - 4\mu\,\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}]\mathbf{v}_k] \\ &\quad + 4\mu^2\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}]\mathbf{v}_k] \\ &\quad - 8\mu^2\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}]\mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k e_{opt,k}] + 4\mu^2\mathrm{E}[e_{opt,k}^2\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k]\end{aligned} \tag{13}$$

With $\mathbf{u}^{\mathrm{T}} = [\mathbf{w}_{opt}^{\mathrm{T}} \;\; \mathbf{p}_{opt}^{\mathrm{T}}]$, we get $e_{opt,k} = -\mathbf{u}^{\mathrm{T}}\mathbf{y}_k$ and

$$\begin{aligned}\mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}] &= \mathbf{\Psi}\,\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]\mathbf{\Psi}, \\ \mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k e_{opt,k}] &= -\mathbf{\Psi}\,\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]\mathbf{u}, \\ \mathrm{E}[e_{opt,k}^2\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k] &= \mathbf{u}^{\mathrm{T}}\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]\mathbf{u}.\end{aligned} \tag{14}$$

Thus, the main task is to simplify the fourth-order moment $\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]$.

Firstly, with $z_k = n_k + \mathbf{h}^{\mathrm{T}}\tilde{\mathbf{a}}_k$, it can be shown that

$$\begin{aligned}\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k &= \textstyle\sum_{i=0}^{N_w-1} n_{k-i}^2 + 2\sum_{i=0}^{N_w-1} n_{k-i}\mathbf{h}^{\mathrm{T}}\tilde{\mathbf{a}}_{k-i} \\ &\quad + \textstyle\sum_{i=0}^{N_w-1}\mathbf{h}^{\mathrm{T}}\tilde{\mathbf{a}}_{k-i}\tilde{\mathbf{a}}_{k-i}^{\mathrm{T}}\mathbf{h} + \mathbf{a}_{k-m_0}^{\mathrm{T}}\mathbf{A}\mathbf{a}_{k-m_0}\end{aligned} \tag{15}$$

where $\mathbf{h} = [h_0 \;\; h_1 \;\ldots\; h_{N_h-1}]^{\mathrm{T}}$ represents the channel bit response, $\tilde{\mathbf{a}}_k = [a_k \;\; a_{k-1} \;\ldots\; a_{k-N_h+1}]^{\mathrm{T}}$, and N_h is the channel length. Next, $\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]$ is partitioned as

$$\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}] = \begin{pmatrix}\mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22}\end{pmatrix} \tag{16}$$

where $\mathbf{Y}_{11} = \mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{T}}]$, $\mathbf{Y}_{12} = -\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{z}_k\mathbf{a}_{k-m_0}^{\mathrm{T}}]$, $\mathbf{Y}_{21} = \mathbf{Y}_{12}^{\mathrm{T}}$, and $\mathbf{Y}_{22} = \mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{a}_{k-m_0}\mathbf{a}_{k-m_0}^{\mathrm{T}}]$. Further, the fourth-order moment of the data bits a_k is given by $\mathrm{E}[a_i a_j a_k a_l] = \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - 2\delta_{ij}\delta_{ik}\delta_{il}$. Since a_k and n_k are mutually independent, it is possible to find an exact expression for $\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]$ (see Appendix A). Next, $\mathbf{R}_y$ is also partitioned as

$$\begin{aligned}\mathbf{R}_y = \mathrm{E}[\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}] &= \begin{bmatrix}\mathbf{R}_z & -\mathbf{R}_{za} \\ -\mathbf{R}_{za}^{\mathrm{T}} & \mathbf{R}_a\end{bmatrix}, \;\; \mathbf{R}_z = \mathrm{E}[\mathbf{z}_k\mathbf{z}_k^{\mathrm{T}}], \\ \mathbf{R}_{za} &= \mathrm{E}[\mathbf{z}_k\mathbf{a}_{k-m_0}^{\mathrm{T}}], \;\; \mathbf{R}_a = \mathrm{E}[\mathbf{a}_{k-m_0}\mathbf{a}_{k-m_0}^{\mathrm{T}}], \\ \mathbf{R}_z(i,j) &= \sigma_n^2\delta_{ij} + \textstyle\sum_{\alpha=0}^{N_h-1} h_\alpha h_{\alpha+i-j}, \\ \mathbf{R}_{za}(i,j) &= h_{m_0+j-i}, \;\; \mathbf{R}_a(i,j) = \delta_{ij}.\end{aligned} \tag{17}$$

Consequently, it is noted that

$$\begin{aligned}\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k] &= \mathrm{tr}(\mathbf{R}_x) = \mathrm{tr}(\mathbf{R}_z) + \mathrm{tr}(\mathbf{A}\mathbf{R}_a) \\ &= N_w\sigma_n^2 + N_w\textstyle\sum_{\alpha=0}^{N_h-1} h_\alpha^2 + \mathrm{tr}(\mathbf{A}).\end{aligned} \tag{18}$$

