A JUSTIFICATION FOR THE IMPROVED PERFORMANCE OF THE MULTI-SPLIT LMS ALGORITHM

R. D. SOUZA, L. S. RESENDE

Communications Research Group
UFSC - Florianópolis - BRAZIL
{richard,leonardo}@eel.ufsc.br

M. G. BELLANGER

Laboratoire d'Electronique et Communication
CNAM - Paris - FRANCE
bellang@cnam.fr

ABSTRACT

This paper presents an analysis that justifies the improved performance of the multi-split LMS algorithm. It is shown that instead of reducing the eigenvalue ratio, the multi-split operation increases the diagonalization factor of the transformed input signal autocorrelation matrix, which assists the power normalized and time-varying step-size LMS algorithm used for updating the single parameters independently. Case studies and simulation results enable us to evaluate the improved performance of the multi-split LMS algorithm.

1. INTRODUCTION

Owing to its simplicity and robustness, the standard LMS algorithm is one of the most widely used algorithms for adaptive signal processing. Unfortunately, its performance in terms of convergence rate and tracking capability depends on the eigenvalue spread of the input signal autocorrelation matrix [3]. Transform domain LMS algorithms, like DCT and DFT, which reduce such a spread, have been used to solve this problem at the expense of a high computational complexity [2].

A low computational burden multi-split preprocessing of the input data vector has also been proposed for improving the performance of the LMS algorithm [1, 4]. After preprocessing, the adaptive FIR filter is realized as a set of zero-order filters connected in parallel, and with each single parameter independently updated.

Our contribution in this paper is to show that, instead of reducing the eigenvalue spread, the multi-split operation increases the diagonalization factor of the transformed input signal autocorrelation matrix. This fact leads us to assert that the improved performance of the multi-split LMS algorithm is due to its individual and independent updating characteristic by means of a power-normalized and time-varying step-size LMS algorithm.

In Section 2 we briefly present the multi-split LMS preprocessing using the linearly-constrained approach proposed in [1], and investigate its effect on the eigenvalue spread of the input data autocorrelation matrix. In Section 3, we consider a diagonalization factor measure and show that the multi-split operation allows us to perform an eigenvalue estimation and a subsequent step-size optmization. We also show that, in the particular case of a filter with only two coefficients, the multi-split LMS algorithm corresponds to the Newton LMS algorithm. In Section 4 we present simulation results that validate our analysis. Finally, in Section 5 we draw some final remarks.

2. MULTI-SPLIT LMS ALGORITHM

Consider the classical scheme of an adaptive transversal filter whose coefficients are represented by the $N \times 1$ vector $W(n)$, where $N = 2^M$. Without loss of generality, all the parameters are assumed to be real. Applying the multi-split operation as defined in [1], we arrive, after M steps with 2^{m-1} splitting operations $(m = 1, 2, \ldots, M)$, at the multi-split scheme shown in Figure 1, where C_{sm} and C_{am} are $2^{M-m+1} \times 2^{M-m}$ matrices, and $w_{\perp i}$, for $i = 0, 1, \ldots, N-1$, represent the single parameters of the resulting zero-order filters.

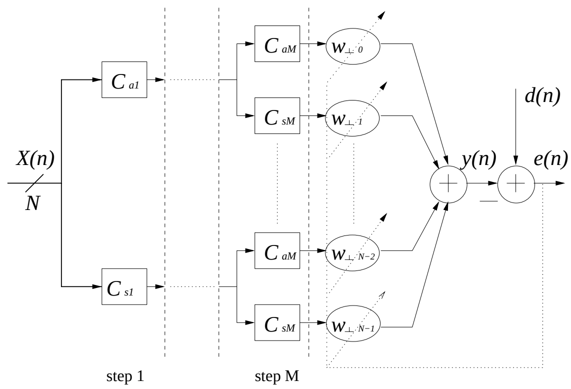

Fig. 1. Multi-split adaptive filtering.

