Nonlinear Quantization Effects in the LMS Algorithm - Analytical Models for the MSE Transient and Convergence Behavior

José Carlos M. Bermudez[†‡] and Neil J. Bershad[‡]

[†]Electronic Instrumentation Lab., Dept of Electrical Engineering, Fed. Univ. of Santa Catarina, Florianopolis, SC 88040-900, Brazil.
[‡]Department of Electrical and Computer Engineering, University of California, Irvine, Irvine, CA 92717, U.S.A..

Abstract

This paper[1] extends conditional moment techniques previously developed for the study of nonlinear versions of the LMS algorithm to the study of the effects of quantizers in the finite precision case. Deterministic nonlinear recursions are derived for the mean and second moment matrix of the weight vector about the Wiener weight for white gaussian data models and small algorithm step sizes μ. These recursions are solved numerically and shown to be in very close agreement with Monte Carlo simulations. Simulation examples are presented which demonstrate the accuracy of the theory in predicting the transient behavior and cancellation performance in steady-state for the quantized LMS algorithm.

1: Introduction

The least mean squares (LMS) algorithm is certainly one of the most popular algorithms for digital implementation of real-time high-speed adaptive filters. Fixed-point arithmetic is prevalent in such applications [1-8]. Many previous publications have studied the effects of a finite precision implementation on the behavior of the LMS algorithm.

Gitlin et al. [1] were the first to address the so called "stopping phenomenon". They compared the digital and analog LMS implementations for the least attainable residual mean-square errors (MSE). Caraiscos and Liu [2] presented steady-state analyses of the roundoff errors for fixed-point and floating-point arithmetic. Their analysis used a linear model for the correlation multiplier. This model approximates the quantization errors by uncorrelated additive white noise sources. Alexander [3] presented a finite precision analysis of the LMS algorithm which included the transient adaptation period. Here, as in [2], a linear model has been used for the quantization operation. The analysis in [3] was based upon an analytical model for the difference between the finite-precision and infinite-precision weight vectors. However, this error vector does not provide direct information regarding the mean-square output error.

Although the linear model is adequate during the early stages of adaptation, its validity lessens as the error decreases and the algorithm converges. However, it is in the convergence region of algorithm operation that the "stopping phenomenon" occurs. This phenomenon cannot be accurately predicted by the linear model. The "stopping phenomenon" is caused by the inputs to the various quantizers in the algorithm dropping below the least significant bit (LSB). It is an inherently nonlinear phenomenon and can be better predicted using a nonlinear model.

This paper extends the conditional moment techniques developed in [4-7] to the study of the nonlinear behavior of the quantizers in LMS adaptation.

1.1: Mathematical model of quantized LMS

The updating equation for the LMS algorithm is given by [1-8]

$$W_L(n+1) = W_L(n) + \mu\, \varepsilon(n) X(n) \qquad (1)$$

where

$\varepsilon(n) = d(n) + z(n) - W_L^T(n) X(n)$,

$X(n) = [x(n), x(n-1), x(n-2), \cdots, x(n-N+1)]^T$

:observed data vector with length N equal to the number of filter taps,

$W_L(n)$:weight vector at time n,

$d(n)$:desired scalar signal

$z(n)$:additive noise

1 This work was supported in part by the Brazilian Research Council (CNPq) under grant No. 201532/93-0.

In real-time high-speed digital signal processing applications, the LMS algorithm is usually implemented in a single serial digital signal processor [8]. In this situation, a quantization step exists after every multiplication. Also, the dominant roundoff errors arise from the quantizations realized at the weight updating step [1], [2], [3], [8]. Thus, finite precision effects shall be considered only in the updating equation.

The implementation of (1) using arbitrary values of μ leads to the LMS updating equation [8]

$$W_{Q_1}(n+1) = W_{Q_1}(n) + Q\left[Q\left[\mu\,\varepsilon(n)\right]X(n)\right] \qquad (2)$$

where $Q[\cdot]$ denotes a quantization operation. Note that the product $\mu\,\varepsilon(n)$ is the first to be quantized. The input signals are assumed to be properly scaled to avoid overflow errors due to additions.

It would be desirable to apply our mathematical techniques directly to (2). However, except for some special cases, it is not possible to obtain analytical expressions for expectations of nonlinear functions of nonlinear functions for the update. A possible candidate approximation for (2) using one less quantizer is the stochastic recursion

$$W_{Q_2}(n+1) = W_{Q_2}(n) + Q\left[\mu\,\varepsilon(n)\right]X(n) \qquad (3)$$

Equation (3) assumes that the quantized value of the double product $\mu\,\varepsilon(n)$ determines convergence. Indeed, it has been determined from extensive simulations [10] that the internal quantization $Q\left[\mu\,\varepsilon(n)\right]$ in (2) tends to determine the convergence properties of the algorithm for typical adaptation step sizes and signal power levels. Thus, an understanding of the theoretical behavior of (3) is of practical significance, since it will also predict the behavior of (2). Hence, this paper focuses in the analysis of the error-modified LMS algorithm given in (3).