Upon inspection of the expressions for $\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]$, $\mathbf{R}_y$ and $\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k]$, it can be deduced that

$$\begin{aligned}\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}] &= \mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k]\mathbf{R}_y \\ &\quad + 2\left\{\mathbf{R}_y\mathbf{\Psi}\mathbf{R}_y - \mathbf{R}_y\begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{G}(\mathbf{A})\end{bmatrix}\mathbf{R}_y - \begin{pmatrix}\mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{pmatrix}\right\}\end{aligned} \tag{19}$$

where $\mathcal{G}_{ij}(\mathbf{A}) = \mathbf{A}(i,j)\delta_{ij}$ is the (i,j)-th element of $\mathcal{G}(\mathbf{A})$. Also, $\mathbf{H}_{11}$, $\mathbf{H}_{12}$ and $\mathbf{H}_{22}$ are given by

$$\mathbf{H}_{22} = \mathrm{diag}\left(\sum_{\alpha=0}^{N_w-1} h_{m_0-\alpha}^2, \sum_{\alpha=0}^{N_w-1} h_{m_0+1-\alpha}^2, \ldots, \sum_{\alpha=0}^{N_w-1} h_{m_0+N_p-1-\alpha}^2\right),$$

$$\mathbf{H}_{12} = -\underbrace{\begin{pmatrix} h_{m_0} & h_{m_0+1} & \cdots & h_{m_0+N_p-1} \\ h_{m_0-1} & h_{m_0} & \cdots & h_{m_0+N_p-2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m_0-N_w+1} & h_{m_0-N_w+2} & \cdots & h_{m_0+N_p-N_w}\end{pmatrix}}_{\text{Toeplitz Matrix}}\mathbf{H}_{22},$$

$$\mathbf{H}_{11} = \mathbf{C}\mathbf{Z}\mathbf{C}^{\mathrm{T}}, \;\; \mathbf{C} = \underbrace{\begin{pmatrix} h_0 & h_1 & \cdots & h_{N_w+N_h-2} \\ h_{-1} & h_0 & \cdots & h_{N_w+N_h-3} \\ \vdots & \vdots & \ddots & \vdots \\ h_{-N_w+1} & h_{-N_w+2} & \cdots & h_{N_h-1}\end{pmatrix}}_{\text{Toeplitz Matrix}},$$

$$\mathbf{Z} = \mathrm{diag}\left(\sum_{\alpha=0}^{N_w-1} h_{0-\alpha}^2, \sum_{\alpha=0}^{N_w-1} h_{1-\alpha}^2, \ldots, \sum_{\alpha=0}^{N_w-1} h_{N_w+N_h-2-\alpha}^2\right) \tag{20}$$

where diag(…) denotes a diagonal matrix consisting of the indicated elements. Note that $\mathbf{H}_{22} = \mathbf{R}_a\mathbf{H}_{22}\mathbf{R}_a$ and $\mathbf{H}_{12} = -\mathbf{R}_{za}\mathbf{H}_{22}\mathbf{R}_a$. This inspires us to write

$$\begin{pmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{pmatrix}=\mathbf{R}_y\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_{22}\end{bmatrix}\mathbf{R}_y+\begin{bmatrix}\boldsymbol{\Phi} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$
$$=\begin{bmatrix}\mathbf{R}_{za}\mathbf{H}_{22}\mathbf{R}_{za}^{\mathrm{T}} & -\mathbf{R}_{za}\mathbf{H}_{22}\mathbf{R}_a\\ -\mathbf{R}_a\mathbf{H}_{22}\mathbf{R}_{za}^{\mathrm{T}} & \mathbf{R}_a\mathbf{H}_{22}\mathbf{R}_a\end{bmatrix}+\begin{bmatrix}\boldsymbol{\Phi} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \quad (21)$$

where $\boldsymbol{\Phi}$ is some unknown $N_w \times N_w$ matrix to be found. The matrices $\mathbf{Z}=\mathrm{diag}(\mathbf{Z}_1,\mathbf{Z}_2,\mathbf{Z}_3)$ and $\mathbf{C}=[\mathbf{C}_1\ \mathbf{C}_2\ \mathbf{C}_3]$ are partitioned such that $\mathbf{Z}_2=\mathbf{H}_{22}$ and $\mathbf{C}_2=\mathbf{R}_{za}$. As a result, $\mathbf{H}_{11}$ can be expressed as

$$\mathbf{H}_{11}=\mathbf{C}_1\mathbf{Z}_1\mathbf{C}_1^{\mathrm{T}}+\mathbf{R}_{za}\mathbf{H}_{22}\mathbf{R}_{za}^{\mathrm{T}}+\mathbf{C}_3\mathbf{Z}_3\mathbf{C}_3^{\mathrm{T}}. \quad (22)$$

