The Binormalized Data-Reusing LMS Algorithm with Optmized Step-Size Sequence

J. A. Apolinário Jr.[1,2], M. L. R. de Campos[1], P. S. R. Diniz[2], and T. I. Laakso[3]

apolin@epq.ime.eb.br, campos@aquarius.ime.eb.br, diniz@lps.ufrj.br, and Timo.Laakso@hut.fi

[1]Instituto Militar de Engenharia
Depto. Eng. Elétrica
Pça Gal. Tibúrcio, 80
Praia Vermelha, Rio de Janeiro, RJ
22.290-270 — Brazil

[2]Universidade Federal do Rio de Janeiro
COPPE - Programa de Engenharia Elétrica
P. O. Box 68504
Rio de Janeiro, RJ
21945-970 — Brazil

[3]Helsinki University of Technology
Laboratory of Telecommunications Technology
P. O. Box 3000
Espoo
FIN-02015 TKK — Finland

Abstract

A new algorithm, the binormalized data-reusing least mean-squares (LMS) algorithm is presented. The new algorithm has been found to converge faster than other LMS-like algorithms, such as the Normalized LMS algorithm and several data-reusing LMS algorithms, in cases where the input signal is strongly correlated. The computational complexity of this new algorithm is only slightly higher than a recently proposed normalized new data-reusing LMS algorithm. Superior performance in convergence speed is, however, followed by a higher misadjustment if the step-size is close to the value which allows the fastest convergence. An optimal step-size sequence for this algorithm is proposed after considering a number of simplifying assumptions. Moreover, this work brings insight in how to deal with these conflicting requirements of fast convergence and minimum steady-state mean-square error (MSE)

1 Introduction

The least mean-squares (LMS) algorithm is very popular and has been widely used due to its simplicity. Nevertheless, its convergence speed is highly dependent on the eigenvalue spread of the input-signal autocorrelation matrix (ratio between the largest and the smallest eigenvalue also known as *condition number*) [1, 2]. Alternative schemes which try to improve convergence speed at the cost of minimum additional computational complexity have been proposed and extensively discussed in the past [1, 3, 4].

The data-reusing LMS (DR-LMS) algorithm, which uses current desired and input signals repeatedly within each iteration is one among such schemes. It can be easily shown that in the limit of infinite data reuses per iteration the DR-LMS and the normalized LMS (NLMS) algorithms yield the same solution [5]. Performance can be further improved with the recently proposed normalized and unnormalized new data-reusing LMS (NNDR-LMS and UNDR-LMS) algorithms [5]. These algorithms reuse the data pair, namely desired and input signals, from previous iterations as well.

In reference [5], a graphical description of the NNDR-LMS and UNDR-LMS algorithms was presented and it was shown that this new class of data-reusing algorithms had prospective better performance than the NLMS algorithm. The geometric description also showed why improvement is achieved when the number of reuses is increased. The new binormalized data-reusing LMS (BNDR-LMS) algorithm described here employs normalization on two orthogonal directions obtained from consecutive data pairs within each iteration. In all simulations carried out with colored input signals, the new algorithm presented faster convergence than all other algorithms mentioned above (case of two data pairs).

Convergence speed is related to the level of mean-squared error (MSE) in steady-state which is controlled by a step-size parameter μ. For $\mu = 1$, we have the fastest convergence and also the highest steady-state MSE when compared to the values of the step-size closer to zero. In [8], it was shown that the BNDR-LMS algorithm converges if the step-size is in the range from zero to two. For practical reasons, the value of μ is kept between zero and one since it was observed that the steady-state MSE was higher and the convergence slower when the step-size was set to a value between one and two. Only after [9] an analysis for the MSE behavior of the BNDR-LMS algorithm was available. In this paper, the expression for the MSE developed in [9] is used to derive an optimal step-size sequence which allows fast convergence and minimum misadjustment.

This paper is organized as follows: Section 2 presents LMS-like algorithms and a graphical illustration of their coefficient updating. Section 3 introduces the new BNDR-LMS algorithm as well as some remarks about its convergence behavior. In section 4, the optimal step-size sequence is derived and several approximations for this optimal sequence are also proposed. Section 5 contains simulation results and Section 6 presents conclusions.