The above multi-split scheme can be viewed as a linear transformation of the input data $X(n)$ denoted by:

$$X_{\perp}(n) = T^t X(n), \tag{1}$$

where:

$$T = \begin{bmatrix} C^t_{aM} C^t_{aM-1} \ldots C^t_{a1} \\ C^t_{sM} C^t_{aM-1} \cdots C^t_{a1} \\ C^t_{aM} C^t_{sM-1} \ldots C^t_{a1} \\ \ddots \\ C^t_{sM} C^t_{sM-1} \ldots C^t_{s1} \end{bmatrix}^t_{N \times N}, \tag{2}$$

is a non-singular matrix, and $T^t \times T = 2^M \cdot I$. The columns of T can be permuted in order to re-arrange the single parameters of Figure 1. One of these permutations is obtained by making:

$$C_{sm} = \begin{bmatrix} J_{M-m} \\ -I_{M-m} \end{bmatrix} \quad \text{and} \quad C_{am} = \begin{bmatrix} I_{M-m} \\ J_{M-m} \end{bmatrix} \tag{3}$$

for $m = 1, 2, \ldots, M$, where J is the reflection matrix. Applying (3) to (2), we get a linear transformation of $X(n)$ with a butterfly structure that is very suitable for VLSI implementation [1].

The autocorrelation matrix of $X_\perp$, $R_{X_\perp X_\perp}$, regardless of the transformation T in use, can be written as a function of the autocorrelation matrix of X, R_{XX}, as:

$$\begin{aligned} \mathrm{E}[X_\perp \cdot X_\perp^t] &= \mathrm{E}[T^t \cdot X \cdot X^t \cdot T] \\ &= 2^M \cdot T^{-1} \cdot \mathrm{E}[X \cdot X^t] \cdot T. \end{aligned} \tag{4}$$

Thus,

$$R_{X_\perp X_\perp} = 2^M \cdot T^{-1} \cdot R_{XX} \cdot T. \tag{5}$$

The matrix $R = T^{-1} \cdot R_{XX} \cdot T$ is said to be similar to R_{XX}, where T is a similarity transformation [6]. As a consequence:

$$\begin{aligned} \det(R - \lambda I) &= \det\left(T^{-1} \cdot (R_{XX} - \lambda I) \cdot T\right) \\ &= \det(R_{XX} - \lambda I), \end{aligned} \tag{6}$$

which means that matrix R has the same characteristic equation and the same eigenvalues as R_{XX}. Therefore:

$$\Lambda(R_{X_\perp X_\perp}) = 2^M \cdot \Lambda(R_{XX}), \tag{7}$$

where $\Lambda(\bullet)$ denotes the eigenvalue diagonal matrix of $\bullet$. Based on (7) we can say that the multi-split operation does not affect the eigenvalue spread of the input data correlation matrix.

According to [3], (7) reflects in the geometric ratios of the learning curves of the LMS algorithm. The ratios for the classical algorithm are given by:

$$r_i = 1 - 2\mu \cdot \lambda_i, \tag{8}$$

for $i = 0, 1 \ldots, N-1$, where μ is the step-size parameter, and convergence is guaranteed if [3]:

$$0 < \mu < \frac{1}{\lambda_{max}}. \tag{9}$$

Based on this and taking into account that the multi-split adaptive scheme has N filters with only one coefficient, we have:

$$r_{i_\perp} = 1 - 2^{M+1} \mu_{i_\perp} \cdot \lambda_i, \tag{10}$$

and convergence for each filter is guaranteed if:

$$0 < \mu_{i_\perp} < \frac{1}{2^M \cdot \lambda_i}. \tag{11}$$

Perfect knowledge of the eigenvalues in (7) allows us to optimize the choice of $\mu_{i_\perp}$ for each one of the N single coefficient filters. In [1] the authors proposed, without further analysis, the application of a power-normalized version of the LMS algorithm independently for each single coefficient:

$$w_{\perp i}(n) = w_{\perp i}(n-1) + \frac{\mu_s}{\tilde{r}_i(n)} x_{\perp i}(n) e(n) \tag{12}$$