The only requirement for writing (22) is given by $0\le m_0, m_0+N_p-1\le N_w+N_h-2$, which is satisfied in practice. From (21) and (22), it can be seen that $\boldsymbol{\Phi}=\mathbf{C}_1\mathbf{Z}_1\mathbf{C}_1^{\mathrm{T}}+\mathbf{C}_3\mathbf{Z}_3\mathbf{C}_3^{\mathrm{T}}$. Thus, (19) can be written as

$$\mathrm{E}[\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}]=\mathrm{tr}(\mathbf{R}_x)\mathbf{R}_y+2(\mathbf{R}_y\boldsymbol{\Psi}\mathbf{R}_y-\mathbf{L}) \quad (23)$$

where $\mathbf{L}=\mathbf{R}_y\left(\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathcal{G}(\mathbf{A})\end{bmatrix}+\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{H}_{22}\end{bmatrix}\right)\mathbf{R}_y+\begin{bmatrix}\boldsymbol{\Phi} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$.

The matrices $\mathbf{A}$, $\mathbf{H}_{22}$ and $\boldsymbol{\Phi}$ are clearly psd. Since the diagonal elements of a psd matrix are nonnegative, $\mathcal{G}(\mathbf{A})$ is also psd. Thus, it is clear that $\mathbf{L}$ is psd. Noting from (5) that $\mathrm{E}[e_{opt,k}\boldsymbol{\Psi}\mathbf{y}_k]=\boldsymbol{\Psi}\mathrm{E}[\mathbf{y}_k(-\mathbf{y}_k^{\mathrm{T}}\mathbf{u})]=-\boldsymbol{\Psi}\mathbf{R}_y\mathbf{u}=\mathbf{0}$, and substituting (23) into (14), we get

$$\mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}]=\mathrm{tr}(\mathbf{R}_x)\mathbf{R}_x+2\mathbf{R}_x^2-2\boldsymbol{\Psi}\mathbf{L}\boldsymbol{\Psi},$$
$$\mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k e_{opt,k}]=2\boldsymbol{\Psi}\mathbf{L}\mathbf{u},$$
$$\mathrm{E}[e_{opt,k}^2\mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k]=\xi_{\min}\mathrm{tr}(\mathbf{R}_x)-2\mathbf{u}^{\mathrm{T}}\mathbf{L}\mathbf{u}. \quad (24)$$

Substituting (24) into (13) yields

$$\mathrm{E}[\mathbf{v}_{k+1}^{\mathrm{T}}\mathbf{v}_{k+1}]=\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{v}_k]-4\mu\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{R}_x\mathbf{v}_k]$$
$$+4\mu^2\mathrm{E}\{\mathbf{v}_k^{\mathrm{T}}[\mathrm{tr}(\mathbf{R}_x)\mathbf{R}_x+2\mathbf{R}_x^2-2\boldsymbol{\Psi}\mathbf{L}\boldsymbol{\Psi}]\mathbf{v}_k\}$$
$$-16\mu^2\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}]\boldsymbol{\Psi}\mathbf{L}\mathbf{u}+4\mu^2\xi_{\min}\mathrm{tr}(\mathbf{R}_x)-8\mu^2\mathbf{u}^{\mathrm{T}}\mathbf{L}\mathbf{u}. \quad (25)$$

If the LCLMS algorithm is convergent, then $\mathrm{E}[\mathbf{v}_{k+1}^{\mathrm{T}}\mathbf{v}_{k+1}]=\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{v}_k]$ and $\mathrm{E}[\mathbf{v}_k]=\mathbf{0}$ as $k\to\infty$. From (25), with $\mu\ne 0$, it can be shown that

$$\mu\xi_{\min}\mathrm{tr}(\mathbf{R}_x)=\xi_{\mathrm{ex}}-\mu\,\mathrm{tr}(\mathbf{R}_x)\xi_{\mathrm{ex}}-2\mu\,\mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x^2)$$
$$+2\mu\,\mathrm{tr}(\mathbf{D}_\infty\mathbf{L})+2\mu\mathbf{u}^{\mathrm{T}}\mathbf{L}\mathbf{u}. \quad (26)$$

Since $\mathbf{L}$ is psd, $\mathrm{tr}(\mathbf{D}_\infty\mathbf{L})=\mathrm{E}[\mathbf{v}_\infty^{\mathrm{T}}\mathbf{L}\mathbf{v}_\infty]\ge 0$ and $\mathbf{u}^{\mathrm{T}}\mathbf{L}\mathbf{u}\ge 0$. Therefore, $2\mu\,\mathrm{tr}(\mathbf{D}_\infty\mathbf{L})+2\mu\mathbf{u}^{\mathrm{T}}\mathbf{L}\mathbf{u}$ and $-2\mu\,\mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x^2)$ would cancel each other partially. Extensive computer simulations indicate that $\mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x^2)-\mathrm{tr}(\mathbf{D}_\infty\mathbf{L})-\mathbf{u}^{\mathrm{T}}\mathbf{L}\mathbf{u}$ is at least an order of magnitude smaller than $\mathrm{tr}(\mathbf{R}_x)\xi_{\mathrm{ex}}$ and thus can be ignored. Using this in (26) yields the misadjustment

$$\mathrm{M}=\xi_{\mathrm{ex}}/\xi_{\min}\approx\mu\,\mathrm{tr}(\mathbf{R}_x)/[1-\mu\,\mathrm{tr}(\mathbf{R}_x)]. \quad (27)$$

