

MEAN-SQUARED ERROR ANALYSIS OF THE BINORMALIZED DATA-REUSING LMS ALGORITHM USING A DISCRETE-ANGULAR-DISTRIBUTION MODEL FOR THE INPUT SIGNAL

Marcello L. R. de Campos,[‡] José A. Apolinário Jr.,^{††} and Paulo S. R. Diniz,[†]

[‡]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
CEP 21945-970, Rio de Janeiro, RJ - Brazil

ABSTRACT

Providing a quantitative mean-squared-error analysis of adaptation algorithms is of great importance for determining their usefulness and for comparison with other algorithms. However, when the algorithm reutilizes previous data, such analysis becomes very involved as the independence assumption cannot be used. In this paper, a thorough mean-squared-error analysis of the binormalized data-reusing LMS algorithm is carried out. The analysis is based on a simplified model for the input-signal vector, assuming independence between the continuous radial probability distribution and the discrete angular probability distribution. Throughout the analysis only parallel and orthogonal input-signal vectors are used in order to obtain a closed-form formula for the excess mean-squared error. The formula agrees closely with simulation results even when the input-signal vector is a delay line. Furthermore, the analysis can be readily extended to other algorithms with expected similar accuracy.

1. INTRODUCTION

Increasing speed of convergence of adaptive filters invariably implies a corresponding increase in computational complexity of the adaptation algorithm. In many applications, gradient-type algorithms are not fast enough for a satisfactory performance and Newton-type algorithms are too complex for the given sampling frequency [1]. In these situations, a compromise solution may be the one which attempts to improve speed of convergence of a gradient-type algorithm while keeping the extra computation to a minimum. The data-reusing LMS (DR-LMS) algorithm is one approach that reutilizes available data repeatedly as many times as possible in order to achieve faster convergence as compared to the conventional LMS algorithm [2]. Improvements to the DR-LMS-algorithm performance can be obtained with an optimized step-size (e.g., NNDR-LMS and UNDR-LMS algorithms [3]).

The binormalized data-reusing LMS (BNDR-LMS) algorithm was introduced and briefly analyzed in [4] and [5]. Superior performance to that of other data-reusing algorithms may be expected due to normalization in orthogonal directions obtained from current and previous data vectors [4][5].

In this paper, a thorough analysis of the mean squared error (MSE) for the BNDR-LMS algorithm is carried out by using a simplified model for the input-signal vector. The model was first ap-

plied successfully to the analysis of the normalized LMS (NLMS) algorithm [6][7] and later to the analysis of a quasi-Newton algorithm [8]. In section II of this paper, the BNDR-LMS algorithm is briefly introduced. In section III, the model for the input-signal vector is described and discussed, and the mean-squared-error analysis is carried out. In section IV simulation results are provided supporting the analysis. Section V presents conclusions.

2. THE BNDR-LMS ALGORITHM

Derivation of the BNDR-LMS algorithm may be carried out from an optimization perspective, or from a geometrical perspective. We will briefly present the algorithm together with a derivation based on the first approach.

Let $\mathcal{S}(k)$ denote the hyperplane which contains all vectors \mathbf{w} such that $\mathbf{x}^T(k)\mathbf{w} = d(k)$. The solution given by the BNDR-LMS algorithm, $\mathbf{w}(k+1)$, is the one which belongs to $\mathcal{S}(k)$ and $\mathcal{S}(k-1)$ and is at a minimum distance from $\mathbf{w}(k)$, i.e., the one which solves

$$\min_{\mathbf{w}(k+1)} \|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2 \quad (1)$$

subjected to

$$\mathbf{x}^T(k)\mathbf{w}(k+1) = d(k) \quad (2)$$

and

$$\mathbf{x}^T(k-1)\mathbf{w}(k+1) = d(k-1) \quad (3)$$

The functional to be minimized is, therefore,

$$\begin{aligned} f[\mathbf{w}(k+1)] = & \|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2 \\ & + \lambda_1[d(k) - \mathbf{x}^T(k)\mathbf{w}(k+1)] \\ & + \lambda_2[d(k-1) - \mathbf{x}^T(k-1)\mathbf{w}(k+1)] \end{aligned} \quad (4)$$

