STEP-SIZE OPTIMIZATION OF THE BNDR-LMS ALGORITHM

J.A. Apolinário Jr.[1,2,3*], *P.S.R. Diniz*[2] , *T.I. Laakso*[1] and *M. L. R. de Campos*[3]
[1] Helsinki University of Technology, Helsinki, FINLAND
e-mail: Timo.Laakso@hut.fi
[2] COPPE/Universidade Federal do Rio de Janeiro, Rio de Janeiro, BRAZIL
e-mail: diniz@coe.ufrj.br
[3] Instituto Militar de Engenharia, Rio de Janeiro, BRAZIL
e-mails: apolin@aquarius.ime.eb.br and campos@aquarius.ime.eb.br

ABSTRACT

The binormalized data-reusing least mean squares (BNDR-LMS) algorithm has been recently proposed and has been shown to have faster convergence than other LMS-like algorithms in cases where the input signal is strongly correlated. This 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 simplicity of the least mean squares (LMS) algorithm has motivated alternative schemes which try to compensate for its main drawback, the dependence on the eigenvalue spread of the input autocorrelation matrix at the expense of a minimum extra computational load [1, 2]. The BNDR-LMS algorithm [3, 4] offers faster convergence than a number of other normalized LMS algorithms for a highly correlated input signals at the cost of a small additional complexity. The MSE after convergence for this algorithm 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 [4], it was shown that the BNDR-LMS algorithm convergences 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 [5] an analysis for the MSE behavior of the BNDR-LMS algorithm was available. In this paper, the expression for the MSE developed in [5] is used to propose an optimal step-size sequence which allows a fast convergence and a minimum misadjustment.

*This work was partially supported by CAPES, Brazil.

The paper is organized as follows. In Section 2 we present the BNDR-LMS algorithm as well as its convergence behavior. Section 3 develops the optimal step-size sequence of $\mu(k)$. In Section 4 several approximations for this optimal sequence are proposed and simulation results are presented. Finally, Section 5 draws some conclusions.

2 THE BNDR-LMS ALGORITHM

The BNDR-LMS algorithm employs normalization on two orthogonal directions obtained from consecutive data pairs within each iteration. These data pairs are the input data vectors and the reference or desired signal at instants k and $k-1$, denoted by $(\boldsymbol{x}(k), d(k))$ and $(\boldsymbol{x}(k-1), d(k-1))$. This algorithm is described by the equations of Table 1. According to the notation used in Table 1, the coefficient vector is given by $\boldsymbol{w}(k)$.

In order to use the results from [5], let us assume that an unknown FIR filter is to be identified by an adaptive filter of the same order (system identification problem), employing the BNDR-LMS algorithm. The 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.

The final expression for the convergence behavior of the BNDR-LMS algorithm obtained in [5] is given in terms of the excess in the MSE defined as the difference between the MSE and the minimum MSE after convergence or $\Delta\xi(k) = \xi(k) - \xi_{min} = E[e^2(k)] - \sigma_n^2$, since we assume here that the minimum mean square error is caused only by additional noise.

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

The term ν_x is known as the *kurtosis* of the input signal and it is, for the sake of simplicity, assumed here

Table 1: The Binormalized Data-Reusing LMS Algorithm [4].

BNDR-LMS
ϵ = small 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
$\quad \{\ e = \boldsymbol{x}_2^T \boldsymbol{w}(k)$
$\quad\quad den = bc - a^2$
$\quad\quad A = (d_1 c + ea - dc - d_2 a)/den$
$\quad\quad B = (d_2 b + da - eb - d_1 a)/den$
$\quad\quad \boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu(A\boldsymbol{x}_1 + B\boldsymbol{x}_2)$
$\quad \}$
$\}$

to be equal to unity. This expression for the excess in the MSE [1] was obtained with the help of a simple model [6] for the input signal vector given by

$$\boldsymbol{x}(k) = s_k r_k \boldsymbol{V}_k \tag{2}$$

where s_k is ± 1 with probability $1/2$, r_k has the same distribution as $\| \boldsymbol{x}(k) \|$ or $E[r_k^2] = (N+1)\sigma_x^2$, and $\boldsymbol{V}_k$ is one of the $N+1$ orthonormal eigenvectors of the autocorrelation matrix $\boldsymbol{R}$ of the input signal. It was assumed in the analysis that each vector $\boldsymbol{V}_i$ occurs with an equal probability of $1/(N+1)$ which means white noise type input.