If the input data a_k were Gaussian distributed with zero mean and unit variance, then $\mathbf{L}=\mathbf{0}$ and the last two terms of (26) disappear. Further, observe that

$$\mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x^2)=\mathrm{tr}(\mathbf{D}'_\infty\boldsymbol{\Lambda}^2)=\sum_{i=0}^{N-1}\lambda_i^2 D'_\infty(i,i),$$
$$\mathrm{tr}(\mathbf{R}_x)\xi_{\mathrm{ex}}=\mathrm{tr}(\mathbf{R}_x)\mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x)=\mathrm{tr}(\boldsymbol{\Lambda})\mathrm{tr}(\mathbf{D}'_\infty\boldsymbol{\Lambda})$$
$$=\sum_{i=0}^{N-1}\lambda_i^2 D'_\infty(i,i)+\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\underbrace{\lambda_i\lambda_j D'_\infty(i,i)}_{i\ne j} \quad (28)$$

where $\mathbf{D}'_k=\mathbf{Q}^{\mathrm{T}}\mathbf{D}_k\mathbf{Q}$ and $D'_k(i,j)$ is the (i,j)-th element of $\mathbf{D}'_k$. Since λ_i and $D'_\infty(i,i)$ are non-negative and almost always positive, $\mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x^2)$ is expected to be negligible relative to $\mathrm{tr}(\mathbf{R}_x)\xi_{\mathrm{ex}}$. Extensive computer simulations indicate that $\mathrm{tr}(\mathbf{D}_\infty\mathbf{R}_x^2)$ is at least an order of magnitude smaller than $\mathrm{tr}(\mathbf{R}_x)\xi_{\mathrm{ex}}$ and thus can be ignored. Using this again yields (27) and therefore it can be concluded that the misadjustment is approximately the same for both binary data and Gaussian data.

Next, the stability of the LCLMS algorithm is investigated by expressing (25) as

$$\mathrm{E}[\mathbf{v}_{k+1}^{\mathrm{T}}\mathbf{v}_{k+1}]=\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{F}\mathbf{v}_k]-16\mu^2\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}]\boldsymbol{\Psi}\mathbf{L}\mathbf{u}$$
$$+4\mu^2\xi_{\min}\mathrm{tr}(\mathbf{R}_x)-8\mu^2\mathbf{u}^{\mathrm{T}}\mathbf{L}\mathbf{u} \quad (29)$$

where $\mathbf{F}=\mathbf{I}-4\mu\mathbf{R}_x+4\mu^2[\mathrm{tr}(\mathbf{R}_x)\mathbf{R}_x+2\mathbf{R}_x^2-2\boldsymbol{\Psi}\mathbf{L}\boldsymbol{\Psi}]$. To determine the range of μ that guarantees the convergence of $\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{v}_k]$, the term $\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}]\boldsymbol{\Psi}\mathbf{L}\mathbf{u}$ is examined. A necessary condition is that the range of μ must guarantee the convergence of $\mathrm{E}[\mathbf{v}_k]$ and this is given by (11). Since $\mathbf{F}$ is psd, it is deduced that $0\le\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{F}\mathbf{v}_k]\le\lambda_{\max}(\mathbf{F})\mathrm{E}[\mathbf{v}_k^{\mathrm{T}}\mathbf{v}_k]$, where $\lambda_{\max}(\mathbf{F})$ is the maximum eigenvalue of $\mathbf{F}$ (Bertsekas, 1999). Thus, from (29), it is also required that $\lambda_{\max}(\mathbf{F})<1$. Let $\mathbf{F}=\mathbf{E}+\mathbf{G}$, where $\mathbf{E}=-8\mu^2\boldsymbol{\Psi}\mathbf{L}\boldsymbol{\Psi}$ and $\mathbf{G}=\mathbf{I}-4\mu\mathbf{R}_x+4\mu^2[\mathrm{tr}(\mathbf{R}_x)\mathbf{R}_x+2\mathbf{R}_x^2]$. If $\lambda_i(\mathbf{G})$ denotes the i-th smallest eigenvalue of $\mathbf{G}$, then

$$\lambda_{\min}(\mathbf{E})+\lambda_k(\mathbf{G})\le\lambda_k(\mathbf{F})\le\lambda_{\max}(\mathbf{E})+\lambda_k(\mathbf{G}) \quad (30)$$