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

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\lambda_1}{2}\mathbf{x}(k) + \frac{\lambda_2}{2}\mathbf{x}(k-1) \quad (5)$$

where

$$\frac{\lambda_1}{2} = \frac{[d(k) - \mathbf{x}^T(k)\mathbf{w}(k)]\|\mathbf{x}(k-1)\|^2}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} - \frac{[d(k-1) - \mathbf{x}^T(k-1)\mathbf{w}(k)]\mathbf{x}^T(k-1)\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} \quad (6)$$

and

$$\frac{\lambda_2}{2} = \frac{[d(k-1) - \mathbf{x}^T(k-1)\mathbf{w}(k)]\|\mathbf{x}(k)\|^2}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} - \frac{[d(k) - \mathbf{x}^T(k)\mathbf{w}(k)]\mathbf{x}^T(k-1)\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} \quad (7)$$

Equations (5)–(7) completely describe the algorithm.

3. MEAN-SQUARED-ERROR ANALYSIS

Analysis of data-reusing algorithms often involves difficulties that are not present in the analyses of other gradient-type or Newton-type algorithms, for reutilization of previous data vectors precludes application of the independence assumption [3]. Cumbersome expressions for the coefficient-error covariance matrix render the analysis not tractable and, consequently, pose difficulties for an accurate description of the algorithm performance [3]. This difficulty is overcome in the analysis presented here with the adoption of a simplified model for the input-signal vector which has discrete angular probability distribution. The excellent agreement between theoretical and simulation results suggests that analysis based on such model should be extended to other adaptation algorithms.

For the purposes of the analysis, we assume that an unknown FIR filter with coefficient vector given by \mathbf{w}_o is to be identified by an adaptive filter of same order employing the BNDR-LMS algorithm, i.e., $d(k)$ is modeled as

$$d(k) = \mathbf{x}^T(k)\mathbf{w}_o + n(k) \quad (8)$$

where $n(k)$ is measurement noise. It is also assumed that input signal and measurement noise are taken from independent and identically distributed zero mean white Gaussian noise processes with variances σ_x^2 and σ_n^2 , respectively.

We are interested in analyzing the behavior of the coefficient vector in terms of a step-size μ . Let

$$\Delta\mathbf{w}(k) = \mathbf{w}(k) - \mathbf{w}_o \quad (9)$$

be the error in the adaptive filter coefficients as related to the ideal coefficient vector. For the BNDR-LMS algorithm as described in (5)–(7), $\Delta\mathbf{w}(k+1)$ is given by

$$\Delta\mathbf{w}(k+1) = \Delta\mathbf{w}(k) + \mu \left[\frac{\lambda_1}{2}\mathbf{x}(k) + \frac{\lambda_2}{2}\mathbf{x}(k-1) \right] \quad (10)$$

From (8) and (5)–(7), we have

$$\Delta\mathbf{w}(k+1) = [\mathbf{I} + \mu\mathbf{A}]\Delta\mathbf{w}(k) + \mu\mathbf{b} \quad (11)$$

where

$$\mathbf{A} = \frac{\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{x}(k-1)\mathbf{x}^T(k-1)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} + \frac{\mathbf{x}(k-1)\mathbf{x}^T(k-1)\mathbf{x}(k)\mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} - \frac{\|\mathbf{x}(k-1)\|^2\mathbf{x}(k)\mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} + \frac{\|\mathbf{x}(k)\|^2\mathbf{x}(k-1)\mathbf{x}^T(k-1)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2} \quad (12)$$

and

$$\mathbf{b} = \frac{n(k)\|\mathbf{x}(k-1)\|^2 - n(k-1)\mathbf{x}^T(k)\mathbf{x}(k-1)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2}\mathbf{x}(k) + \frac{n(k-1)\|\mathbf{x}(k)\|^2 - n(k)\mathbf{x}^T(k-1)\mathbf{x}(k)}{\|\mathbf{x}(k)\|^2\|\mathbf{x}(k-1)\|^2 - [\mathbf{x}^T(k)\mathbf{x}(k-1)]^2}\mathbf{x}(k-1) \quad (13)$$