where μ_s is a constant step-size, $e(n) = d(n) - y(n)$,

$$y(n) = \sum_{i=0}^{N-1} x_{\perp i}(n) w_{\perp i}(n-1), \tag{13}$$

and:

$$\tilde{r}_i(n) = \gamma \tilde{r}_i(n-1) + \frac{1}{n}(|x_{\perp i}(n)|^2 - \gamma \tilde{r}_i(n-1)), \tag{14}$$

for $i = 0, 1, \ldots, N-1$. When the autocorrelation matrix of $X_\perp$ is diagonal, the eigenvalues can be perfectly estimated, and the step-sizes (11) can be independently optimized. In this case, the multi-split LMS algorithm performas as well as the Newton LMS algorithm [3]. We analyze this case in the next section.

3. DIAGONALIZATION ANALYSIS

Consider a filter $W(n)$ with only 2 coefficients. To perform the multi-split operation in this case, we can apply the transformation:

$$T_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^t \tag{15}$$

to the input vector $X(n)$, whose autocorrelation matrix is

$$R_{XX} = \begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix}, \tag{16}$$

where $r_{01} = r_{10}$ and $r_{00} = r_{11}$. Then, the autocorrelation matrix of the split vector $X_\perp(n) = T^t \cdot X(n)$ is

$$R_{X_\perp X_\perp} = 2 \cdot \begin{bmatrix} r_{00} + r_{01} & 0 \\ 0 & r_{00} - r_{01} \end{bmatrix}. \tag{17}$$

The eigenvalues of the matrix $R_{X_\perp X_\perp}$ are $\lambda_1 = 2 \cdot (r_{00} + r_{01})$ and $\lambda_2 = 2 \cdot (r_{00} - r_{01})$, which is the same as the power in each tap of $W_\perp(n)$. Thus, equation (14) actually estimates the eigenvalues of $R_{X_\perp X_\perp}$, which, in matrix form, is:

$$\tilde{\Lambda} = \begin{bmatrix} \tilde{r}_0 & 0 \\ 0 & \tilde{r}_1 \end{bmatrix}. \tag{18}$$

Then, the updating equations (12) can be also written in matrix form as:

$$W_\perp(n) = W_\perp(n-1) + \mu_s \tilde{\Lambda}^{-1} X_\perp(n) e(n). \tag{19}$$

Note that, since in this case $\tilde{\Lambda}$ is a perfect estimate of the autocorrelation matrix, equation (19) is the same as the Newton LMS updating equation [3].

When the number of coefficients in the filter $W(n)$ is increased to the next powers of 2, the autocorrelation matrix $R_{X_\perp X_\perp}$ is not diagonal. It has the form:

$$R_{X_\perp X_\perp} = \begin{bmatrix} S_{K \times K} & 0_{K \times K} \\ 0_{K \times K} & A_{K \times K} \end{bmatrix}, \tag{20}$$

where $K = N/2$, and the eigenvalues of $R_{X_\perp X_\perp}$ are the combination of the eigenvalues of S and A, since $R_{X_\perp X_\perp}$ can be written as the direct sum $S \oplus A$ [6]. Even though matrix $R_{X_\perp X_\perp}$ is not diagonal, its *diagonalization factor* is increased. The diagonalization factor is defined in [4] as:

$$\gamma(R_{X_\perp X_\perp}) = \frac{\mathrm{trace}(R_{X_\perp X_\perp})}{\sum |\text{elements of } R_{X_\perp X_\perp}| - \mathrm{trace}(R_{X_\perp X_\perp})}. \tag{21}$$

In this case, equation (19) is an approximation of the Newton LMS. The closeness of this approximation can be measured by the increase in the diagonalization factor produced by the multi-split operation.

Theorem of the Circle of Gerschgorin [6]: Each eigenvalue λ (real or complex) of a matrix B, $n \times n$, satisfies at least one of the inequalities:

$$|\lambda - b_{ii}| \leq y_i, \quad \text{where} \quad y_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}| \quad (i = 1, \ldots, n). \tag{22}$$

Thus, increasing the trace of the matrix $R_{X_\perp X_\perp}$ by a factor that is larger than the increase of the sum of the module of the elements not in its main diagonal, increases the accuracy of the eigenvalues estimate given by equation (14). That is exactly the effect of the multi-split operation, as can be seen from comparing the diagonalization factor of R_{XX} and $R_{X_\perp X_\perp}$.