for all k (Golub and Van Loan, 1989). Since $\mathbf{F}$ is psd and $\mathbf{E}$ is negative semidefinite, choosing μ to make the eigenvalues of $\mathbf{G}$ less than 1 will also ensure that the eigenvalues of $\mathbf{F}$ are less than 1. It can be shown that to guarantee stability, it is necessary that $0<\mu<1/[\mathrm{tr}(\mathbf{R}_x)+2\lambda_{\max}]$. Since $\lambda_{\max}<\mathrm{tr}(\mathbf{R}_x)$, a conservative and convenient range is

$$0<\mu<1/[3\mathrm{tr}(\mathbf{R}_x)]. \quad (31)$$

If the input data a_k were Gaussian distributed with zero mean and unit variance, then $\mathbf{L}=\mathbf{E}=\mathbf{0}$ and $\mathbf{F}=\mathbf{G}$. Clearly, this implies that binary data provides greater stability range than Gaussian data. This is verified by simulations, presented in Section 4.

3. LCLMS ALGORITHM WITH TWO STEP-SIZES

Section 2 assumed that $\mu_w=\mu_p=\mu$. In this section, the assumption is discarded and (4) is rewritten as

$$\mathbf{v}_{k+1}=\boldsymbol{\Psi}[(\mathbf{I}-2\boldsymbol{\Omega}\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}})\mathbf{v}_k+2e_{opt,k}\boldsymbol{\Omega}\mathbf{y}_k] \quad (32)$$

where $\boldsymbol{\Omega}=\mathrm{diag}(\mu_w\mathbf{I}_w,\mu_p\mathbf{I}_p)$, $\mathbf{I}_w$ and $\mathbf{I}_p$ are identity matrices of sizes N_w and N_p, respectively, and $\mu_w,\mu_p>0$. Since $\boldsymbol{\Psi}\boldsymbol{\Omega}=\boldsymbol{\Omega}\boldsymbol{\Psi}$, (32) is rewritten as

$$\mathbf{v}_{k+1}=(\mathbf{I}-2\boldsymbol{\Omega}\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}})\mathbf{v}_k+2e_{opt,k}\boldsymbol{\Omega}\mathbf{x}_k$$
$$\Rightarrow\quad\tilde{\mathbf{v}}_{k+1}=(\mathbf{I}-2\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^{\mathrm{T}})\tilde{\mathbf{v}}_k+2e_{opt,k}\tilde{\mathbf{x}}_k \quad (33)$$

where $\tilde{\mathbf{v}}_k=\boldsymbol{\Omega}^{-1/2}\mathbf{v}_k$ and $\tilde{\mathbf{x}}_k=\boldsymbol{\Omega}^{1/2}\mathbf{x}_k$. Comparing (33) with (7) shows that the derivations that follow would be very similar to the case where $\mu_w=\mu_p=\mu$. The correlation matrix of importance

in the analysis of $\widetilde{\mathbf{v}}_k$ is $\mathbf{R}_{\widetilde{x}} = \mathrm{E}[\widetilde{\mathbf{x}}_k \widetilde{\mathbf{x}}_k^{\mathrm{T}}]$. Note that $\mathbf{R}_{\widetilde{x}} = \mathbf{\Omega}^{1/2}\mathbf{\Psi}\mathbf{R}_y\mathbf{\Psi}\mathbf{\Omega}^{1/2} = \mathbf{\Psi}\mathbf{\Omega}^{1/2}\mathbf{R}_y\mathbf{\Omega}^{1/2}\mathbf{\Psi}$, where $\mathbf{\Omega}^{1/2}\mathbf{R}_y\mathbf{\Omega}^{1/2}$ is positive definite since $\mu_w, \mu_p > 0$.

Using similar steps as before yields misadjustment as

$$\mathrm{M} = \frac{\xi_{\mathrm{ex}}}{\xi_{\min}} \approx \frac{\mathrm{tr}(\mathbf{R}_{\widetilde{x}})}{1-\mathrm{tr}(\mathbf{R}_{\widetilde{x}})} = \frac{\mu_w \mathrm{tr}(\mathbf{R}_z) + \mu_p \mathrm{tr}(\mathbf{A})}{1-[\mu_w \mathrm{tr}(\mathbf{R}_z) + \mu_p \mathrm{tr}(\mathbf{A})]} \quad (34)$$

and the sufficient conditions to guarantee stability as

$$\mu_w, \mu_p > 0, \;\; \mathrm{tr}(\mathbf{R}_{\widetilde{x}}) = \mu_w \mathrm{tr}(\mathbf{R}_z) + \mu_p \mathrm{tr}(\mathbf{A}) < 1/3 \,. \quad (35)$$

When $\mu_w = \mu_p = \mu$, $\mathrm{tr}(\mathbf{R}_{\widetilde{x}}) = \mu \mathrm{tr}(\mathbf{R}_x)$. This shows that (34) and (35) are consistent with (27) and (31), respectively. Lastly, the expression $\mathrm{tr}(\mathbf{R}_{\widetilde{x}}) = \mu_w \mathrm{tr}(\mathbf{R}_z) + \mu_p \mathrm{tr}(\mathbf{A})$ is convenient for practical use since $\mathrm{tr}(\mathbf{R}_z)$ is the sum of the powers of the signal samples at the equalizer input and $\mathrm{tr}(\mathbf{A}) = N_p - N_c$ is a known constant.