3 THE OPTIMAL STEP-SIZE SEQUENCE

In this section the optimal step-size sequence for the given problem is derived. From the expression of $\Delta\xi(k+1)$, we will follow an approach similar to that used in [6] and rewrite (1) assuming that up to time k we have the optimal sequence $\mu^*(0)$ to $\mu^*(k-1)$ already available and also the optimal quantities $\Delta\xi^*(k)$ and $\Delta\xi^*(k-1)$.

$$\begin{aligned} \Delta\xi(k+1) &= \left[1 + \frac{\mu(k)(\mu(k)-2)}{N+1}\right] \Delta\xi^*(k) \\ &+ \frac{N\mu(k)(1-\mu(k))^2(\mu(k)-2)}{(N+1)^2} \Delta\xi^*(k-1) \\ &+ \frac{(1+N(\mu(k)-2)^2)\mu(k)^2}{(N+1)^2} \sigma_n^2 \end{aligned} \tag{3}$$

If we now compute the derivative of $\Delta\xi(k+1)$ with respect to $\mu(k)$ and make it equal to zero, we obtain after some algebraic manipulation

$$\begin{aligned} \mu^*(k) &= 1 - \sqrt{1 - \frac{\Delta\xi^*(k) + \Delta\xi^*(k-1)}{2(\Delta\xi^*(k-1) + \sigma_n^2)}} \\ &= 1 - \sqrt{1 - \frac{\xi^*(k) + \xi^*(k-1) - 2\sigma_n^2}{2\xi^*(k-1)}} \end{aligned} \tag{4}$$

It is worth mentioning that (4) is in accordance with the situation when the convergence is reached; in that case we have $\xi^*(k) = \xi^*(k-1) = \sigma_n^2$ and therefore we have $\mu^*(k) = 0$ as expected. Moreover, if we have $\sigma_n^2 = 0$ the value of $\mu^*(k)$ will be close to one (admitting that $\Delta\xi^*(k) \approx \Delta\xi^*(k-1)$) even after convergence, which means that we should have maximum speed of convergence with minimum misadjustment if the noise is zero.

For the normalized LMS (NLMS) algorithm, a recursive formula for $\mu^*(k)$ in terms of $\mu^*(k-1)$ and the order N was obtained in [6]. Unfortunately, a similar expression could not be obtained for the BNDR-LMS algorithm. Instead, a simple algorithm was used to produce the optimal step-size sequence. This algorithm is presented in Table 2[1] and has one important initialization parameter with a strong influence on the behavior of $\mu^*(k)$. This parameter is the ratio $\frac{\sigma_d^2}{\sigma_n^2}$ where the numerator is the variance of the reference signal.

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
$\{\ \mu(k) = 1 - \sqrt{1 - \frac{\Delta\xi(k) + \Delta\xi(k-1)}{2(\Delta\xi(k-1)+\sigma_n^2)}}$
$\quad aa = \left[1 + \frac{\mu(k)(\mu(k)-2)}{N+1}\right]$
$\quad bb = \frac{N\mu(k)(1-\mu(k))^2(\mu(k)-2)}{(N+1)^2}$
$\quad cc = \frac{(1+N(\mu(k)-2)^2)\mu(k)^2}{(N+1)^2}\sigma_n^2$
$\quad \Delta\xi(k+1) = aa\Delta\xi(k) + bb\Delta\xi(k-1) + cc$
$\}$

We next present in Fig. 1 the curves of $\mu(k)$ for different values of what should be called in this case (desired) signal to noise ratio or $SNR = 10log\frac{\sigma_d^2}{\sigma_n^2}$ 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 fixed at unity.

[1] Note that the asterisk (*) was dropped out from the optimal values for simplicity only.

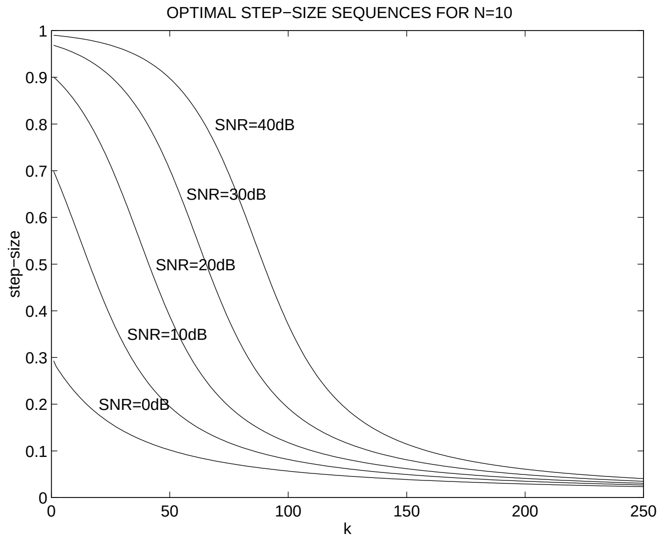

Figure 1: Optimal $\mu(k)$ sequences for the BNDR-LMS algorithm.

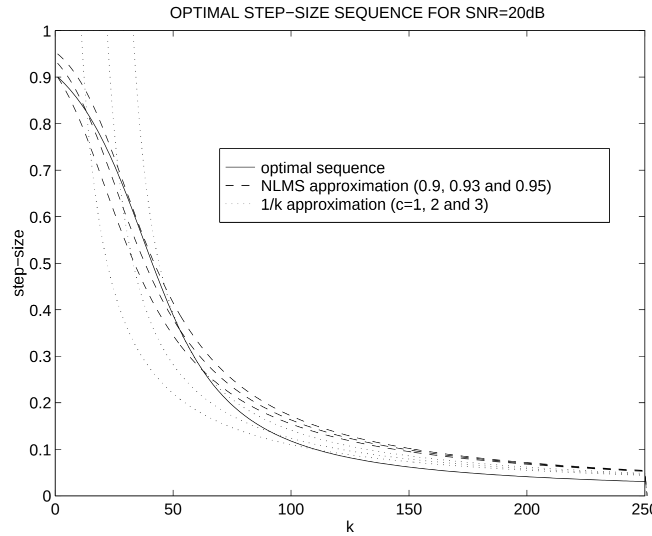

Figure 2: Optimal step-size sequence and two classes of approximation sequences.