2 LMS, DR-LMS, NLMS and NDR-LMS Algorithms

For the LMS algorithm, the coefficient vector $\boldsymbol{w}$ is updated in the opposite direction of the gradient vector obtained from instantaneous squared output error, i.e.,

$$\boldsymbol{w}_{LMS}(k+1) = \boldsymbol{w}_{LMS}(k) - \mu \nabla_w [e^2(k)] \tag{1}$$

where

$$e(k) = d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}_{LMS}(k) \tag{2}$$

is the output error, $d(k)$ is the desired signal, $\boldsymbol{x}(k)$ is the input-signal vector containing the $N+1$ most recent input-signal samples, and μ is the step size. The coefficient-updating equation is

$$\boldsymbol{w}_{LMS}(k+1) = \boldsymbol{w}_{LMS}(k) + \mu e(k)\boldsymbol{x}(k) \tag{3}$$

For the DR-LMS with L data reuses, the coefficients are updated as

$$\boldsymbol{w}_{i+1}(k) = \boldsymbol{w}_i(k) + \mu e_i(k)\boldsymbol{x}(k) \tag{4}$$

for $i = 0, \ldots, L$; where

$$e_i(k) = d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}_i(k), \tag{5}$$

$$\boldsymbol{w}_0(k) = \boldsymbol{w}_{DR-LMS}(k), \tag{6}$$

and

$$\boldsymbol{w}_{DR-LMS}(k+1) = \boldsymbol{w}_{L+1}(k). \tag{7}$$

Note that if $L = 0$ these equations correspond to the LMS algorithm and that μ is the step-size.

The NLMS algorithm normalizes the step-size such that the relation expressed by $\boldsymbol{x}^T(k)\boldsymbol{w}_{NLMS}(k+1) = d(k)$ is always satisfied, i.e.,

$$\boldsymbol{w}_{NLMS}(k+1) = \boldsymbol{w}_{NLMS}(k) + \frac{e(k)}{\boldsymbol{x}^T(k)\boldsymbol{x}(k) + \epsilon}\boldsymbol{x}(k) \tag{8}$$

where ϵ is a very small number used to avoid division by zero. Normalization for this algorithm implies a line search in the opposite direction of the gradient towards the minimum of the instantaneous squared output error.

For the NNDR-LMS algorithm with L data reuses, the coefficient vector is updated by the following relations:

$$\boldsymbol{w}_{i+1}(k) = \boldsymbol{w}_i(k) + \frac{e_i(k)}{\boldsymbol{x}^T(k-i)\boldsymbol{x}(k-i) + \epsilon}\boldsymbol{x}(k-i) \tag{9}$$

for $i = 0, \ldots, L$; where

$$e_i(k) = d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}_i(k), \tag{10}$$

$$\boldsymbol{w}_0(k) = \boldsymbol{w}_{NNDR-LMS}(k), \tag{11}$$

and

$$\boldsymbol{w}_{NNDR-LMS}(k+1) = \boldsymbol{w}_{L+1}(k). \tag{12}$$

For the sake of comparison, our interest is in one single reuse such that $L = 1$. Figure 1 illustrates geometrically the updating of the coefficient vector in a two-dimensional problem for all algorithms discussed above, starting from an arbitrary $\boldsymbol{w}(k)$.

Let $\mathcal{S}(k)$ denote the hyperplane which contains all vectors $\boldsymbol{w}$ such that $\boldsymbol{x}^T(k)\boldsymbol{w} = d(k)$. In a noise-free perfect-modeling situation, $\mathcal{S}(k)$ contains the optimal coefficient vector, $\boldsymbol{w}_o$. Furthermore, it can be easily shown that $\boldsymbol{x}(k)$ and, consequently, $\nabla_w[e^2(k)]$ are orthogonal to the hyperplane $\mathcal{S}(k)$.

The solution given by the DR-LMS algorithm, $\boldsymbol{w}(k+1)$, iteratively approaches $\mathcal{S}(k)$ by following the direction given by $\boldsymbol{x}(k)$ (see 3 in Figure 1). This solution would reach $\mathcal{S}(k)$ in the limit, as the number of data reuses, L, goes to infinity [5]. The NLMS algorithm performs a line search to yield the solution $\boldsymbol{w}(k+1) \in \mathcal{S}(k)$ in a single step (see 4 in Figure 1). The algorithms presented in [5] use more than one hyperplane, i.e., data pair $(\boldsymbol{x}, d)$, in order to produce a solution $\boldsymbol{w}(k+1)$ (see 5 and 6 in Figure 1) that is closer to $\boldsymbol{w}_o$ than the solution obtained with only the current data pair $(\boldsymbol{x}(k), d(k))$. For a noise-free perfect-modeling situation, $\boldsymbol{w}_o$ is at the intersection of $N+1$ hyperplanes constructed with linearly independent input-signal vectors. In this case, the orthogonal-projections algorithm [6] yields the solution $\boldsymbol{w}_o$ in $N+1$ iterations. This algorithm may be viewed as a normalized data-reusing orthogonal algorithm which utilizes $N+1$ data pairs $(\boldsymbol{x}, d)$.

