Analysis of the Hierarchical LMS Algorithm

Vítor H. Nascimento, *Member, IEEE*

***Abstract*—We analyze the recently proposed hierarchical least mean-square (HLMS) algorithm, providing expressions for its steady-state mean-square error (MSE). We find conditions for the hierarchical structure to be equivalent to the optimal (full-length) Wiener solution. When these conditions are not satisfied, we show that HLMS will compute biased estimates. Our analysis also shows that even when these conditions hold, the MSE obtained using HLMS may be much larger than that obtained using LMS, since the potentially large MSEs at the subfilters in the first hierarchical level directly affect the output MSE.**

***Index Terms*—Adaptive filters, estimation, least mean-square methods, least square methods, stochastic systems.**

I. Introduction

As IS WELL KNOWN, the least mean-square (LMS) algorithm has several desirable properties, which explains the algorithm's widespread use: LMS is easily implemented, has a low computational cost, is robust to numerical errors, and has good tracking performance. Its one drawback is its slow initial convergence, especially in situations where there is strong correlation between the entries of the regressor vector [1], [2].

An approach to fight this problem was proposed recently: the hierarchical LMS (HLMS) algorithm [6], [7]. This method attempts to obtain a faster convergence by splitting the LMS filter into several (level-1) independent LMS subfilters (see Fig. 1). The output of each subfilter is sent to another (level-2) LMS subfilter, which weighs each subfilter output to obtain an overall output (other intermediate levels may also be used).

In this letter, we present an analysis of the HLMS algorithm, showing under which conditions the overall filter response converges to the Wiener solution. We also provide a stochastic analysis and derive expressions for the overall mean-square error (MSE) under these conditions. Our analysis shows that in many situations, the HLMS algorithm converges in the mean to a *biased* estimate of the optimal Wiener solution. In addition, there is a tendency for the level-2 subfilter to amplify the misadjustment error in the level-1 subfilters, which results in a high output MSE.

II. HLMS Algorithm

To keep the analysis short, we discuss in this letter only a two-level version of the HLMS algorithm (Fig. 1), although more levels would be possible [6]. In the figure, level 1 has two length-3 subfilters, and level 2 consists of a single length-2 subfilter. The overall filter length is $M = 6$.

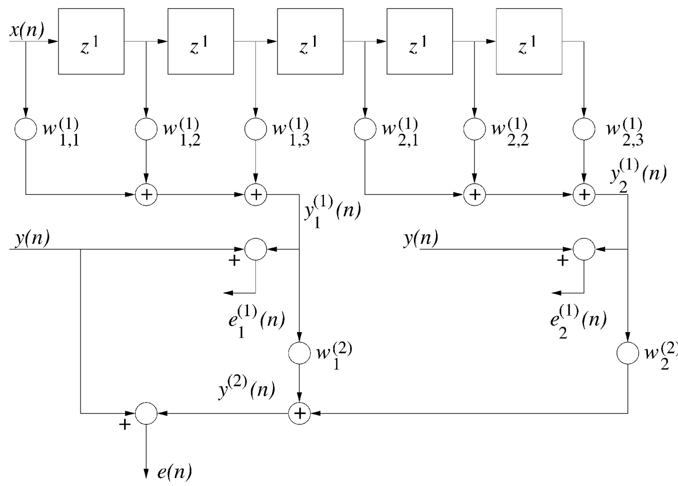

Fig. 1. HLMS structure with two hierarchical levels (level 1: two length-3 subfilters, level 2: one length-2 subfilter).

In general, we split a filter with $M = M_1 \cdot M_2$ taps into (level 1) M_2 length-M_1 subfilters. The ith subfilter in level 1 has coefficient vector $\boldsymbol{W}_i^{(1)}(n) = [w_{i,1}^{(1)}(n) \ \dots w_{i,M_1}^{(1)}(n)]^T$ (superscript ^{T} denotes vector transposition). The output of the ith level-1 subfilter is $y_i^{(1)}(n)$, and the error is $e_i^{(1)}(n)$. The single subfilter in level 2 has coefficients $\boldsymbol{W}^{(2)}(n) = [w_1^{(2)}(n) \ \dots \ w_{M_2}^{(2)}(n)]^T$, output $y^{(2)}(n)$, and error $e(n)$.