Analysis of convergence in the mean showed that the algorithm described above is unbiased and stable provided that $0 < \mu < 2$ [5]

Although $\Delta\mathbf{w}(k)$ converges in average to zero as k goes to infinite, which characterizes unbiasedness of the estimate, consistency of coefficient estimates can seldom be achieved for nonvanishing values of μ . In general, an excess of MSE, which depends on the second-order statistics of vector $\Delta\mathbf{w}(k)$, will be present. The excess of MSE is defined as [1]

$$\xi_{exc} = \lim_{k \rightarrow \infty} \xi(k) - \xi_{min} \quad (14)$$

where $\xi(k) = E[e^2(k)]$ and ξ_{min} is the minimum mean-squared error due to nonexact-modeling or presence of additive noise, or both [1]. The difference $\Delta\xi(k) = \xi(k) - \xi_{min}$ is known as excess in the MSE [1] and can be expressed as

$$\Delta\xi(k) = \text{tr}\{\mathbf{R}\text{cov}[\Delta\mathbf{w}(k)]\} \quad (15)$$

3.1. The Input-Signal-Vector Model

Evaluation of (15) becomes very involved if the input-signal vector is considered a delay line, even after using the independence assumption. An interesting alternative is the use of a simplified model for the input-signal vector $\mathbf{x}(k)$ which can be consistent with the first- and second-order statistics of a general input signal, but has a reduced and countable number of possible directions of excitation. This model was introduced in [6] and was successfully employed in [7] and [8]. The input-signal vector for the model is

$$\mathbf{x}(k) = s_k r_k \mathbf{V}_k \quad (16)$$

where:

- s_k is ± 1 with probability of occurrence $1/2$;
- r_k^2 has the same probability distribution function of $\|\mathbf{x}(k)\|^2$, or, for the case of interest, is a sample of an independent process with χ -square distribution of $(N+1)$ degrees of freedom, $E[r_k^2] = (N+1)\sigma_x^2$;

- \mathbf{V}_k is equal to one of the $N + 1$ orthonormal eigenvectors of \mathbf{R} , denoted $\mathcal{V}_i, i = 1, \dots, N + 1$. We will also assume that for a white Gaussian input signal, \mathbf{V}_k is uniformly distributed and, consequently, if $P(\cdot)$ denotes the probability of occurrence of event (\cdot) , then

$$P(\mathbf{V}_k = \mathcal{V}_i) = \frac{1}{N + 1} \quad (17)$$

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

$$\begin{aligned} \Delta\xi(k + 1) &= \Delta\xi(k + 1) \Big|_{\mathbf{x}(k) \parallel \mathbf{x}(k-1)} P[\mathbf{x}(k) \parallel \mathbf{x}(k - 1)] \\ &+ \Delta\xi(k + 1) \Big|_{\mathbf{x}(k) \perp \mathbf{x}(k-1)} P[\mathbf{x}(k) \perp \mathbf{x}(k - 1)] \end{aligned} \quad (18)$$

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

It is easy to verify that the BNDR-LMS algorithm behaves exactly like the NLMS algorithm when the input signal vector at instants k and $k - 1$ are parallel. In this case, the excess of MSE is given by [7]

$$\Delta\xi(k + 1)_{\parallel} = \left[1 + \frac{\mu(\mu - 2)}{N + 1} \right] \Delta\xi(k) + \frac{\mu^2 \sigma_n^2}{(N + 2 - \nu_x)} \quad (19)$$

where $\nu_x = E[x^4(k)/\sigma_x^4]$ is the *kurtosis* of the input signal, which varies from 1 for a binary distribution to 3 for a Gaussian distribution to ∞ for a Cauchy distribution [7]. It must be stressed, however, that (19) holds only for $\nu_x \ll N + 1$ [7].