In the next section, the new binormalized data-reusing LMS algorithm will be described. This algorithm combines data reusing, orthogonal projections of two consecutive gradient directions, and normalization in order to achieve faster convergence when compared to other LMS-like algorithms. At each iteration, the BNDR-LMS yields the solution $\boldsymbol{w}(k+1)$ which is at the intersection of hyperplanes $\mathcal{S}(k)$ and $\mathcal{S}(k-1)$ and at a minimum distance from $\boldsymbol{w}(k)$ (see 7 in Figure 1). The algorithm can also be viewed as a simplified version of the orthogonal projections algorithm which utilizes just two previous consecutive directions.

3 The BNDR-LMS Algorithm

In order to state the problem, we recall that the solution which belongs to $\mathcal{S}(k)$ and $\mathcal{S}(k-1)$ at a minimum distance from $\boldsymbol{w}(k)$ is the one that solves

$$\min_{\boldsymbol{w}(k+1)} \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 \tag{13}$$

subjected to

$$\boldsymbol{x}^T(k)\boldsymbol{w}(k+1) = d(k) \tag{14}$$

and

$$\boldsymbol{x}^T(k-1)\boldsymbol{w}(k+1) = d(k-1) \tag{15}$$

The functional to be minimized is, therefore,

$$\begin{aligned} f[\boldsymbol{w}(k+1)] \quad &= \quad [\boldsymbol{w}(k+1) - \boldsymbol{w}(k)]^T[\boldsymbol{w}(k+1) - \boldsymbol{w}(k)] \\ &\quad +\lambda_1[\boldsymbol{x}^T(k)\boldsymbol{w}(k+1) - d(k)] \\ &\quad +\lambda_2[\boldsymbol{x}^T(k-1)\boldsymbol{w}(k+1) - d(k-1)] \end{aligned} \tag{16}$$

which, for linearly independent input-signal vectors $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$, has the unique solution

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + (-\lambda_1/2)\boldsymbol{x}(k) + (-\lambda_2/2)\boldsymbol{x}(k-1) \tag{17}$$

where

$$-\lambda_1/2 = \frac{num1}{den} \tag{18}$$

and

$$-\lambda_2/2 = \frac{num2}{den} \tag{19}$$

with:

$$\begin{aligned}
num1 &= [d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\boldsymbol{x}^T(k-1)\boldsymbol{x}(k-1) \\
&\quad -[d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)]\boldsymbol{x}^T(k)\boldsymbol{x}(k-1) \\
num2 &= [d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)]\boldsymbol{x}^T(k)\boldsymbol{x}(k) \\
&\quad -[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\boldsymbol{x}^T(k-1)\boldsymbol{x}(k) \\
den &= \boldsymbol{x}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k-1) \\
&\quad -[\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2
\end{aligned} \tag{20}$$

It can be verified by simulations that the excess of mean-square error (MSE) for the BNDR-LMS algorithm is close to the variance of the measurement noise when there is no modeling error in a system-identification example. In order to control this excess of MSE, a step-size μ will be introduced. It is worth mentioning that the maximum convergence rate is usually obtained with $\mu = 1$. The BNDR-LMS algorithm is summarized in Table 1.

3.1 Geometrical Derivation

This algorithm can be alternatively derived from a purely geometrical reasoning. The first step is to reach a preliminary solution, $\boldsymbol{w}_1(k)$, which belongs to $\mathcal{S}(k)$ and is at a minimum distance from $\boldsymbol{w}(k)$. This is achieved by the NLMS algorithm starting from $\boldsymbol{w}(k)$, i.e.,

$$\boldsymbol{w}_1(k) = \boldsymbol{w}(k) + \frac{e(k)}{\boldsymbol{x}^T(k)\boldsymbol{x}(k)}\boldsymbol{x}(k) \tag{21}$$

In the second step, $\boldsymbol{w}_1(k)$ is updated in a direction orthogonal to the previous one, therefore belonging to $\mathcal{S}(k)$, until the intersection with $\mathcal{S}(k-1)$ is reached. This is achieved by the NLMS algorithm starting from $\boldsymbol{w}_1(k)$ and following the direction $\boldsymbol{x}_1^{\perp}(k)$ which is the projection of $\boldsymbol{x}(k-1)$ onto $\mathcal{S}(k)$.

$$\boldsymbol{w}(k+1) = \boldsymbol{w}_1(k) + \frac{e_1(k)}{\boldsymbol{x}_1^{\perp T}(k)\boldsymbol{x}_1^{\perp}(k)}\boldsymbol{x}_1^{\perp}(k) \tag{22}$$

where

$$\boldsymbol{x}_1^{\perp}(k) = \left[\mathbf{I} - \frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\boldsymbol{x}^T(k)\boldsymbol{x}(k)}\right]\boldsymbol{x}(k-1) \tag{23}$$

and

$$e_1(k) = d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}_1(k) \tag{24}$$

The use of $\boldsymbol{x}_1^{\perp}(k)$ obtained from $\boldsymbol{x}(k-1)$ assures that the minimum-distance path is chosen.