To render the analysis more tractable, the following typical assumptions are made [5], [7]:

a) The data vector $X(n)$ is statistically independent over time. Hence, the present weight and data vectors are statistically independent. Also, $x(n)$ is a stationary zero-mean independent Gaussian sequence. Thus, the data covariance matrix $R_{XX} = E\left[X(n)X^T(n)\right] = \sigma_x^2 I$;

b) The desired data $d(n)$ is a stationary zero-mean Gaussian sequence, correlated with $X(n)$;

c) The noise sequence $z(n)$ is zero-mean, Gaussian and statistically independent of any other signal in the system.

Using these assumptions, the analysis leads to results that are representative of several practical applications [3], [5]. In digital data transmission, the sequences of binary digits are independent which leads to a white data model as a valid representation. The white Gaussian data case is also very common in system identification and channel equalization [3]. Even in applications with correlated data (such as voiceband echo cancellation), the white Gaussian data analysis retains sufficient information about the behavior of the adaptive process for the theoretical results to serve as reliable design guidelines [5], [7].

2: LMS algoritm with quantized update

A two's complement rounding quantizer with step-size Δ is assumed in the analysis. Fig. 1 shows the quantizer input/output relationship. Noticeable roundoff effects occur only for quantizer input magnitudes below Δ. Thus, the simplified quantizer model shown in Fig. 2 is used.

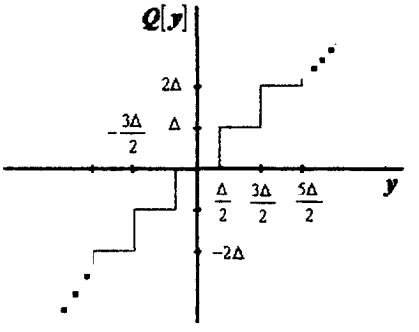

Fig. 1- Quantizer input/output relation

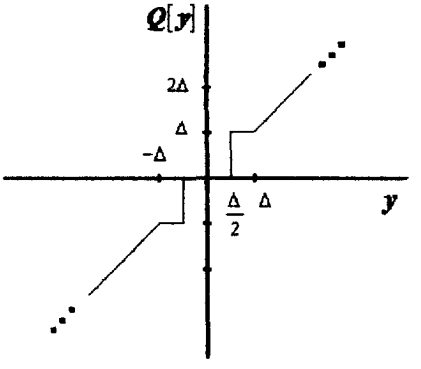

Fig. 2- Quantizer model

2.1: Mean behavior

It is mathematically more convenient to investigate the statistics of (3) about the optimum Wiener weight vector $W_0 = R_{XX}^{-1} R_{dX}$, where $R_{dX} = E[d(n)X(n)]$ and $R_{XX} = E[X(n)X^T(n)]$. Here, $E[\cdot]$ denotes statistical expectation.

Letting $V(n) = W_{Q_2}(n) - W_0$ and inserting into (3) yields

$$V(n+1) = V(n) + Q[\mu\,\varepsilon(n)]X(n) \tag{4}$$

Averaging both sides of (4) yields

$$\begin{aligned} E[V(n+1)] &= E[V(n)] \\ &+ E\left[Q\left[\mu\left\{d(n)+z(n)-V^T(n)X(n)-W_0^T X(n)\right\}\right]X(n)\right] \end{aligned} \tag{5}$$

The expectations in (5) are taken in two steps; first on the data and then on $V(n)$. Conditioning on $V(n)$ yields

$$E[V(n+1)|V(n)] = V(n) + E[Q[\mu\,\varepsilon(n)]X(n)|V(n)] \tag{6}$$

Now, conditioned on $V(n)$ and using (a)-(c), the error

$$\varepsilon(n) = d(n) + z(n) - V^T(n)X(n) - W_0^T X(n) \tag{7}$$

is a zero-mean Gaussian variable. Squaring and averaging (7), the MSE is given by

$$E[\varepsilon^2(n)] = \xi_0 + \sigma_x^2\,\mathrm{tr}[K_{VV}(n)] \tag{8}$$

where $\xi_0 = \sigma_d^2 + \sigma_z^2 - R_{dX}^T R_{XX}^{-1} R_{dX}$ is the MSE using the Wiener filter and $K_{VV}(n) = E[V(n)V^T(n)]$ is the correlation matrix of the weight-error vector $V(n)$.

