Colored Input-Signal Analysis of Normalized Data-Reusing LMS Algorithms

M. L. R. de Campos,[‡] *J. A. Apolinário Jr.,*[†‡] *P. S. R. Diniz,*[†]
campos@ime.eb.br, apolin@ime.eb.br, diniz@lps.ufrj.br

[‡]Departamento de Engenharia Elétrica
Instituto Militar de Engenharia
Praça General Tibúrcio, 80
CEP 22290-270, Rio de Janeiro, RJ - Brazil

[†]Programa de Engenharia Elétrica
COPPE — Univ. Federal do Rio de Janeiro
P. O. Box 68564 — FAX: +55 21 290-6626
CEP 21945-970, Rio de Janeiro, RJ - Brazil

Abstract—In this paper, a simplified model is used to qualitatively and quantitatively analyze two normalized data-reusing LMS algorithms when the input sequence is colored. The proposed approach relates the eigenvalue distribution of the input-signal autocorrelation matrix with probabilities of occurrence of subsequent parallel or orthogonal input-signal vectors. Closed-form formulas that describe steady-state mean-squared error (MSE) are provided and confronted with simulation results.

I. Introduction

Despite the great popularity enjoyed by the least-mean-squares (LMS) algorithm for adaptive filtering, its performance with respect to convergence speed may be far from acceptable for many applications. This is the case when the input data sequence is correlated. However, due to increased throughput capacity of recently developed digital-signal-processing machines, more elaborate algorithms which present faster convergence even for highly correlated signals have been proposed. Data-reusing algorithms [1] present an interesting approach in a sense that computational complexity can be adjusted according to availability of time for processing. Algorithm complexity, in this case, is between that of Newton-type algorithms and that of the conventional LMS algorithm.

In their most simple and common implementation, data-reusing LMS (DR-LMS) algorithms utilize the current input-signal vector as many times as it is possible within each iteration in order to obtain better estimates of the coefficients. Better convergence rates may be achieved if current as well as previous input-signal vectors are employed [1]. Algorithms which belong to this class will be referred to as normalized new data-reusing LMS (NNDR-LMS) and unnormalized new data-reusing LMS (UNDR-LMS) algorithms, depending on whether the step sizes taken are normalized or not. This nomenclature was chosen to maintain consistency with previous publications related to these algorithms [1].

The binormalized data-reusing LMS (BNDR-LMS) algorithm [2, 3] bears close relationship with the NNDR-LMS algorithm with the difference that the previous input-signal vector is transformed prior to reutilization. At each iteration the previous input-signal vector is projected onto a subspace orthogonal to the subspace spanned by the current input-signal vector. Such transformation allows simultaneous normalizations with respect to current and previous data. Simulations for colored input sequences have shown that the BNDR-LMS algorithm may converge faster than the NNDR-LMS and the NLMS algorithms with a minor increase in computational complexity.

In [4], a simple model for the input signal vector has been introduced in order to simplify analysis of convergence behavior of the LMS and normalized LMS (NLMS) algorithms. The model assumes independent discrete angular distribution and continuous radial distribution for the input-signal vector. The angular distribution is such that input-signal vectors at consecutive time instants can only be parallel or orthogonal, and the radial distribution assures that accuracy is preserved at least for the first two statistical moments.

II. Data-Reusing LMS Algorithms

A. NNDR-LMS Algorithm

For the NNDR-LMS algorithm with L data reuses, the coefficient vector is updated as

$$\boldsymbol{w}_{i+1}(k) = \boldsymbol{w}_i(k) + \frac{e_i(k)}{\|\boldsymbol{x}(k-i)\|^2 + \epsilon}\boldsymbol{x}(k-i)$$

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

$$e_i(k) = d(k-i) - \boldsymbol{x}^T(k-i)\boldsymbol{w}_i(k)$$
$$\boldsymbol{w}_0(k) = \boldsymbol{w}_{NNDR}(k)$$
$$\boldsymbol{w}_{NNDR}(k+1) = \boldsymbol{w}_{L+1}(k)$$

0-7803-5008-1/98/$10.00©1998 IEEE.