4. SIMULATION RESULTS

In this section, simulation results are presented to support the theoretical analyses in the previous sections. In the simulations, the perpendicular magnetic recording channel is used. Its bit response h_i, $i = 0, 1, \ldots, (N_h - 1)$, is given by $h_i = h_{s,i} - h_{s,i-1}$, where $h_{s,i} = h_s(t)\big|_{t=(i-\Delta)T}$, $h_s(t) = A \tanh[(\ln 3)t / T_{50}]$, T is the bit duration, A is half the signal amplitude, T_{50} is the time that $h_s(t)$ takes to rise from $-A/2$ to $+A/2$, and Δ is a shift index. Parameters are set as $A = 0.5$, $T = 1$, $T_{50} = 2T$, $\Delta = 19$, $N_h = 40$, $N_w = 15$, $N_p = 5$, and $m_0 = 25$, unless stated otherwise. In the design, the first tap-weight of the PR target is constrained to unity, i.e. 'monic constraint' (Moon and Zeng, 1995). The remaining tap-weights are initialized to zero. Lastly, the signal-to-noise ratio is set to $\mathrm{SNR(dB)} = 10\log_{10}(\sum_{i=0}^{N_h-1} h_i^2 / \sigma_n^2) = 20\,\mathrm{dB}$.

We first investigate the validity of approximation (34), which can be rewritten as

$$\mathrm{tr}(\mathbf{R}_{\widetilde{x}}) = \mu_w \mathrm{tr}(\mathbf{R}_z) + \mu_p \mathrm{tr}(\mathbf{A}) = \mathrm{M}/(\mathrm{M}+1). \quad (36)$$

The step-sizes μ_w and μ_p are chosen according to (36) for M = 10%. The value of μ_w / μ_p is varied to investigate its effect on the transient and steady-state behavior of the algorithm.

Fig. 2 shows the learning curves for various values of μ_w / μ_p. Each plot is based on an ensemble average of 10,000 independent simulation runs. Observe that the steady-state behavior is roughly the same for the three cases, and there is a close match between the learning curves at steady-state and the theoretical value of $\xi_{\min} + \xi_{\mathrm{ex}}$, thus validating (34). Further, the use of distinct step-sizes provides faster convergence for a given misadjustment. It is of interest to investigate the various parameters that relate the value of μ_w / μ_p to the convergence rate.

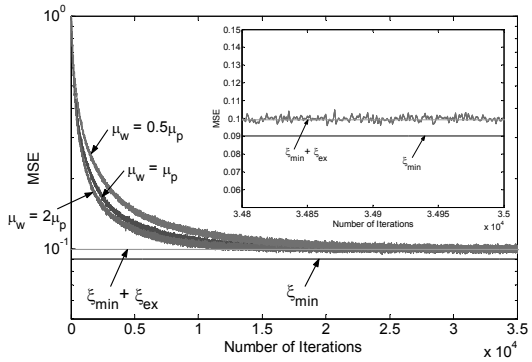

Fig. 2. Learning curves for M = 10% and different values of μ_w / μ_p. Inset: Steady-state behavior.

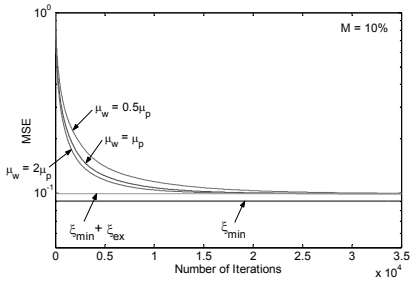

Fig. 3. Theoretical approximate learning curves.

Firstly, an approximate learning curve is derived by noting from (12) that $\xi_k = \xi_{\min} + \mathrm{tr}(\widetilde{\mathbf{D}}_k \mathbf{R}_{\widetilde{x}})$, where $\widetilde{\mathbf{D}}_k = \mathrm{E}[\widetilde{\mathbf{v}}_k \widetilde{\mathbf{v}}_k^{\mathrm{T}}]$. With $\mathbf{R}_{\widetilde{x}} = \widetilde{\mathbf{Q}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{Q}}^{\mathrm{T}}$ where $\widetilde{\mathbf{\Lambda}}$ is a diagonal matrix consisting of the eigenvalues $\widetilde{\lambda}_0, \widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{N-1}$ of $\mathbf{R}_{\widetilde{x}}$ and the columns of $\widetilde{\mathbf{Q}}$ contain the corresponding orthonormal eigenvectors, it can be shown that $\xi_k = \xi_{\min} + \sum_{i=0}^{N-1} \widetilde{\lambda}_i \mathrm{E}[(\widetilde{v}'_{k,i})^2]$, where $\widetilde{v}'_{k,i}$ is the i-th element of $\widetilde{\mathbf{v}}'_k = \widetilde{\mathbf{Q}}^{\mathrm{T}} \widetilde{\mathbf{v}}_k$. When the step-size parameter is small, the following approximation