For the case where $\mathbf{x}(k)$ and $\mathbf{x}(k - 1)$ are always orthogonal, we have, for $\mathbf{R} = \sigma_x^2 \mathbf{I}$, i.e., white-noise input signals (see Appendix A),

$$\begin{aligned} \Delta\xi(k + 1)_{\perp} &= \left[1 + \frac{\mu(\mu - 2)}{N + 1} \right] \Delta\xi(k) \\ &+ \frac{\mu(1 - \mu)^2(\mu - 2)}{N + 1} \Delta\xi(k - 1) + \frac{\mu^2(\mu - 2)^2}{N + 2 - \nu_x} \sigma_n^2 \end{aligned} \quad (20)$$

A final expression for the excess in the MSE may now be obtained from (19) and (20) combined and weighted accordingly, as suggested in (18). For a white input signal, the probabilities of $\mathbf{V}_k = \mathbf{V}_{k-1}$ and $\mathbf{V}_k \neq \mathbf{V}_{k-1}$ are equal to $\frac{1}{N+1}$ and $\frac{N}{N+1}$, respectively. The excess in the MSE is, therefore, given by

$$\begin{aligned} \Delta\xi(k + 1) &= \left[1 + \frac{\mu(\mu - 2)}{N + 1} \right] \Delta\xi(k) \\ &+ \frac{N\mu(1 - \mu)^2(\mu - 2)}{(N + 1)^2} \Delta\xi(k - 1) \\ &+ \frac{\mu^2 [1 + N(\mu - 2)^2]}{(N + 1)(N + 2 - \nu_x)} \sigma_n^2 \end{aligned} \quad (21)$$

The solution for $k \rightarrow \infty$ provides the magnitude of the excess of MSE

$$\xi_{exc} = \frac{\mu(N + 1) [1 + N(2 - \mu)^2] \sigma_n^2}{(N + 2 - \nu_x)(2 - \mu) [(2N + 1) - N\mu(2 - \mu)]} \quad (22)$$

4. SIMULATION RESULTS

Several simulations were run in order to verify how well the formulas obtained describe the behavior of the mean-squared error when the input-signal vector is a delay line. The application chosen was system identification, where a 63rd-order plant was to be identified by a same-order adaptive filter employing the BNDR-LMS algorithm. Additive measurement-noise variance was -60dB relative to the input-signal variance. The input signal in this case was white Gaussian noise and the excess of MSE was measured for different values of the step-size (μ varied from 0.1 to 1.9). The results are depicted in Fig. 1, where we can see a very close match between theoretical and simulation results, indicating accuracy of the analysis and suitability of the model used even when the input-signal-vector angular probability distribution is continuous and not independent of its radial distribution, as is the case for the delay line.

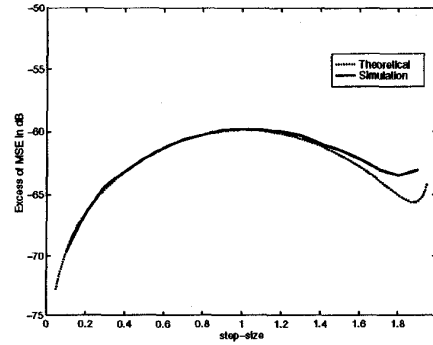


Figure 1: Excess of MSE for $N = 63$ as a function of μ .

5. CONCLUSIONS

The BNDR-LMS algorithm was analyzed with respect to the excess of MSE. The analysis was carried out with the assumption that the input-signal vector had discrete angular probability distribution independent from the continuous radial probability distribution. The closed-form formula obtained provides accurate results even when the input-signal vector is a delay line of samples of a white Gaussian process. The accuracy was verified in simulations for several values of the step-size in the region of stability.

APPENDIX A

1. Equation (20):

In the derivation of (20) $\mathbf{x}(k)$ and $\mathbf{x}(k - 1)$ were replaced by $s_k r_k \mathbf{V}_k$ and $s_{k-1} r_{k-1} \mathbf{V}_{k-1}$, respectively, with $\mathbf{V}_k \perp \mathbf{V}_{k-1}$. Furthermore, a second-order approximation for $E[1/r_k^2]$ was used [7], i.e.,

$$\begin{aligned} E \left[\frac{1}{\|\mathbf{x}(k)\|^2} \right] &= E \left[\frac{1}{\|\mathbf{x}(k-1)\|^2} \right] \\ &= E \left[\frac{1}{r_k^2} \right] \approx \frac{1}{(N + 2 - \nu_x) \sigma_x^2} \end{aligned} \quad (23)$$

where ν_x is the kurtosis of the input signal.