It is easy to show that if the BNDR-LMS algorithm was modified to utilize $k \mod (N+2)$ orthogonal directions, instead of two orthogonal directions, the resulting algorithm would be the orthogonal-projections algorithm described in [6] requiring reinitialization after every $N+1$ iterations.

Note that the requirement of linear independence of consecutive input-signal vectors $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$, necessary to ensure existence and uniqueness of the solution, is also manifested here. If $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$ are linearly dependent, then we cannot find $\boldsymbol{x}_1^{\perp}(k) \in \mathcal{S}(k)$ and the algorithm yields $\boldsymbol{w}(k+1) = w_1(k)$.

3.2 Mean-square Error Analyzis

Let us assume that an unknown FIR filter is to be identified by an adaptive filter of the same order, employing the BNDR-LMS algorithm. Input signal and measurement noise are assumed to be independent and identically distributed zero-mean white-noise with variances σ_x^2 and σ_n^2.

Assuming that the minimum mean-square error was caused by additive noise only, an expression for the MSE convergence behavior of the BNDR-LMS algorithm was obtained in [9] in terms of the excess in the MSE, defined as the difference between the MSE and the minimum MSE after convergence, i.e., $\Delta\xi(k) = \xi(k) - \xi_{min} = E[e^2(k)] - \sigma_n^2$.

$$\begin{aligned}
\Delta\xi(k+1) \quad &= \quad \left[1 + \frac{\mu(\mu-2)}{N+1}\right]\Delta\xi(k) \\
&+ \frac{N\mu(1-\mu)^2(\mu-2)}{(N+1)^2}\Delta\xi(k-1) \\
&+ \frac{[1+N(\mu-2)^2]\mu^2}{(N+1)(N+2-\nu_x)}\sigma_n^2
\end{aligned} \tag{25}$$

where ν_x is the *kurtosis* of the input signal.

4 Optimal Step-Size Sequence

In this section the optimal step-size sequence for the given problem is derived. We will follow an approach similar to that used in [4] assuming that up to time k the optimal sequences $\mu_\circ(0)$ to $\mu_\circ(k-1)$ and $\Delta\xi_\circ(0)$ to $\Delta\xi_\circ(k)$ are available. For the sake of simplicity, the *kurtosis* of the input signal is assumed equal to one in (25), i.e.,

$$\begin{aligned}\Delta\xi(k+1) &= \left[1+\frac{\mu(k)(\mu(k)-2)}{N+1}\right]\Delta\xi_{\circ}(k) \\ &+\frac{N\mu(k)(1-\mu(k))^2(\mu(k)-2)}{(N+1)^2}\Delta\xi_{\circ}(k-1) \\ &+\frac{(1+N(\mu(k)-2)^2)\mu(k)^2}{(N+1)^2}\sigma_n^2 \qquad (26)\end{aligned}$$

By differentiating $\Delta\xi(k+1)$ with respect to $\mu(k)$ and setting the result equal to zero, we obtain

$$\begin{aligned}\mu_{\circ}(k) &= 1-\sqrt{1-\frac{\Delta\xi_{\circ}(k)+\Delta\xi_{\circ}(k-1)}{2[\Delta\xi_{\circ}(k-1)+\sigma_n^2]}} \\ &= 1-\sqrt{1-\frac{\xi_{\circ}(k)+\xi_{\circ}(k-1)-2\sigma_n^2}{2\xi_{\circ}(k-1)}} \qquad (27)\end{aligned}$$

It is worth mentioning that (27) is in accordance with the situation when convergence is reached; in that case $\xi_{\circ}(k)=\xi_{\circ}(k-1)=\sigma_n^2$ and $\mu_{\circ}(k)=0$, as expected. Moreover, from the above relation, if $\sigma_n^2=0$ and admitting that $\Delta\xi_{\circ}(k)\approx\Delta\xi_{\circ}(k-1)$, $\mu_{\circ}(k)$ is close to one.

For the normalized LMS (NLMS) algorithm, a recursive formula for $\mu_{\circ}(k)$ in terms of $\mu_{\circ}(k-1)$ and the order N was obtained in [4]. For the BNDR-LMS algorithm, a routine based on (26) and (27) is presented in Table 2[1]. This routine has an important initialization parameter with a strong influence on the behavior of $\mu_{\circ}(k)$. This parameter is the ratio $\frac{\sigma_d^2}{\sigma_n^2}$ where the numerator is the variance of the reference signal (c.f. Figure 3).

[1]Note that, for simplicity, the *circle* ($\circ$) was dropped from the optimal values.