The inputs to the filter are the sequences $\{x(n)\}$ (the regressor sequence) and $\{y(n)\}$ (the desired sequence), which we assume to be real and zero-mean. From the regressor sequence, we define the vectors $\boldsymbol{X}(n) = [x(n) \ \dots \ x(n-M+1)]^T$, $\boldsymbol{X}^{(2)}(n) = [y_1^{(1)}(n) \ \dots \ y_{M_2}^{(1)}(n)]^T$, and split $\boldsymbol{X}(n) = [\boldsymbol{X}_1^{(1)^T}(n) \ \dots \ \boldsymbol{X}_{M_2}^{(1)^T}(n)]^T$, where each $\boldsymbol{X}_i^{(1)}(n)$ is a length-M_1 vector. With these definitions, we have $y_i^{(1)}(n) = \boldsymbol{W}_i^{(1)^T}(n)\boldsymbol{X}_i^{(1)}(n)$, $e_i^{(1)}(n) = y(n) - y_i^{(1)}(n)$, $y^{(2)}(n) = \boldsymbol{W}^{(2)^T}(n)\boldsymbol{X}^{(2)}(n)$, and $e(n) = y(n) - y^{(2)}(n)$. The update equations for each subfilter is

$$\boldsymbol{W}_i^{(1)}(n+1) = \boldsymbol{W}_i^{(1)}(n) + \mu_1 e_i^{(1)}(n)\boldsymbol{X}_i^{(1)}(n) \qquad (1)$$

$$\boldsymbol{W}^{(2)}(n+1) = \boldsymbol{W}^{(2)}(n) + \mu_2 e(n)\boldsymbol{X}^{(2)}(n) \qquad (2)$$

where μ_1 and μ_2 are the step sizes for levels 1 and 2, respectively.

III. Optimal Overall Solution

In this section, we find the optimal (Wiener) solution for each subfilter, and we compare the optimal overall filter with the full

Manuscript received August 9, 2002; revised August 14, 2002. This work was supported by FAPESP—São Paulo State Research Council under Grant 2000/09569-6 and by CNPq—Brazilian Research Council.

The author is with the Electronic Systems Engineering Department, Escola Politécnica, Universidade de São Paulo, São Paulo 05508-900, Brazil (e-mail: vitor@lps.usp.br).

Digital Object Identifier 10.1109/LSP.2002.807863

length Wiener solution. That is, we use the hierarchical filter structure, but choose the best weights for each subfilter, assuming knowledge of the statistics of the input signals, without adaptation.

We begin by defining the regressor covariance matrix R

$$R \triangleq \mathrm{E}(\boldsymbol{X}(n)\boldsymbol{X}^T(n)) = \begin{bmatrix} R_1 & R_{1,2} & \dots & R_{1,M_2} \\ R_{1,2}^T & R_2 & \dots & R_{2,M_2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{1,M_2}^T & R_{2,M_2}^T & \dots & R_{M_2} \end{bmatrix}$$

where $\mathrm{E}(\cdot)$ represents the expectation operator, $R_i \triangleq \mathrm{E}(\boldsymbol{X}_i^{(1)}(n)\boldsymbol{X}_i^{(1)^T}(n))$, and $R_{i,j} \triangleq \mathrm{E}(\boldsymbol{X}_i^{(1)}(n)\boldsymbol{X}_j^{(1)^T}(n))$. Note that since R is a symmetric matrix, $R_{i,j} = R_{j,i}^T$.