$\boldsymbol{x}(k)$ is the input-signal vector, and $d(k)$ is the desired signal at time instant k. The set of equations displayed above can be rewritten such that

$$\boldsymbol{w}_{NNDR}(k+1) = \boldsymbol{w}_{NNDR}(k) + \mu\boldsymbol{\Delta w}(k)$$

where a convergence factor $\mu \leq 1$ is introduced to control misadjustment while maintaining the original update direction.

B. BNDR-LMS Algorithm

For the BNDR-LMS algorithm, assuming linearly independent input-signal vectors $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$, the coefficient vector is updated as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu\left[\frac{\lambda_1}{2}\boldsymbol{x}(k) + \frac{\lambda_2}{2}\boldsymbol{x}(k-1)\right]$$

where

$$\frac{\lambda_1}{2} = \frac{[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\|\boldsymbol{x}(k-1)\|^2}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} - \frac{[d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)]\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}$$

and

$$\frac{\lambda_2}{2} = \frac{[d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)]\|\boldsymbol{x}(k)\|^2}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} - \frac{[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}$$

Analyses in the mean and in the mean-squared of the coefficient error of the BNDR-LMS algorithm have been carried out based on the assumption of white Gaussian input sequence and showed that stability is guaranteed for $\mu \leq 2$ [2, 3].

III. Input-Signal Model

In order to avoid the enormous complexity involved in the analysis of data-reusing algorithms, a simplified model for the input-signal vector that can be consistent with the first- and second-order statistics of a general input signal vector, but has a reduced and countable number of possible directions of excitation, is as follows:

$$\boldsymbol{x}(k) = s_k r_k \boldsymbol{V}_k \tag{1}$$

where:

- s_k is ± 1 with probability of occurrence $1/2$;
- r_k^2 has the same probability distribution function of $\|\boldsymbol{x}(k)\|^2$, or, for the case of interest, is a sample of an independent process with χ-square distribution of $(N+1)$ degrees of freedom, $E[r_k^2] = (N+1)\sigma_x^2$;
- $\boldsymbol{V}_k$ is equal to one of the $N+1$ orthonormal eigenvectors of $\boldsymbol{R}$, denoted $\mathcal{V}_i$, $i = 1, \ldots, N+1$.

IV. Mean-Squared-Error Analysis

Measurement noise or inaccurate order estimation cause, in general, a fluctuation of the coefficient estimates around their mean values. This excess of MSE is the usual choice for comparing steady-state behavior of algorithms and is defined as

$$\xi_{exc} = \lim_{k\to\infty} \xi(k) - \xi_{min}$$

where $\xi(k) = E[e^2(k)]$ and ξ_{min} is the minimum mean-squared error [5]. The difference $\Delta\xi(k) = \xi(k) - \xi_{min}$ is known as excess in the MSE [5] and can be expressed as

$$\Delta\xi(k) = \text{tr}\{\boldsymbol{R}\,\text{cov}[\Delta\boldsymbol{w}(k)]\}$$

For the given input-signal model, we may express $\Delta\xi(k+1)$ as

$$\begin{aligned}\Delta\xi(k+1) = {} & \Delta\xi(k+1)\Big|_{\boldsymbol{x}(k)\,\|\,\boldsymbol{x}(k-1)} \\ & \times P[\boldsymbol{x}(k) \,\|\, \boldsymbol{x}(k-1)] \\ & + \Delta\xi(k+1)\Big|_{\boldsymbol{x}(k)\,\perp\,\boldsymbol{x}(k-1)} \\ & \times P[\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)]\end{aligned} \tag{2}$$

Conditions $\boldsymbol{x}(k) \,\|\, \boldsymbol{x}(k-1)$ and $\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)$ in the adopted model are equivalent to $\boldsymbol{V}_k = \boldsymbol{V}_{k-1}$ and $\boldsymbol{V}_k \neq \boldsymbol{V}_{k-1}$, respectively, for $\boldsymbol{V}_k$ and $\boldsymbol{V}_{k-1}$ can only be parallel or orthogonal to each other.