5 Simulation Results

In order to test the BNDR-LMS algorithm, simulations were carried out for a system identification problem. The system order N was equal to 10, the input signal was correlated noise such that the input-signal autocorrelation matrix had a conditioning number around 55 and input-signal to observation-noise ratio SNR equal to 150dB. The learning curves (MSE in dB) for the NLMS, the NNDR-LMS (one reuse), and the BNDR-LMS are depicted in Figure 2, corresponding to an average of 200 realizations.

In this example we can clearly verify the superior performance of the BNDR-LMS algorithm in terms of speed of convergence when compared to the NLMS and the NNDR-LMS (with one single reuse) algorithms. Simulations for the conventional LMS algorithm and for the DR-LMS algorithm were also carried out for the same setup, but their performances were, as expected, inferior compared to that of the NLMS algorithm and the results were omitted from Figure 2.

In order to test performance of the algorithms in terms of mean-square error after convergence, we measured the excess of MSE (MSE - MSE$_{min}$) in dB. The MSE$_{min}$ is the variance of the observation noise, set equal to 10^{-6} in this experiment. The results are summarized in Table 3 where we can also observe the excess of MSE in dB for a nonstationary environment. In this case, observation noise was set to zero and the system (plant) coefficients varied according to $\boldsymbol{w}_o(k) = \boldsymbol{w}_o(k-1) + \boldsymbol{v}$, where $\boldsymbol{v}$ is a vector whose elements were random numbers with zero mean and variance equal to 10^{-6}. As we can see from Table 3 the BNDR-LMS algorithm performed closely to the NLMS and the NNDR-LMS algorithms in both stationary and nonstationary environments.

In terms of computational complexity, Table 4 shows comparisons among these three algorithms. Note that $p = N + 1$ is the number of coefficients.

We present in Figure 3 the curves of $\mu(k)$ for values of desired signal to observation noise ratio, $SNR = 10log\frac{\sigma_d^2}{\sigma_n^2}$, from 0 to 40 dB. Note that for $\sigma_n^2 = 0$ (noiseless case), the SNR goes to infinity and the step-size would remain constant, equal to one.

We will next demonstrate the superior performance obtained with the proposed adaptive step-size scheme which in real time can be computed *a priori* and stored in memory or computed. For this last option, an approximation of the curve is of great interest. We will use here two classes of sequences also proposed in [4]. They were chosen due to their simplicity and, as will be seen later, lead to good results. The first class is the optimal sequence for the NLMS algorithm. It is given by

$$\mu(k) = \mu(k-1)\frac{1 - \frac{\mu(k-1)}{N+1}}{1 - \frac{\mu^2(k-1)}{N+1}} \tag{28}$$

For the NLMS algorithm, the correct initialization for this sequence is given by $\mu(0) = 1 - \frac{\sigma_n^2}{\sigma_d^2}$. However, in our case we can choose an initial value for the step-size such that the two sequences are close, as will be seen.

The second class of sequences (referred to hereafter as the $1/k$ approximation) is quite simple and was also used in [4]. This sequence is given by

$$\mu(k) = \begin{cases} 1 & \text{if } 0 \leq k \leq c(N+1) \\ max\{\mu_{min}, \frac{1}{1-c+\frac{k}{N+1}}\} & \text{if } k > c(N+1) \end{cases} \tag{29}$$

The parameter c will be related to the SNR of the optimal sequence. A minimum step-size was introduced here (it can be used in all sequences as well) in order to provide tracking capability to the algorithm.

For the first simulation, we used a white noise input signal in a system identification setup with $N = 10$, $\sigma_n^2 = 10^{-2}$ and $SNR = 20dB$. Figure 4 shows the optimal step-size sequence obtained with the algorithm described in Table 2 and other curves from the two classes of approximations used.

From Figure 4, we can guess which curve to use. If we use the least norm of the difference between the optimal and the approximation sequences as a criterion to decide which curve to implement, the chosen parameters for this example will be $\mu(0) = 0.93$ and $c = 3$.

With these parameters we have run a simulation with a fixed step-size, an optimal step-size and the two approximations. The learning curves (average of 1000 runs) are depicted in Figure 5 where we can see that the same fast convergence and the same small steady-state MSE are shared by the three time-varying step-size sequences used. The fixed step-size was set to one and, as expected, has the highest misadjustment.

A second experiment was carried out in order to evaluate the performance of this optimal sequence in case where the input signal is correlated. The same setup was used but with an input signal having a condition number around 180. Figure 6 shows us that, also in cases of correlated input signal, the proposed step-size sequence has a good performance.