It is well known from linear estimation theory that, given two zero-mean stationary sequences $\{\boldsymbol{X}(n)\}$ and $\{y(n)\}$, there is a vector $\boldsymbol{\Omega}$ such that [4]

$$y(n) = \boldsymbol{\Omega}^T \boldsymbol{X}(n) + v(n) \tag{3}$$

where $v(n)$ is a noise uncorrelated with $\boldsymbol{X}(n)$, and with variance $\sigma_0^2 = \mathrm{E}y^2(n) - \boldsymbol{\Omega}^T R \boldsymbol{\Omega}$. The vector $\boldsymbol{\Omega}$ may be computed by solving the system of $M = M_1 \cdot M_2$ linear equations

$$\mathrm{E}(\boldsymbol{X}(n)\boldsymbol{X}^T(n))\boldsymbol{\Omega} = \mathrm{E}(\boldsymbol{X}(n)y(n)). \tag{4}$$

These results can also be applied to the M_2 level-1 subfilters, as follows: assume that $\boldsymbol{\Omega}$ satisfies (3) and (4). Then, in order to compute the optimal solution $\boldsymbol{\Omega}_i^{(1)}$ for the ith level-1 subfilter, we need R_i and

$$\begin{aligned} \boldsymbol{p}_i^{(1)} &\triangleq \mathrm{E}(\boldsymbol{X}_i^{(1)}(n)y(n)) = \mathrm{E}\left[\boldsymbol{X}_i^{(1)}(n)(\boldsymbol{X}^T(n)\boldsymbol{\Omega} + v(n))\right] \\ &= [\, R_{i,1} \quad \dots \quad R_{i,i-1} \quad R_i \quad R_{i,i+1} \quad \dots \quad R_{i,M_2} \,]\, \boldsymbol{\Omega}. \end{aligned}$$

R_i is invertible if R is nonsingular [3]; therefore, the optimal weight vector $\boldsymbol{\Omega}_i^{(1)}$ for the ith level-1 subfilter is given by

$$\boldsymbol{\Omega}_i^{(1)} = R_i^{-1}\, [\, R_{i,1} \quad \dots \quad R_i \quad \dots \quad R_{i,M_2} \,]\, \boldsymbol{\Omega}.$$

If we split $\boldsymbol{\Omega}$ in M_2 subvectors, such that $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_1^T \ \dots \ \boldsymbol{\Omega}_{M_2}^T]^T$, then

$$\boldsymbol{\Omega}_i^{(1)} = \boldsymbol{\Omega}_i + \sum_{k=1, k\neq i}^{M_2} R_i^{-1} R_{i,k} \boldsymbol{\Omega}_k. \tag{5}$$

Note that, in general, $\boldsymbol{\Omega}_i^{(1)} \neq \boldsymbol{\Omega}_i$.

The optimal level-2 regressor vector is therefore given by

$$\boldsymbol{X}^{(2)}(n) = \left[\boldsymbol{X}_1^{(1)^T}(n)\boldsymbol{\Omega}_1^{(1)} \ \dots \ \boldsymbol{X}_{M_2}^{(1)^T}(n)\boldsymbol{\Omega}_{M_2}^{(1)}\right]. \tag{6}$$

In order to compute the optimal level-2 coefficients, we must evaluate $S \triangleq \mathrm{E}(\boldsymbol{X}^{(2)}(n)\boldsymbol{X}^{(2)^T}(n))$

$$S = \begin{bmatrix} \boldsymbol{\Omega}_1^{(1)^T} R_1 \boldsymbol{\Omega}_1^{(1)} & \dots & \boldsymbol{\Omega}_1^{(1)^T} R_{1,M_2} \boldsymbol{\Omega}_{M_2}^{(1)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}_{M_2}^{(1)^T} R_{1,M_2}^T \boldsymbol{\Omega}_1^{(1)} & \dots & \boldsymbol{\Omega}_{M_2}^{(1)^T} R_{M_2} \boldsymbol{\Omega}_{M_2}^{(1)} \end{bmatrix} \tag{7}$$

and also the cross correlation $\boldsymbol{p}^{(2)} \triangleq \mathrm{E}(\boldsymbol{X}^{(2)}(n)y(n))$

$$\boldsymbol{p}^{(2)} = \left[\boldsymbol{\Omega}_1^{(1)^T} R_1 \boldsymbol{\Omega}_1^{(1)} \quad \dots \quad \boldsymbol{\Omega}_{M_2}^{(1)^T} R_{M_2} \boldsymbol{\Omega}_{M_2}^{(1)}\right]^T. \tag{8}$$

The optimal level-2 coefficient vector $\boldsymbol{\Omega}^{(2)}$ must satisfy $S\boldsymbol{\Omega}^{(2)} = \boldsymbol{p}^{(2)}$. (We did not write $\boldsymbol{\Omega}^{(2)} = S^{-1}\boldsymbol{p}^{(2)}$, since in general S may be singular. When the filters are allowed to adapt, S becomes nonsingular.)