For $\mathbf{R} = \sigma_x^2 \mathbf{I}$, using (11) and (15) the expression for $\Delta\xi(k)$ may be rewritten as

$$\begin{aligned} \Delta\xi(k+1) &= \sigma_x^2 \text{tr} \left(E \left\{ [\mathbf{I} + \mu \mathbf{A}] \Delta \mathbf{w}(k) \right. \right. \\ &\quad \left. \left. \times \Delta \mathbf{w}^T(k) [\mathbf{I} + \mu \mathbf{A}] \right\} \right) \\ &\quad + \sigma_x^2 \text{tr} \left(E \left\{ \mu [\mathbf{I} + \mu \mathbf{A}] \Delta \mathbf{w}(k) \mathbf{b}^T \right\} \right) \\ &\quad + \sigma_x^2 \text{tr} \left(E \left\{ \mu \mathbf{b} \Delta \mathbf{w}^T(k) [\mathbf{I} + \mu \mathbf{A}] \right\} \right) \\ &\quad + \sigma_x^2 \text{tr} \left(E \left[\mu^2 \mathbf{b} \mathbf{b}^T \right] \right) \\ &= \rho_1 + \rho_2 + \rho_3 + \rho_4 \end{aligned} \quad (24)$$

Evaluating each of these terms separately we obtain

$$\begin{aligned} \rho_1 &= \sigma_x^2 \text{tr} \left\{ \text{cov}[\Delta \mathbf{w}(k)] \right\} \\ &\quad - \mu \sigma_x^2 \text{tr} \left\{ E \left[\frac{\mathbf{x}(k-1) \mathbf{x}^T(k-1)}{\|\mathbf{x}(k-1)\|^2} \right. \right. \\ &\quad \left. \left. \times \frac{\Delta \mathbf{w}(k) \Delta \mathbf{w}^T(k)}{\|\mathbf{x}(k-1)\|^2} \right] \right\} \\ &\quad - \mu \sigma_x^2 \text{tr} \left\{ E \left[\frac{\mathbf{x}(k) \mathbf{x}^T(k) \Delta \mathbf{w}(k) \Delta \mathbf{w}^T(k)}{\|\mathbf{x}(k)\|^2} \right] \right\} \\ &\quad - \mu \sigma_x^2 \text{tr} \left\{ E \left[\frac{\Delta \mathbf{w}(k) \Delta \mathbf{w}^T(k)}{\|\mathbf{x}(k-1)\|^2} \right. \right. \\ &\quad \left. \left. \times \frac{\mathbf{x}(k-1) \mathbf{x}^T(k-1)}{\|\mathbf{x}(k-1)\|^2} \right] \right\} \\ &\quad - \mu \sigma_x^2 \text{tr} \left\{ E \left[\frac{\Delta \mathbf{w}(k) \Delta \mathbf{w}^T(k) \mathbf{x}(k) \mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^2} \right] \right\} \\ &\quad + \mu^2 \sigma_x^2 \text{tr} \left(E \left\{ \frac{\mathbf{x}(k-1) \mathbf{x}^T(k-1) \Delta \mathbf{w}(k)}{\|\mathbf{x}(k-1)\|^2} \right. \right. \\ &\quad \left. \left. \times \frac{\Delta \mathbf{w}^T(k) \mathbf{x}(k-1) \mathbf{x}^T(k-1)}{\|\mathbf{x}(k-1)\|^2} \right\} \right) \\ &\quad + \mu^2 \sigma_x^2 \text{tr} \left\{ E \left[\mathbf{x}(k) \mathbf{x}^T(k) \Delta \mathbf{w}(k) \right. \right. \\ &\quad \left. \left. \times \Delta \mathbf{w}^T(k) \mathbf{x}(k-1) \mathbf{x}^T(k-1) \right] \right\} \\ &\quad + \mu^2 \sigma_x^2 \text{tr} \left\{ E \left[\mathbf{x}(k-1) \mathbf{x}^T(k-1) \Delta \mathbf{w}(k) \right. \right. \\ &\quad \left. \left. \times \Delta \mathbf{w}^T(k) \mathbf{x}(k) \mathbf{x}^T(k) \right] \right\} \\ &\quad + \mu^2 \sigma_x^2 \text{tr} \left(E \left\{ \frac{\mathbf{x}(k) \mathbf{x}^T(k) \Delta \mathbf{w}(k)}{\|\mathbf{x}(k)\|^2} \right. \right. \\ &\quad \left. \left. \times \frac{\Delta \mathbf{w}^T(k) \mathbf{x}(k) \mathbf{x}^T(k)}{\|\mathbf{x}(k)\|^2} \right\} \right) \\ &= \psi_1 + \psi_2 + \dots + \psi_9 \end{aligned} \quad (25)$$