In the case of $N = 4$, the autocorrelation matrix R_{XX} is

$$\begin{bmatrix} r_{00} & r_{01} & r_{02} & r_{03} \\ r_{01} & r_{00} & r_{01} & r_{02} \\ r_{02} & r_{01} & r_{00} & r_{01} \\ r_{03} & r_{02} & r_{01} & r_{00} \end{bmatrix}.$$

Suppose that $|r_{01}| \geq |r_{02}| \geq |r_{03}|$. Then, due to the *Theorem of the Circle of Gerschgorin*, the minimum error on the estimate of the eigenvalues given by the main diagonal of R_{XX} is given by:

$$y_1 = |r_{01}| + |r_{02}| + |r_{03}|. \qquad (23)$$

The autocorrelation matrix $R_{X_\perp X_\perp}$ is

$$2 \cdot \begin{bmatrix} (2r_{00} + 3r_{01} + 2r_{02} + r_{03}) & (r_{03} - r_{01}) & 0 & 0 \\ (r_{03} - r_{01}) & (2r_{00} - r_{01} - 2r_{02} + r_{03}) & 0 & 0 \\ 0 & 0 & (2r_{00} + r_{01} - 2r_{02} - r_{03}) & (r_{03} - r_{01}) \\ 0 & 0 & (r_{03} - r_{01}) & (2r_{00} - 3r_{01} + 2r_{02} - r_{03}) \end{bmatrix},$$

and the error on the estimate of the eigenvalues given by the main diagonal of $R_{X_\perp X_\perp}$ are at most:

$$y_{\perp i} = 2 \times |r_{03} - r_{01}|, \qquad (24)$$

for $i = 1, 2, 3, 4$. While the eigenvalues of $R_{X_\perp X_\perp}$ are 2^2 times larger than the eigenvalues of R_{XX}, the estimation error of $\lambda_\perp$, in the worst case, is twice the estimation error of λ. So, the efficiency of the estimate given by equation (14) is considerably increased.

Example 1: Consider a similar case as the one in the Example of Figure 6.2 from [3, Chapter 6], where the input signal is $x(n) = \sin(2 \times \pi / L)$, the desired response is $d(n) = 2 \times \cos(2 \times \pi / L)$, the filter $W(n)$ has four coefficients, the power of the random signal added to $x(n)$ in the input of the filter is $\phi = 0.01$, and $L = 16$ samples per signal cycle. Then, the autocorrelation matrix of the input vector $X(n)$ has diagonalization factor:

$$\gamma = \frac{2.04}{4.566} = 0.446.$$

Applying the multi-split operation to the input vector $X(n)$ yields the split input vector $X_\perp(n)$, whose autocorrelation matrix has diagonalization factor:

$$\gamma_\perp = \frac{8.16}{2.164} = 3.77.$$

Thus, the diagonalization factor was increased by more than 8 times, and, according to the *Theorem of the Circle of Gercshgorin*, the accuracy of the estimate given by equation (14) was also considerably increased.

When $N > 4$, the analysis is not so simple, and it becomes tedious to analyze each case individually. But making use of equation (20), it can be shown that the split operation also increases the diagonalization factor for $N > 4$. It is also important to note that the improvement in the diagonalization factor γ is also a function of the amount of correlation in $X(n)$. In order to explore this idea, let us consider the following example.

Example 2: Consider that $x(n)$ is the output of the 2kft-AWG26 HDSL channel as presented in [5], when the input is a random signal with unit variance and zero mean, and $N = 8$. The autocorrelation matrix of $X(n)$ has diagonalization factor:

$$\gamma = \frac{11.22}{9.48} = 1.18,$$

while the autocorrelation matrix of vector $X_\perp(n)$ has diagonalization factor:

$$\gamma_\perp = \frac{89.76}{27.62} = 3.25.$$

Here, the diagonalization factor was increased 2.75 times, also improving considerably the estimate on the eigenvalues given by equation (14), but by a smaller factor than the improvement in Example 1. Note also that γ here is 2.65 times greater than in Example 1.