$$\xi_k \approx \xi_{\min} + \xi_{\mathrm{ex}} + \sum_{i=0}^{N-1} \widetilde{\lambda}_i (1 - 2\widetilde{\lambda}_i)^{2k} (\mathrm{E}[\widetilde{v}'_{0,i}])^2 \quad (37)$$

applies (Farhang-Boroujeny, 2000).

Fig. 3 shows the theoretical approximate learning curves based on (37) for various values of μ_w / μ_p. Figs. 2 and 3 show that the convergence behavior of the simulated learning curves are predicted correctly by the theoretical ones. In fact, extensive simulations indicate that the theoretical and simulated learning curves agree in all cases.

For a given misadjustment M, (36) shows that μ_w and μ_p are constrained such that an increase in μ_w results in a decrease in μ_p, and vice versa. Since large step-sizes result in fast convergence, the fastest overall convergence can be achieved when the equalizer and PR target converge at the same speed. With the monic constraint, the LCLMS algorithm amounts to constraining the first PR target coefficient to unity and adapting the remaining coefficients using the standard unconstrained LMS algorithm. Based on the principle of the normalized LMS algorithm, which provides fast convergence, it is seen from (1) that the normalized step-sizes are given by

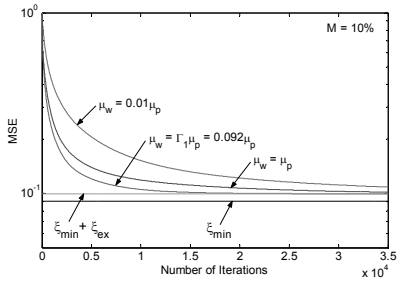

Fig. 4. Learning curves with $A = 2$.

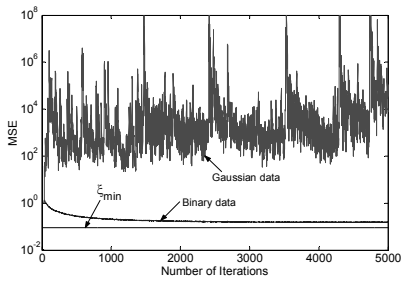

Fig. 5. Comparison of MSE curves with binary data and Gaussian data for M = 70%.

$$\mu_w = \overline{\mu}_w / (\mathbf{z}_k^{\mathrm{T}}\mathbf{z}_k), \quad \mu_p = \overline{\mu}_p / (\mathbf{a}_{k-m_0}^{\mathrm{T}}\widetilde{\mathbf{I}}\mathbf{a}_{k-m_0}) \quad (38)$$

where $\widetilde{\mathbf{I}}$ is an identity matrix with the first diagonal element set to zero. Clearly, $\mathbf{z}_k^{\mathrm{T}}\mathbf{z}_k \approx \mathrm{tr}(\mathbf{R}_z)$ when the equalizer is long enough. Also, $\mathbf{a}_{k-m_0}^{\mathrm{T}}\widetilde{\mathbf{I}}\mathbf{a}_{k-m_0} = \mathrm{tr}(\widetilde{\mathbf{I}})$. The terms $\overline{\mu}_w$ and $\overline{\mu}_p$ control the convergence rates of the equalizer and PR target, respectively (Farhang-Boroujeny, 2000). Thus, setting $\overline{\mu}_w = \overline{\mu}_p$ makes them converge at the same speed. This leads to $\mu_w \mathrm{tr}(\mathbf{R}_z) = \mu_p \mathrm{tr}(\widetilde{\mathbf{I}})$, which can be written as

$$\mu_w = \Gamma_1 \mu_p \quad \text{with} \quad \Gamma_1 = (N_p - N_c)/\mathrm{tr}(\mathbf{R}_z). \quad (39)$$

Fig. 4 shows the theoretical learning curves for various values of μ_w/μ_p. Observe that the step-size ratio given by (39) results in the fastest convergence.

Lastly, the stability ranges for binary data and Gaussian data (with zero mean and unit variance) at the channel input are investigated. Fig. 5 shows the corresponding MSE curves, with each plot based on an ensemble average of 10,000 independent simulation runs. From (36), with $\mu_w = \mu_p = \mu$, the step-size μ is calculated for M = 70%. Observe that the step-size range required to guarantee the stability of the LCLMS algorithm is larger for binary data compared to Gaussian data. This corroborates the conjecture made at the end of Section 2.