A. BNDR-LMS Algorithm

As it was presented in [3], the BNDR-LMS algorithm behaves exactly like the NLMS algorithm when the input-signal vectors at instants k and $k-1$ are parallel. In this case, the excess of MSE is given by [4]

$$\Delta\xi(k+1)_{\|} = \left[1 + \frac{\mu(\mu-2)}{N+1}\right]\Delta\xi(k) + \frac{\mu^2\sigma_n^2}{(N+2-\nu_x)} \tag{3}$$

where $\nu_x = E[x^4(k)/\sigma_x^4]$ is the *kurtosis* of the input signal [4].

For the case where $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$ are orthogonal, we have, for $\boldsymbol{R} = \sigma_x^2\boldsymbol{I}$, i.e., white-noise input signals [3],

$$\Delta\xi(k+1)_{\perp} = \left[1 + \frac{\mu(\mu-2)}{N+1}\right]\Delta\xi(k) + \frac{\mu(1-\mu)^2(\mu-2)}{N+1}\Delta\xi(k-1) + \frac{\mu^2(\mu-2)^2}{N+2-\nu_x}\sigma_n^2 \tag{4}$$

B. NNDR-LMS Algorithm

A simple geometrical description of the NNDR-LMS and BNDR-LMS algorithms, as, e.g., the one provided in [3], shows that the two algorithms take identical steps from iteration $(k-1)$ to iteration k when the input-signal vectors at these time instants are orthogonal, i.e., $\boldsymbol{x}(k-1) \perp \boldsymbol{x}(k)$. Therefore, we can safely assign (4) to both algorithms.

When $\boldsymbol{x}(k-1) \parallel \boldsymbol{x}(k)$, it can be easily verified that the coefficient vector is updated as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu \frac{d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)}{\|\boldsymbol{x}(k-1)\|^2} \boldsymbol{x}(k-1)$$

Consequently, the excess in the MSE in this case reduces to

$$\Delta\xi(k+1)_{\parallel} = \Delta\xi(k) - (2\mu - \mu^2)\sigma_x^2 \times E\left\{ \frac{[\boldsymbol{x}^T(k-1)\boldsymbol{\Delta w}(k)]^2}{\|\boldsymbol{x}(k-1)\|^2} \right\} + \mu^2\sigma_n^2\sigma_x^2 E\left[\frac{1}{\|\boldsymbol{x}(k-1)\|^2} \right]$$

Substituting (1) in the equation above an expression identical to (3) results. Therefore, assuming the simplified model for the input-signal vector, the NNDR-LMS and BNDR-LMS algorithms present the same excess in the MSE.

C. Excess of MSE

A final expression for the excess in the MSE may now be obtained from (3) and (4) combined and weighted accordingly, as suggested in (2). The solution for $k \to \infty$ provides the magnitude of the excess of MSE. If the angular distribution of the input-signal vector is considered uniform, the probabilities of $\boldsymbol{V}_k = \boldsymbol{V}_{k-1}$ and $\boldsymbol{V}_k \neq \boldsymbol{V}_{k-1}$ are equal to $\frac{1}{N+1}$ and $\frac{N}{N+1}$, respectively. The reasoning used to extend the analysis to colored input signals is based on the fact that only the angular distribution of $\boldsymbol{x}(k)$ need be changed in order to incorporate different probabilities for the directions given by the $(N+1)$ eigenvectors of $\boldsymbol{R}$. In other words, (2)–(4) are maintained and only probabilities $P[\boldsymbol{x}(k) \parallel \boldsymbol{x}(k-1)]$ and $P[\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)]$ need be recalculated. Each eigenvector of $\boldsymbol{R}$, denoted as $\mathcal{V}_i$, $i = 1, \ldots, N+1$ will now have the following probability of occurrence [4]