The optimal subfilters for levels 1 and 2 are equivalent to a single length-M filter, $\boldsymbol{\Omega}_{\mathrm{eq}}$. Denoting the entries of $\boldsymbol{\Omega}^{(2)}$ by $\omega_l^{(2)}$, for $1 \leq l \leq M_2$, we have

$$\boldsymbol{\Omega}_{\mathrm{eq}} = \left[\omega_1^{(2)}\boldsymbol{\Omega}_1^{(1)^T} \quad \dots \quad \omega_{M_2}^{(2)}\boldsymbol{\Omega}_{M_2}^{(1)^T}\right]^T.$$

In general, this equivalent filter will *not* be equal to the optimal length-M filter $\boldsymbol{\Omega}$. Thus, the optimal coefficients for the HLMS algorithm will result in an MSE *larger* than that obtained by the optimal solution, $\boldsymbol{\Omega}$. There are situations, however, for which the optimal HLMS coefficients result in an equivalent overall filter that is equal to $\boldsymbol{\Omega}$. One case is when only one of the $\boldsymbol{\Omega}_i$ is nonzero—i.e., when the optimal filter has an effective length equal to M_1. The other situation is when $R_{i,j} = 0$ for $i \neq j$. The first case would result in S being singular, so for now we consider only the second.

When $R_{i,j} = 0$ for $i \neq j$, the expressions for $\boldsymbol{\Omega}_i^{(1)}$ (5) and S (7) reduce to

$$\boldsymbol{\Omega}_i^{(1)} = \boldsymbol{\Omega}_i, \qquad S = \begin{bmatrix} \boldsymbol{\Omega}_1^T R_1 \boldsymbol{\Omega}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{\Omega}_{M_2}^T R_{M_2} \boldsymbol{\Omega}_{M_2} \end{bmatrix}.$$

Expression (8) for $\boldsymbol{p}^{(2)}$ does not change, and thus, when S is invertible, $\boldsymbol{\Omega}^{(2)} = [1 \ 1 \ \dots \ 1]^T$, and $\boldsymbol{\Omega}_{\mathrm{eq}} = \boldsymbol{\Omega}$.

IV. Stochastic Analysis of the HLMS Algorithm

In order to continue our analysis of the HLMS algorithm, we now evaluate the output MSE, i.e.,

$$J(n) \triangleq \mathrm{E}e(n)^2 = \mathrm{E}(y^{(2)}(n) - y(n))^2$$

with the usual independence assumptions [2] (recall that—for LMS—these assumptions give reasonable approximations for small step sizes [5]).

1) The sequences $\{y(n)\}$ and $\{\boldsymbol{X}(n)\}$ are independent, identically distributed (i.i.d.).
2) The noise sequence $\{v(n) = y(n) - \boldsymbol{\Omega}^T\boldsymbol{X}(n)\}$ is also i.i.d. and is independent of $\{\boldsymbol{X}_n\}$.
   In addition to the independence assumptions, we assume that $\boldsymbol{\Omega}_{\mathrm{eq}} = \boldsymbol{\Omega}$.
3) $R_{i,j} = 0$ for $i \neq j$, so that $\boldsymbol{\Omega}_{\mathrm{eq}} = \boldsymbol{\Omega}$, and $\boldsymbol{\Omega}_i^{(1)} = \boldsymbol{\Omega}_i$.