In the next section we will present simulations showing that the performance of the multi-split LMS algorithm can be seen as half-way between the classical LMS and the RLS algorithm performances. We use the RLS algorithm in the simulations since it can be considered as a stochastic approximation of the iterative deterministic Newton algorithm [7].

4. SIMULATION RESULTS

Now we make use of simulations to compare the performances of the Multi-Split LMS (MS), the classical LMS and the RLS algorithms.

In the following simulations, we consider an adaptive system as the one in Figure 2, where the channel input is binary, with $b(n) = \pm 1$, and $v(n)$ is an additive noise with variance $\sigma_v^2 = 0.001$.

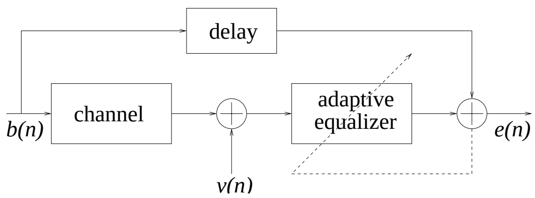

Fig. 2. Adaptive equalizer

In Figures 3, 4, and 5 we compare the ensemble mean square error (MSE) performances of the LMS, the MS and the RLS algorithms, when applied to the minimization of the estimation error in Figure 2, for three different channels. The step-sizes μ and μ_s, for the LMS and the MS algorithms, respectively, were set to $1/2N$.

In the first two cases we considered the raised cosine channel [2] with coefficients described by:

$$c_j = \begin{cases} \frac{1}{2}\left(1 + \cos\left(\frac{2\pi}{\alpha}(j-2)\right)\right), & j = 1, 2, 3. \\ 0, & \text{otherwise} \end{cases}, \qquad (25)$$

where α controls the eigenvalue spread χ of the autocorrelation matrix of the channel input vector $X(n)$, and N is the number of tap-weights of the equalizer.

Figure 3 presents the ensemble MSE when $\alpha = 2.9$ and $N = 8$, and Figure 4 presents the case where $\alpha = 3.5$ and $N = 8$. The poorer performance of the LMS algorithm in Figure 3 compared

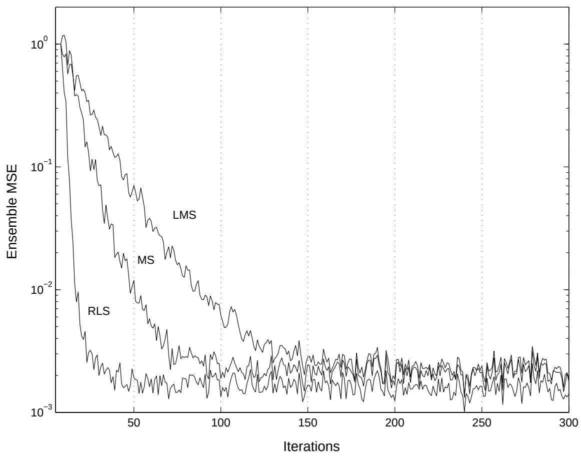

Fig. 3. Ensemble mean square error versus number of iterations. Raised cosine channel - α=2.9, N=8.