$$P(\boldsymbol{V}_k = \mathcal{V}_i) = \frac{\lambda_i}{tr(R)} \tag{5}$$

where λ_i is the eigenvalue associated to the eigenvector $\mathcal{V}_i$. From the input signal we can obtain all necessary eigenvalues and eigenvectors such that we can compute

$$\begin{aligned} P[\boldsymbol{x}(k) \parallel \boldsymbol{x}(k-1)] &= P[\boldsymbol{V}_k \parallel \boldsymbol{V}_{k-1}] \\ &= P\left[\boldsymbol{V}_k \parallel \boldsymbol{V}_{k-1} \,|_{\boldsymbol{V}_{k-1} = \boldsymbol{V}_1}\right] \times P[\boldsymbol{V}_{k-1} = \boldsymbol{V}_1] + \cdots \\ &\quad + P\left[\boldsymbol{V}_k \parallel \boldsymbol{V}_{k-1} \,|_{\boldsymbol{V}_{k-1} = \boldsymbol{V}_{N+1}}\right] \times P[\boldsymbol{V}_{k-1} = \boldsymbol{V}_{N+1}] \\ &= \sum_{i=1}^{N+1} \left(\frac{\lambda_i}{tr(R)} \right)^2 \end{aligned} \tag{6}$$

and

$$P[\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)] = 1 - P[\boldsymbol{x}(k) \parallel \boldsymbol{x}(k-1)] \tag{7}$$

Equations (6) and (7) are in accordance with the white-input situation, for this case all eigenvalues are equal to σ_x^2 such that $P(\boldsymbol{V}_k = \mathcal{V}_i) = \frac{1}{N+1}$ as already mentioned. When the input signal is correlated through a first-order allpole filter and modeled with (1) and (5), the excess of MSE is given by (2)–(4) with probabilities given by (6) and (7). Although (3)–(4) have been obtained based on a white Gaussian model for the input signal, simulations have shown that our reasoning is valid when the input signal is generated according to (1) with probabilities given by (5) and λ_i obtained from the input signal. Moreover, for $\mu = 1$ and a modeled input signal where only parallel or perpendicular vectors may occur, the BNDR-LMS algorithm degrades to the NLMS algorithm and the steady-state MSE becomes independent of the radial distribution of $\boldsymbol{x}(k)$ [4]. This is perfectly described by (3) and (4), supporting the validity of our reasoning.

V. Simulation Results

An experiment was designed to test the accuracy of the expressions derived in the analysis when the input signal is correlated. A 10th-order FIR adaptive filter was used to identify an unknown 10th-order FIR digital filter with coefficients randomly chosen from a zero-mean unity-variance white process with uniform distribution. Simulations with 10,000 iterations were averaged in an ensemble of 50 runs. The coefficients of the unknown system were constant throughout the ensemble, but for each run the input sequence was obtained from an independent and identically distributed zero-mean white Gaussian process filtered by an allpole lowpass digital filter with pole $\gamma = 0.9$. The eigenvalues were obtained from the autocorrelation matrix whose elements are given by $\sigma_\eta^2 \gamma^{|i-j|}(1-\gamma)/(1+\gamma)$ with σ_η^2 such that $\sigma_x^2 = 1$. Zero-mean white Gaussian noise with variance $\sigma_n^2 = 10^{-6}$ was used to corrupt the unknown-system output. Figure 1 shows the excess of MSE as a function of the step-size μ for the NLMS, NNDR-LMS, and BNDR-LMS algorithms in the setup described above, as well as the results expected from theory. For the NLMS algorithm, input-signal correlation does not affect its steady-state behavior [4]. As a consequence, (3) was used to obtain ξ_{exc}. For the NNDR-LMS and BNDR-LMS algorithm, $P_{\|}$ and $P_{\perp}$ were obtained from (6) and (7).