A. Analysis of Level-1 Subfilters

Under the above assumptions, we can use standard independence theory results to evaluate $\boldsymbol{J_i}^{(1)}(n) \triangleq \mathrm{E}(e_i^{(1)}(n))^2 = \mathrm{E}(y_i^{(1)}(n) - y(n))^2$. For this, we need to compute the minimum variance of the estimation error for each of the level-1 subfilters, i.e., [defining $v_i^{(1)}(n) = y(n) - \boldsymbol{\Omega}_i^{(1)^T}\boldsymbol{X}_i^{(1)}(n) = y(n) - \boldsymbol{\Omega}_i^T\boldsymbol{X}_i^{(1)}(n)$]

$$\begin{aligned} J_{0,i}^{(1)} &= \mathrm{E}(v_i^{(1)}(n))^2 = \mathrm{E}(y(n) - \boldsymbol{\Omega}_i^T\boldsymbol{X}_i^{(1)}(n))^2 \\ &= \mathrm{E}(\boldsymbol{\Omega}^T\boldsymbol{X}(n) + v(n) - \boldsymbol{\Omega}_i^T\boldsymbol{X}_i^{(1)}(n))^2. \end{aligned}$$

From assumption 3) above, (5), and recalling that $v(n)$ is independent of $\boldsymbol{X}(n)$ [and thus also of $\boldsymbol{X}_i^{(1)}(n)$], we have

$$J_{0,i}^{(1)} = \sigma_0^2 + \sum_{k=1,k\neq i}^{M_2} \boldsymbol{\Omega}_k^T R_i \boldsymbol{\Omega}_k. \tag{9}$$

Note that $J_{0,i}^{(1)}$ may be orders of magnitude larger than σ_0^2.

When we try to approximate $\boldsymbol{\Omega}_i^{(1)} = \boldsymbol{\Omega}_i$ using the LMS recursion (1) with step size μ_1, the estimation error will be, under the above assumptions, larger than $J_{0,i}^{(1)}$. Using standard results from LMS theory [1] we obtain

$$\lim_{n\to\infty} J_i^{(1)}(n) \approx J_{0,i}^{(1)}\left(1+\frac{\mu_1 \operatorname{Tr}(R_i)}{2}\right) \tag{10}$$

where $\operatorname{Tr}(A)$ is the trace of matrix A, i.e., the sum of the main diagonal elements. Although (10) is approximate (it holds only for "sufficiently small" μ_1), it shows that the excess MSE $J_{\mathrm{ex},i}^{(1)}$ is proportional to the optimum LMS error $J_{0,i}^{(1)}$

$$J_{\mathrm{ex},i}^{(1)} \approx \frac{\mu_1 \operatorname{Tr}(R_i)}{2} J_{0,i}^{(1)}. \tag{11}$$

This has an important effect on the behavior of the HLMS algorithm, as we shall see soon.

B. Analysis of the Level-2 Subfilter

Using (1) to estimate $\boldsymbol{\Omega}^{(1)}$, the statistics of $y_i^{(1)}(n)$ are no longer stationary, which makes the analysis considerably more involved. We study here only the behavior of the level-2 subfilter after steady state is reached, i.e., we assume that the level-1 subfilters have already converged, with MSE given by (10). With this assumption, we must find the new steady-state level-2 autocorrelation matrix S_a. Defining $\Delta\boldsymbol{w}_i^{(1)}(n) = \boldsymbol{\Omega}_i^{(1)} - \boldsymbol{w}_i^{(1)}(n)$, and recalling that under our assumptions $\boldsymbol{X}_i^{(1)}(n)$ is independent of $\Delta\boldsymbol{w}_i^{(1)}(n)$, in steady state

$$\begin{aligned}\mathrm{E}\left(y_i^{(1)}(n)\right)^2 &= \mathrm{E}\left[\left(\boldsymbol{\Omega}_i - \Delta\boldsymbol{w}_i^{(1)}(n)\right)^T \boldsymbol{X}_i^{(1)}(n)\right]^2\\ &= \boldsymbol{\Omega}_i^T R_i \boldsymbol{\Omega}_i + \\ &\quad + \operatorname{Tr}\left(R_i \mathrm{E}\left(\Delta\boldsymbol{w}_i^{(1)}(n)\Delta\boldsymbol{w}_i^{(1)^T}(n)\right)\right).\end{aligned}$$

The second term in this expression is exactly the steady-state excess MSE given by (11). Since the $J_{\mathrm{ex},i}^{(1)}$ are strictly positive, S_a is invertible and equal to

$$S_a = \begin{bmatrix} \boldsymbol{\Omega}_1^T R_1 \boldsymbol{\Omega}_1 + J_{\mathrm{ex},1}^{(1)} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \boldsymbol{\Omega}_{M_2}^T R_{M_2} \boldsymbol{\Omega}_{M_2} + J_{\mathrm{ex},M_2}^{(1)} \end{bmatrix}.$$