4 SIMULATION RESULTS

In this section the results of a few experiments will demonstrate the superior performance obtained with the proposed adaptive step-size scheme. In a practical implementation the optimal sequence can be computed *a priori* and stored in memory or computed on the fly. For this last option, since a recursive and compact formula is not available, an approximation of the curve is of great interest. We will use here two classes of sequences also proposed in [6]. 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

$$\mu(k) = \mu(k-1)\frac{1-\frac{\mu(k-1)}{N+1}}{1-\frac{\mu^2(k-1)}{N+1}} \quad (5)$$

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 [6]. This sequence is given by

$$\mu(k) = \begin{cases} 1 & \text{if } 0 \leq k \leq c(N+1) \\ max\{\mu_{min}, \frac{1}{1-c+\frac{k}{N+1}}\} & \text{if } k > c(N+1) \end{cases} \quad (6)$$

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 a 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 2 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 Fig. 2, 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 Fig. 3 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 (ratio between the largest and the smallest eigenvalue of the input signal autocorrelation matrix) around 180. Fig. 4 shows us that, even for a 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 (3) as described in the algorithm of Table 2. We have also made an experiment using the following estimator:

$$\xi(k+1) = \lambda\xi(k) + (1-\lambda)e^2(k) \quad (7)$$

This experiment has shown us that a reasonable value for λ is around 0.96. The advantage of this alternative

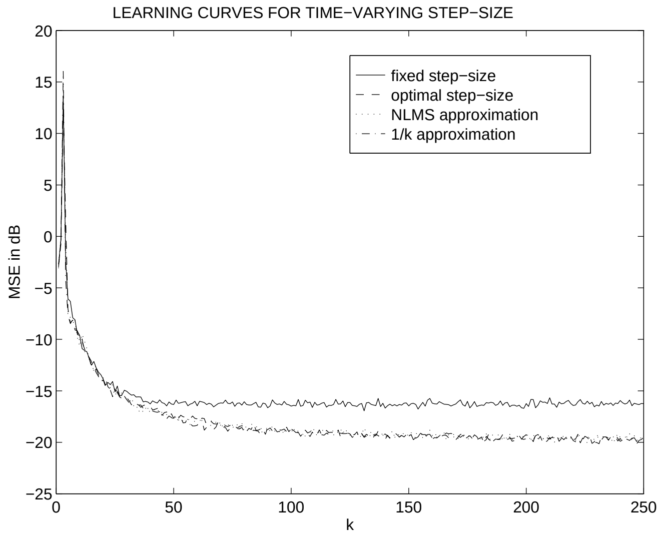

Figure 3: Learning curves for the fixed step-size, the optimal step-size and its two approximations.

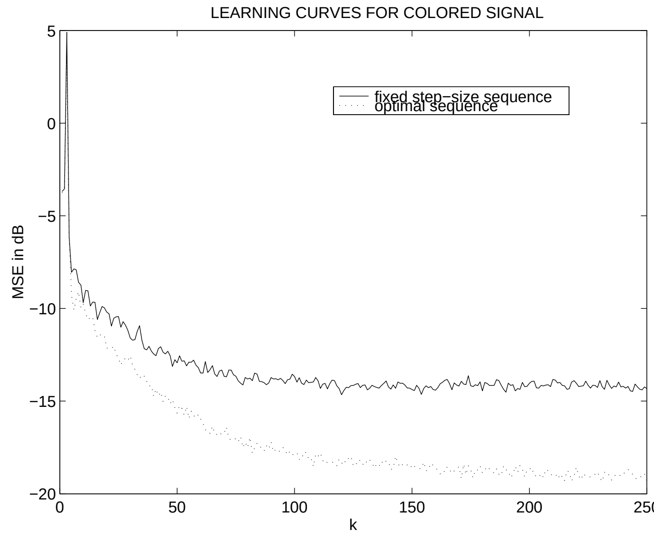

Figure 4: Comparing the learning curves for the case of colored input signal.

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 (4) 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 (8) shows this expression.

$$\mu(k) = \frac{\frac{\xi(k)+\xi(k-1)-2\sigma_n^2}{2\xi(k-1)}}{1+\sqrt{1-\frac{\xi(k)+\xi(k-1)-2\sigma_n^2}{2\xi(k-1)}}} \tag{8}$$

5 CONCLUSIONS

This paper 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. In another simulation 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] 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.

[4] 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.

[5] 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," Proceedings of the *International Conference on Acoustics, Speech, and Signal Processing*, Seattle, USA, 1998.

[6] 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.