5. CONCLUSION

In this paper, a novel approach is proposed to carry out convergence analysis of the LCLMS algorithm with a single step-size parameter. An approximate expression for misadjustment is derived. It is also shown that the step-size range required to guarantee stability is larger for binary data compared to Gaussian data. The analysis is extended to accommodate distinct step-sizes for equalizer and PR target. A rule of thumb is established for choosing the optimal step-size ratio to achieve fast convergence.

REFERENCES

Bertsekas, D.P. (1999). *Nonlinear Programming*, pp. 660–661. Athena Scientific, Massachusetts.

Claasen, T.A.C.M. and W.F.G. Mecklenbräuker (1981). Comparison of the convergence of two algorithms for adaptive FIR digital filters. *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-29**, 670–678.

Cowen, C.C. (1997). *Introduction to Matrix Analysis for Engineering and Science*, pp. 290–296. West Pickle Press, Indiana.

Farhang-Boroujeny, B. (2000). *Adaptive Filters: Theory and Applications*, Chap. 6. John Wiley & Sons, Chichester.

Feuer, A. and E. Weinstein (1985). Convergence analysis of LMS filters with uncorrelated Gaussian data. *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-33**, 222–230.

Frost, O.L. III (1972). An algorithm for linearly constrained adaptive array processing. *Proc. IEEE*, **60**, 926–935.

Godara, L.C. and A. Cantoni (1986). Analysis of constrained LMS algorithm with application to adaptive beamforming using perturbation sequences. *IEEE Trans. Antennas Propagat.*, **AP-34**, 368–379.

Golub, G.H. and C.F. Van Loan (1989). *Matrix Computations*, p. 411. The Johns Hopkins University Press, Baltimore, MD.

Moon, J. and W. Zeng (1995). Equalization for maximum likelihood detectors. *IEEE Trans. Magn.*, **31**, 1083–1088.

APPENDIX A

Noting that $\mathrm{E}[n_k^4] = 3\sigma_n^4$, the various expressions in (16) can be expanded and simplified to

$$\mathbf{Y}_{11}(i,j) = \mathrm{tr}(\mathbf{A})\sigma_n^2\delta_{ij} + \mathrm{tr}(\mathbf{A})\sum_{\alpha=0}^{N_h-1} h_\alpha h_{\alpha+i-j}$$
$$+ 2\sum_{\alpha=0}^{N_p-1}\sum_{\beta=0}^{N_p-1} \underbrace{h_{m_0+\alpha-i}h_{m_0+\beta-j}\mathbf{A}(\alpha,\beta)}_{\alpha\neq\beta} + N_w\sigma_n^4\delta_{ij} + 2\sigma_n^4\delta_{ij}$$
$$+ N_w\sigma_n^2\sum_{\alpha=0}^{N_h-1} h_\alpha h_{\alpha+i-j} + 4\sigma_n^2\sum_{\alpha=0}^{N_h-1} h_\alpha h_{\alpha+i-j} + N_w\sigma_n^2\sum_{\alpha=0}^{N_h-1} h_\alpha^2\delta_{ij}$$
$$+ N_w\sum_{\alpha=0}^{N_h-1} h_\alpha h_{\alpha+i-j}\sum_{\beta=0}^{N_h-1} h_\beta^2 + 2\sum_{\gamma=0}^{N_w-1}\sum_{\alpha=0}^{N_h-1}\sum_{\beta=0}^{N_h-1} \underbrace{h_\alpha h_\beta h_{\alpha+i-\gamma}h_{\beta+j-\gamma}}_{\beta\neq\alpha+i-j},$$

$$\mathbf{Y}_{12}(i,j) = -\mathrm{tr}(\mathbf{A})h_{m_0+j-i} - 2\sum_{\alpha=0}^{N_p-1} \underbrace{h_{m_0+\alpha-i}\mathbf{A}(\alpha,j)}_{\alpha\neq j}$$
$$- N_w\sigma_n^2 h_{m_0+j-i} - 2\sigma_n^2 h_{m_0+j-i} - N_w h_{m_0+j-i}\sum_{\alpha=0}^{N_h-1} h_\alpha^2$$
$$- 2\sum_{\alpha=0}^{N_h-1} \underbrace{h_\alpha \sum_{\beta=0}^{N_w-1} h_{i+\alpha-\beta}h_{m_0+j-\beta}}_{\alpha\neq m_0+j-i},$$

$$\mathbf{Y}_{22}(i,j) = [\mathrm{tr}(\mathbf{A}) + N_w\sigma_n^2 + N_w\sum_{\alpha=0}^{N_h-1} h_\alpha^2]\delta_{ij}$$
$$+ [2\mathbf{A}(i,j) + 2\sum_{\alpha=0}^{N_w-1} h_{m_0+i-\alpha}h_{m_0+j-\alpha}][1-\delta_{ij}].$$