A final remark is the possibility to use an estimator for $\xi(k)$ instead of calculating

$\Delta\xi(k)$ using (26) as described in the algorithm of Table 2. We have also made an experiment using the following estimator:

$$\xi(k+1) = \lambda\xi(k) + (1-\lambda)e^2(k) \tag{30}$$

This experiment has shown us that a reasonable value for λ is around 0.96. The advantage of this alternative approach is the possibility of fast tracking of sudden and strong changes in the environment. In this case, the instantaneous error becomes high and the estimated $\xi(k+1)$ is increased such that the value of μ approaches the unity again and a fast re-adaptation starts.

When using this approach, it is worth remembering that, since equation (27) is of the type $1-\sqrt{1-x}$, the step-size $\mu(k)$ can be written as $\frac{x}{1+\sqrt{1-x}}$ which is a numerically less sensitive expression. Equation (31) shows (27) rewritten with this numerically better expression.

$$\mu(k) = \frac{\frac{\xi(k)+\xi(k-1)-2\sigma_n^2}{2\xi(k-1)}}{1+\sqrt{1-\frac{\xi(k)+\xi(k-1)-2\sigma_n^2}{2\xi(k-1)}}} \tag{31}$$

6 Conclusions

This paper introduced the BNDR-LMS algorithm which has faster convergence than a number of other LMS-like algorithms when the input signal is highly correlated. A geometric interpretation of the algorithm was also provided showing that the coefficients are updated in two normalized steps following orthogonal directions. The relationship between the BNDR-LMS algorithm and the orthogonal-projections algorithm was clarified. Simulations carried out in a system identification application showed that the BNDR-LMS algorithm compared favorably with other LMS-like algorithms in terms of speed of convergence. Moreover, the more correlated is the input signal, the better the new algorithm performs. This improvement is clearly verified in cases of high signal to noise ratio.

This work also addressed the optimization of the step-size of the BNDR-LMS algorithm when the input signal is uncorrelated. An optimal sequence was proposed and a simple algorithm to find this sequence was introduced. Alternative approximation sequences were also presented and their initialization parameters compared. Simulations carried out in a system identification problem showed the good performance of the optimal step-size sequence as well as the possibility of using alternatives sequences obtained with less effort and with similar efficiency. It was possible to observe that the same step-size sequence, optimal for the white noise input, can also be used in applications where a highly correlated input signal is present.

References

[1] P.S.R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation.* Kluwer Academic Press, Boston, MA, USA, 1997.

[2] S. Haykin, *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1996.

[3] S. Roy and J.J. Shynk, “Analysis of the data-reusing LMS algorithm,” *32nd Midwest Symposium Circuits and Systems*, pp. 1127-1130, Urbana-Champaign,

IL, USA, 1990.

[4] D. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2811-2825, Sept. 1993.

[5] B.A. Schnaufer, *Practical Techniques for Rapid and Reliable Real-Time Adaptive Filtering,* Ph.D. Thesis, Urbana-Champaign, IL, USA, 1995.

[6] G.C. Goodwin and K.S. Sin, *Adaptive Filtering Prediction and Control.* Prentice-Hall, Englewood Cliffs, NJ, USA, 1984.

[7] J. A. Apolinário Jr., M. L. R. de Campos and P. S. R. Diniz, "The Binormalized Data-Reusing LMS Algorithm," Proceedings of the *XV Simpósio Brasileiro de Telecomunicações*, Recife, Brazil, 1997.

[8] J. A. Apolinário Jr., M. L. R. de Campos and P. S. R. Diniz, "Convergence analysis of the Binormalized Data-Reusing LMS Algorithm," Proceedings of the *European Conference on Circuit Theory and Design*, Budapest, Hungary, 1997.

[9] M. L. R. de Campos, J. A. Apolinário Jr. and P. S. R. Diniz, "Mean-squared error analysis of the Binormalized Data-Reusing LMS algorithm using a discrete-angular-distribution model for the input signal," accepted to the International Conference on Acoustics, Speech, and Signal Processing, Seattle, USA, 1998.

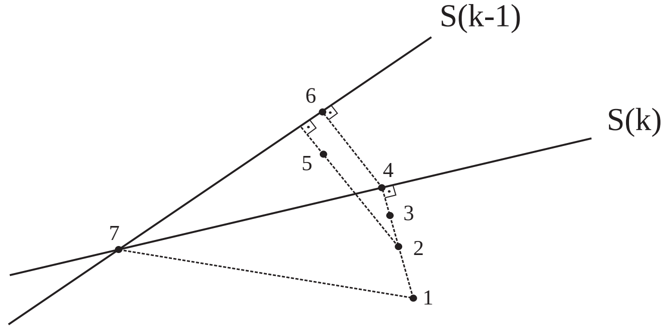

Figure 1: Updating the coefficient vector:
1. $\boldsymbol{w}(k)$;
2. $\boldsymbol{w}_{LMS}(k+1)$ (first step of DR-LMS and UNDR-LMS);
3. $\boldsymbol{w}_{DR-LMS}(k+1)$;
4. $\boldsymbol{w}_{NLMS}(k+1)$ (first step of NNDR-LMS);
5. $\boldsymbol{w}_{UNDR-LMS}(k+1)$;
6. $\boldsymbol{w}_{NNDR-LMS}(k+1)$;
7. $\boldsymbol{w}_{BNDR-LMS}(k+1)$.