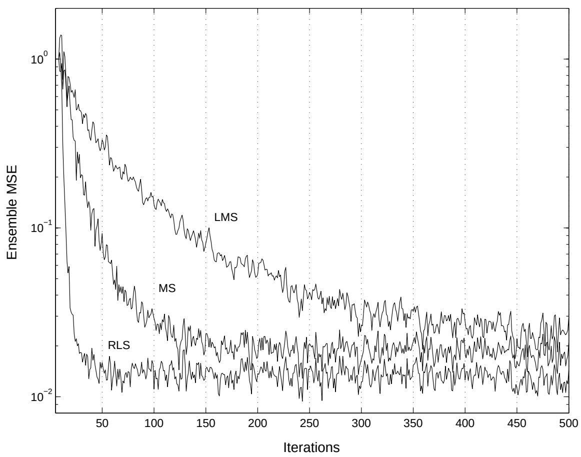

Fig. 4. Ensemble mean square error versus the number of iterations. Raised cosine channel - α=3.5, N=8.

with Figure 4 is due to the increase in the eigenvalues spread χ [2]. We can also observe that the performance of the MS algorithm relative to the RLS algorithm did not suffer from the increase in the eigenvalue spread.

In Figure 5 we considered the 2kft-AWG26 HDSL channel presented in [5]. The equalizer has $N = 32$ taps. This channel is the one that introduces the largest distortion and the largest eigenvalue spread among the three examples shown here. The relative performance of the MS algorithm to the RLS algorithm is about the same as in Figures 3 and 4, while the performance of the LMS algorithm was even more degradated.

5. FINAL REMARKS

We presented an analysis of the performance of the multi-split LMS algorithm. Our analysis is different from the one presented in [4] in the sense that we consider the linearly-constrained approach proposed in [1], and that we compare its performance with the LMS and the RLS algorithms.

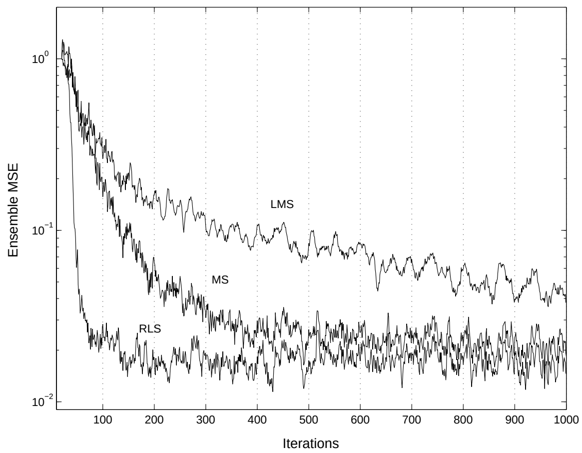

Fig. 5. Ensemble mean square error versus the number of iterations - 2kft-AWG26 HDSL channel - N=32.

It was shown that the multi-split operation increases the diagonalization factor of the input signal autocorrelation matrix, allowing us to perform an estimation of its eigenvalues from the input data. Based on this eigenvalues estimation we view the multi-split LMS algorithm as a power-normalized time-varying step-size algorithm.

For the special case of $N = 2$ coefficients, we showed that the multi-split and the Newton LMS algorithms are equivalent. Simulation results validate our analysis and confirm that the performance of the multi-split LMS is in between the RLS and the LMS algorithms.

The presented analytical approach provides additional insight into how the multi-split operation improves the performance of the LMS algorithm, justifying its choice as a powerful low complexity updating algorithm for adaptive filters.

6. REFERENCES

[1] L. S. Resende, J. M. T. Romano and M. G. Bellanger, “Multi-split adaptive filtering,” *IEEE ICASSP’01*, May 2001

[2] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, Third Edition, 1996.

[3] B. Widrow and S. D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, 1985.

[4] K. F. Wan and P. C. Ching, “Multilevel split-path adaptive filtering and its unification with discrete Walsh transform application,”*IEEE Trans. Circuits Syst II*, vol. 44, No.2, pp. 147-151, Feb. 1997.

[5] D. Yellin, A. Vardy and O. Amrani, “Joint equalization and coding for intersymbol interference channels,”*IEEE Trans. Inform. Theory*, Vol. 43, No. 2, pp. 409-425, Mar. 1997.

[6] B. Noble and J. W. Daniel. *Applied Linear Algebra*. Prentice-Hall, 1977.

[7] G-O. Glentis, K. Berberidis, and S. Theodoridis, “Efficient least squares adaptive algorithms for FIR transversal filtering,” *IEEE Signal Proc. Magazine*, July 1999.