The cross correlation vector $\boldsymbol{p}^{(2)}$ does not change and is still given by (8). The optimal level-2 weights are now

$$\boldsymbol{\Omega}_a^{(2)} = \left[\frac{\boldsymbol{\Omega}_1^T R_1 \boldsymbol{\Omega}_1}{\boldsymbol{\Omega}_1^T R_1 \boldsymbol{\Omega}_1 + J_{\mathrm{ex},1}^{(1)}} \quad \cdots \quad \frac{\boldsymbol{\Omega}_{M_2}^T R_{M_2} \boldsymbol{\Omega}_{M_2}}{\boldsymbol{\Omega}_{M_2}^T R_{M_2} \boldsymbol{\Omega}_{M_2} + J_{\mathrm{ex},M_2}^{(1)}}\right]^T. \tag{12}$$

This expression shows that the level-2 subfilter gives less weight to level-1 subfilters with large excess MSE (in comparison to the optimum variance of $y_i^{(1)}(n)$). While this seems a sensible thing to do, it also implies that the HLMS algorithm converges to a *biased* overall solution, even if the hierarchical structure can describe the optimal Wiener solution. The bias worsens as μ_1 is increased.

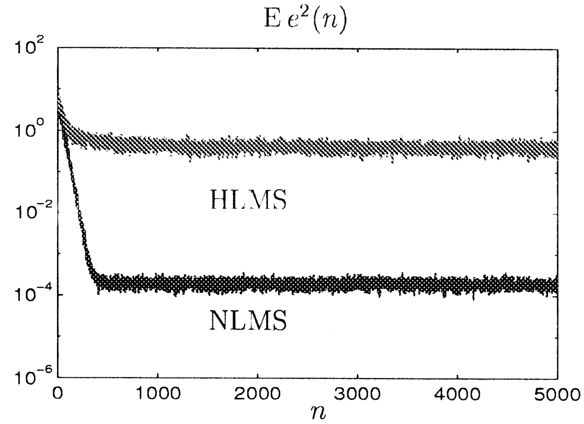

Fig. 2. HLMS applied to i.i.d. inputs, with step sizes $\mu_1 = \mu_2 = 10^{-3}$. For comparison, we also plot the NLMS learning curve with $\mu = 1$. Average of 50 curves.

Since the optimum estimation error variance $J_{0,i}^{(1)}$ of the level-1 subfilters may be much larger than σ_0^2, this bias may be quite significant unless μ_1 is maintained relatively small. This is contrary to the goal of the hierarchical structure, which is to obtain faster convergence. The examples in Section V show that this problem may be serious.

Before turning to the examples, let us find an approximate expression for the overall MSE for HLMS. First, we assume that the level-2 coefficients are fixed and equal to the optimum ones given by (12), and find the minimum MSE obtainable when the level-1 subfilters adapt

$$\begin{aligned}J_0^{(2)} &= \mathrm{E}\left(y(n) - y^{(2)}(n)\right)^2 = \mathrm{E}y^2(n) - \boldsymbol{\Omega}_a^{(2)^T}\boldsymbol{p}^{(2)}\\ &= \sigma_0^2 + \sum_{i=1}^{M_2} \frac{\boldsymbol{\Omega}_i^T R_i \boldsymbol{\Omega}_i}{\boldsymbol{\Omega}_i^T R_i \boldsymbol{\Omega}_i + J_{\mathrm{ex},i}^{(1)}} J_{\mathrm{ex},i}^{(1)}.\end{aligned} \tag{13}$$

The inputs to the level-2 subfilter are not i.i.d. even if $\{\boldsymbol{X}(n)\}$ is an i.i.d. sequence. However, for small μ_2 we may approximate the level-2 MSE using the independence theory formula, which gives

$$J^{(2)} \approx J_0^{(2)}\left(1+\frac{\mu_2 \operatorname{Tr}(S_a)}{2}\right). \tag{14}$$