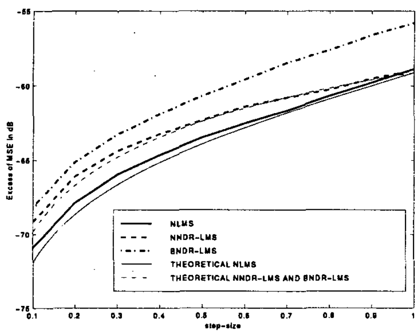

Figure 1: Excess of MSE for colored input signals.

A. Discussion

As pointed out in [4] the steady-state performance of the NLMS algorithm seems to be immune to input-signal correlation, which was verified in the experiment. Excess of MSE for the NNDR-LMS and BNDR-LMS algorithms differ in practice, specially for values of μ close to 1. Although not predicted in theory, this result can be explained in the light of a graphical interpretation of both algorithms. The amount of correlation in the input sequence is taken into account by the analysis method introduced here through a proper choice of the probability of occurrence of parallel subsequent input-signal vectors. This extreme situation implies "total correlation," and in this case, the NNDR-LMS and BNDR-LMS algorithms do behave identically. However, in a more practical scenario, this situation has zero probability of occurrence. It is just an extrapolation of the likely event of a small angle different from zero being formed between subsequent input-signal vectors. Let $\mathcal{S}(k)$ denote the hyperplane which contains all vectors $\boldsymbol{w}$ such that $\boldsymbol{x}^T(k)\boldsymbol{w} = d(k)$. For the NNDR-LMS algorithm, the coefficients are updated in a direction given by two normalized steps in the directions of $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$, which are orthogonal to $\mathcal{S}(k)$ and $\mathcal{S}(k-1)$, respectively [3]. The smaller the angle formed by the vectors, the better the approximation that considers this angle equal to zero, and the better is the agreement between theoretical and practical results. On the other hand, for the BNDR-LMS algorithm, the coefficients are updated in the direction of the intersection of hyperplanes $\mathcal{S}(k)$ and $\mathcal{S}(k-1)$. In this case, a small angle not equal to zero implies a huge step taken to update the coefficients. Therefore, the smaller the angle, the worse the approximation, because the algorithm behavior has a discontinuity as the angle formed by $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$ approaches zero.

VI. Conclusions

By means of the use of a simplified model for the input-signal vector the analysis of two normalized data-reusing adaptation algorithms, namely NNDR-LMS and BNDR-LMS algorithms, was carried out for colored input signals. The model is based on discrete angular and continuous radial distributions of orthogonal vectors, rendering vector manipulation tractable. The association with the scenario where the input sequence is correlated is done through a proper selection of probabilities of occurrence of parallel and orthogonal subsequent vectors. The analysis provided insight on the behavior of the algorithms in steady-state and closed-form formulas which describe the excess of MSE could be obtained. Simulation results showed accuracy of the method and validated the analysis.

References

[1] B. A. Schnaufer, *Practical Techniques for Rapid and Reliable Real-Time Adaptive Filtering.* PhD thesis, University of Illinois at Urbana-Champaign, Urbana-Champaign, USA, 1995.

[2] J. A. Apolinário Jr., M. L. R. de Campos, and P. S. R. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," in *Proc. European Conference on Circuit Theory and Design*, Budapest, Hungary, pp. 972–977, 1997.

[3] 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 for the *International Conference on Acoustics, Speech, and Signal Processing*, Seattle, USA, May 1998.

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

[5] P. S. R. Diniz, *Adaptive Filtering — Algorithms and Practical Implementation.* Norwell, MA: Kluwer Academic Publishers, 1997.