where

$$\psi_1 = \Delta\xi(k) \quad (26)$$

$$\psi_2 = -\frac{\mu(1-\mu)^2}{(N+1)} \Delta\xi(k-1) - \frac{\mu^3 \sigma_n^2}{(N+2-\nu_x)} \quad (27)$$

$$\psi_3 = -\frac{\mu}{(N+1)} \Delta\xi(k) \quad (28)$$

$$\psi_4 = \psi_2 \quad \text{and} \quad \psi_5 = \psi_3 \quad (29)$$

$$\psi_6 = -\mu \psi_2 \quad (30)$$

$$\psi_7 = \psi_8 = 0 \quad (31)$$

$$\psi_9 = \frac{\mu^2}{N+1} \Delta\xi(k) \quad (32)$$

Therefore,

$$\begin{aligned} \rho_1 &= \left(1 + \frac{\mu(\mu-2)}{N+1} \right) \Delta\xi(k) \\ &\quad + \frac{(1-\mu)^2 \mu(\mu-2)}{N+1} \Delta\xi(k-1) + \frac{\mu^3(\mu-2)}{N+2-\nu_x} \sigma_n^2 \end{aligned} \quad (33)$$

Similarly,

$$\begin{aligned} \rho_2 &= \frac{\mu^2(1-\mu)}{(N+2-\nu_x)} \sigma_n^2 \\ &= \rho_3 \end{aligned} \quad (34)$$

$$\rho_4 = \frac{2\mu^2 \sigma_n^2}{(N+2-\nu_x)} \quad (35)$$

From (33)–(35) the difference equation for $\Delta\xi(k)$ is finally obtained as in (20).

6. REFERENCES

- [1] P. S. R. Diniz, *Adaptive Filtering — Algorithms and Practical Implementation*. Boston, MA: Kluwer Academic Press, 1997.
- [2] W. K. Jenkins, A. W. Hull, J. C. Strait, B. A. Schnauffer, and X. Li, *Advanced Concepts in Adaptive Signal Processing*. Boston, MA: Kluwer Academic Press, 1996.
- [3] B. A. Schnauffer, *Practical Techniques for Rapid and Reliable Real-Time Adaptive Filtering*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana-Champaign, USA, 1995.
- [4] J. A. Apolinário Jr., M. L. R. de Campos, and P. S. R. Diniz, "The binormalized data-reusing LMS algorithm," in *Proc. XV Simpósio Brasileiro de Telecomunicações*, in portuguese, Recife, Brazil, 1997.
- [5] 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, 1997.
- [6] D. T. Stocck, "On the convergence behaviour of the LMS and NLMS algorithms," in *Proc. 5th European Signal Processing Conference*, Barcelona, Spain, pp. 197–200, 1990.
- [7] D. T. Stocck, "On the convergence behaviour of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2811–2825, Sept. 1993.
- [8] M. L. R. de Campos and A. Antoniou, "A new quasi-Newton adaptive filtering algorithm," *IEEE Trans. Circuits and Systems — Part II*, Nov. 1997.