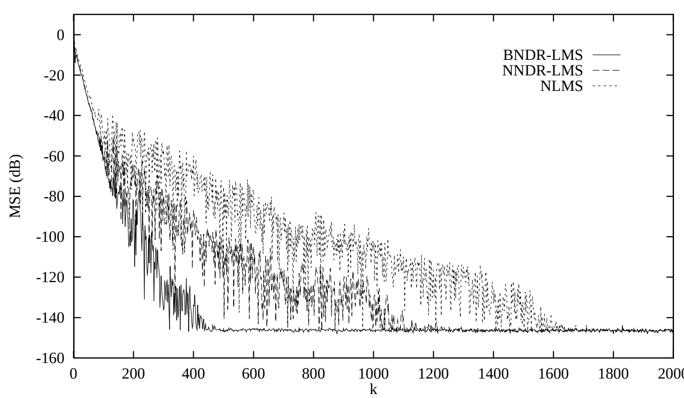

Figure 2: Learning curves of the following algorithms: NLMS, NNDR-LMS and BNDR-LMS.

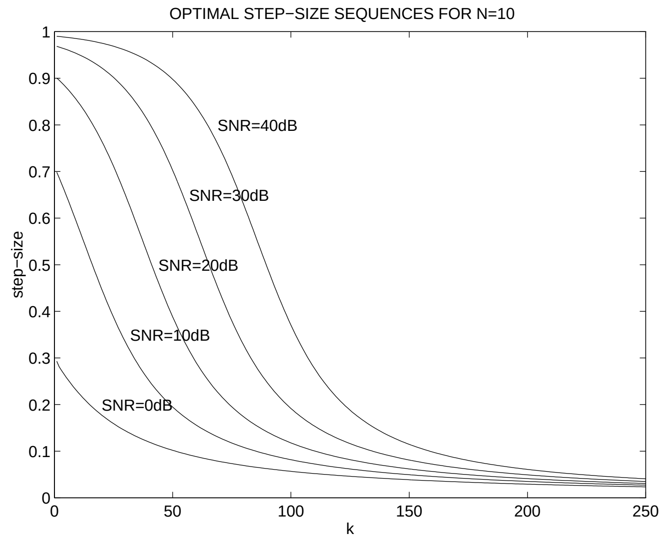

Figure 3: Optimal $\mu(k)$ sequences for the BNDR-LMS algorithm.

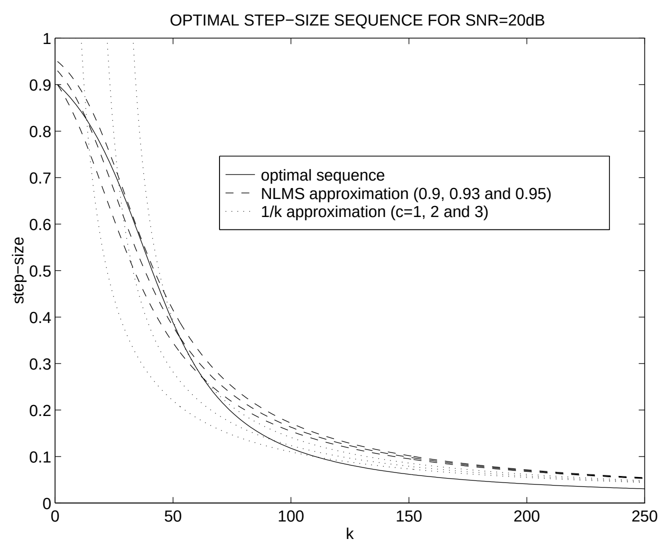

Figure 4: Optimal step-size sequence and two classes of approximation sequences.

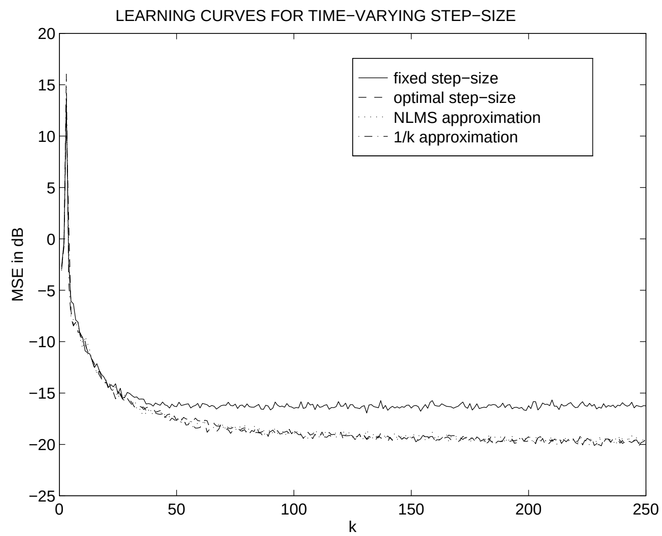

Figure 5: Learning curves for the fixed step-size, the optimal step-size and its two approximations.