V. Simulations

Our first example is a length-32 filter with $M_1 = 4$ and $M_2 = 8$. The input sequence $\{\boldsymbol{X}(n)\}$ is Gaussian, with autocorrelation $R = I$. The optimum vector $\boldsymbol{\Omega}$ has entries $(\Omega)_k = e^{-0.1(k-1)}$, and the noise $v(n)$ has variance 10^{-4}. The learning curves for HLMS with step sizes $\mu_1 = \mu_2 = 10^{-3}$ and also for normalized-LMS with $\mu = 1$ are shown in Fig. 2. Using expressions (9), (11), (13), and (14), we conclude that $J_{0,i}^{(1)}$ varies from 2.5–5.5. $J_{\mathrm{ex},i}^{(1)}$ varies from 0.049–0.11. The theoretical output MSE is $J^{(2)} = 0.4250$; from the ensemble-average learning curve, we obtained $J^{(2)} \approx 0.4085$. If μ_1 is reduced to 10^{-4},

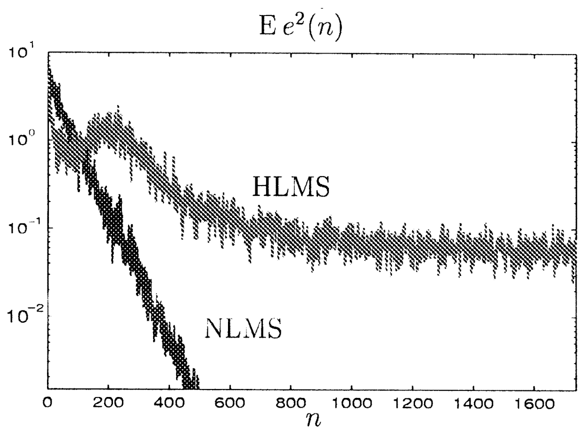

Fig. 3. HLMS applied to equalization problem, with step sizes $\mu_1 = 0.005$ and $\mu_2 = 0.05$, and NLMS with $\mu = 1$. Average of 30 curves.

from (14) we find $J^{(2)} = 7.7 \cdot 10^{-3}$. Simulating this situation, an experimental $J^{(2)} \approx 7.1 \cdot 10^{-3}$ was obtained.

Next, we have a channel equalization example. The channel has a transfer function $C(z) = 1/(1 - 0.7z^{-1} + 0.2z^{-5})$ and a white Gaussian input with unitary variance, with Gaussian output noise with variance 10^{-4}.

Notice that in this example, $R_{ij} \neq 0$, so our analysis in Section IV does not hold. Note, however, how the HLMS algorithm, although converging initially very fast, has $\mathrm{E}e^2(n)$ increasing between $50 \leq n \leq 200$. Later, the algorithm starts reducing $\mathrm{E}e^2(\cdot)$ again, but stops at a higher level than NLMS. This kind of behavior can be seen in [6, Figs. 2 and 3], but for the curve labeled "LMS". There was perhaps a mislabeling in [6], since this would be a very unusual behavior for LMS, which is well known to converge monotonically [2].

VI. Conclusion

We analyzed the performance of the recently proposed HLMS algorithm, providing expressions for the steady-state MSE under certain conditions. Our analysis shows that in general, the HLMS algorithm computes biased estimates for the optimum length-M estimation filter. This bias may be quite large and worsens as the convergence speed of the level-1 subfilters is increased. This means that in some applications, the reason for using HLMS (faster convergence) may be achieved only at the cost of a much larger MSE. However, other simulations show that, in a situation where the noise level is very high, this effect may be masked, making HLMS' performance reasonable.

References

[1] P. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Norwell, MA: Kluwer Academic, 1997.
[2] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[3] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, MA: Cambridge Univ. Press, 1987.
[4] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[5] J. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963–993, May–June 1979.
[6] T.-K. Woo, "Fast hierarchical least mean square algorithm," *IEEE Signal Processing Lett.*, vol. 8, pp. 289–291, Nov. 2001.
[7] ——, "HRLS: A more efficient RLS algorithm for adaptive FIR filtering," *IEEE Commun. Lett.*, vol. 5, pp. 81–84, Mar. 2001.