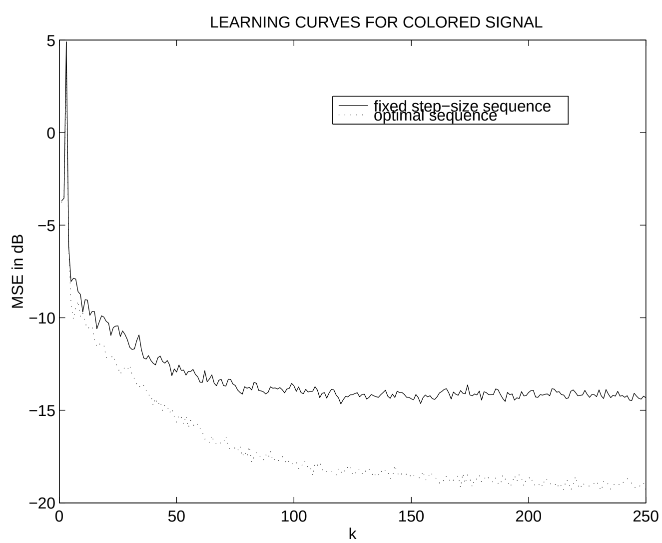

Figure 6: Comparing the learning curves for the case of colored input signal.

Table 1: The Binormalized Data-Reusing LMS Algorithm [8].

BNDR-LMS
$\epsilon =$ small positive value for each k $\{\ \boldsymbol{x}_1 = \boldsymbol{x}(k)$ $\quad \boldsymbol{x}_2 = \boldsymbol{x}(k-1)$ $\quad d_1 = d(k)$ $\quad d_2 = d(k-1)$ $\quad a = \boldsymbol{x}_1^T \boldsymbol{x}_2$ $\quad b = \boldsymbol{x}_1^T \boldsymbol{x}_1$ $\quad c = \boldsymbol{x}_2^T \boldsymbol{x}_2$ $\quad d = \boldsymbol{x}_1^T \boldsymbol{w}(k)$ $\quad$ if $a^2 == bc$ $\quad \{\ \boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu(d_1 - d)\boldsymbol{x}_1/(b+\epsilon)$ $\quad \}$ $\quad$ else $\quad \{\ e = \boldsymbol{x}_2^T \boldsymbol{w}(k)$ $\qquad den = bc - a^2$ $\qquad A = (d_1 c + ea - dc - d_2 a)/den$ $\qquad B = (d_2 b + da - eb - d_1 a)/den$ $\qquad \boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu(A\boldsymbol{x}_1 + B\boldsymbol{x}_2)$ $\quad \}$ $\}$

Table 2: Algorithm for computing the optimal step-size sequence.

$\boldsymbol{\mu(k)}$ **of the BNDR-LMS algorithm**
$\Delta\xi(0) = \Delta\xi(-1) = \sigma_d^2$ σ_n^2 = noise variance N = adaptive filter order $\mu(0) = 1$ for each k $\{\ \mu(k) = 1 - \sqrt{1 - \frac{\Delta\xi(k)+\Delta\xi(k-1)}{2(\Delta\xi(k-1)+\sigma_n^2)}}$ $aa = \left[1 + \frac{\mu(k)(\mu(k)-2)}{N+1}\right]$ $bb = \frac{N\mu(k)(1-\mu(k))^2(\mu(k)-2)}{(N+1)^2}$ $cc = \frac{(1+N(\mu(k)-2)^2)\mu(k)^2}{(N+1)^2}\sigma_n^2$ $\Delta\xi(k+1) = aa\Delta\xi(k) + bb\Delta\xi(k-1) + cc$ $\}$

Table 3: Excess Mean-Square Error.

Algorithm Type	$(\text{MSE - MSE}_{min})_{dB}$	
	Stationary	Nonstationary
NLMS	-59.09	-39.15
NNDR-LMS	-59.40	-39.42
BNDR-LMS	-58.60	-39.45

Table 4: Comparison of computational complexity.

ALG.	ADD	MULT.	DIV.
NLMS	3p-1	3p	1
NNDR-LMS	6p-2	6p	2
BNDR-LMS	7p+3	7p+2	2