Using Price's theorem [9] for Gaussian variates, the expectation on the right side of (6) yields [10]

$$E[V(n+1)|V(n)] = V(n) - \mu\,E[Q'[\mu\,\varepsilon(n)]|V(n)]R_{XX}V(n) \tag{9}$$

with

$$\begin{aligned} &E[Q'[\mu\,\varepsilon(n)]|V(n)] \\ &= 1 + \sqrt{\frac{2}{\pi}}\frac{\Delta}{\mu\,\sigma_{\varepsilon|V}}\,e^{-\frac{\Delta^2}{8\mu^2\sigma_{\varepsilon|V}^2}} - \mathrm{erf}\left(\frac{\Delta}{\sqrt{2}\mu\,\sigma_{\varepsilon|V}}\right) \end{aligned} \tag{10}$$

where $\sigma_{\varepsilon|V}^2 = E[\varepsilon^2(n)|V(n)]$, $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ and $Q'[y] = \frac{dQ[y]}{dy}$.

Averaging (9) over $V(n)$ and neglecting the correlation between $V(n)$ and $E[Q'[\mu\,\varepsilon(n)]|V(n)]$ yields

$$\begin{aligned} E[V(n+1)] &= E[V(n)] \\ &-\mu\,E\left[E[Q'[\mu\,\varepsilon(n)]|V(n)]\right]R_{XX}\,E[V(n)] \end{aligned} \tag{11}$$

For small μ, the weight fluctuations are small. Hence, $\sigma_{\varepsilon|V}^2(V(n))$ is concentrated near its mean. Thus, an accurate approximation for the expectation over $V(n)$ in (11) is obtained by replacing $\sigma_{\varepsilon|V}^2$ in (10) by its mean

$$E[\sigma_{\varepsilon|V}^2] = \sigma_\varepsilon^2 = E[\varepsilon^2(n)] = \xi_0 + \sigma_x^2\,\mathrm{tr}[K_{VV}(n)] \tag{12}$$

Thus, combining (9), (10), (11) and (12) yields

$$\begin{aligned} E[V(n+1)] = \Bigg\{1 - \mu\sigma_x^2\Bigg[1 + \sqrt{\frac{2}{\pi}}\frac{\Delta}{\mu\,\sigma_\varepsilon}\,e^{-\frac{\Delta^2}{8\mu^2\sigma_\varepsilon^2}} \\ -\mathrm{erf}\left(\frac{\Delta}{\sqrt{2}\mu\,\sigma_\varepsilon}\right)\Bigg]\Bigg\}E[V(n)] \end{aligned} \tag{13}$$

This recursion describes the mean behavior of the weight error vector during the transient adaptation phase. For $\Delta = 0$, (13) reduces to the infinite precision mean behavior equation for the white data case. Since (13) is a function of $\mathrm{tr}[K_{VV}(n)]$, a recursion is now derived to describe the statistical behavior of the weight fluctuations and the excess MSE.

2.2: Second moment behavior

Postmultiplying (4) by its transpose and averaging on the data (conditioned on $V(n)$) yields

$$\begin{aligned} E[V(n+1)V^T(n+1)|V(n)] &= V(n)V^T(n) \\ &+ E[Q[\mu\,\varepsilon(n)]V(n)X^T(n)|V(n)] \\ &+ E[Q[\mu\,\varepsilon(n)]X(n)V^T(n)|V(n)] \\ &+ E[Q^2[\mu\,\varepsilon(n)]X(n)X^T(n)|V(n)] \end{aligned} \tag{14}$$

Using Price's Theorem [9], [10] on the first two expectations on the right side of (14) yields

$$E\left[V(n+1)V^T(n+1)\middle|V(n)\right] = V(n)V^T(n)$$
$$-\mu E\left[Q'\left[\mu\varepsilon(n)\right]\middle|V(n)\right]V(n)V^T(n)R_{XX}$$
$$-\mu E\left[Q'\left[\mu\varepsilon(n)\right]\middle|V(n)\right]R_{XX}V(n)V^T(n) \quad (15)$$
$$+E\left[Q^2\left[\mu\varepsilon(n)\right]X(n)X^T(n)\middle|V(n)\right]$$

To determine a recursive equation for $K_{VV'}(n)$, the conditional expectations in (15) are evaluated and then averaged over $V(n)$. With the same reasoning used in section 2.1, it can be shown [10] that

$$\mathrm{tr}\left[K_{VV'}(n+1)\right] = (1-A)\,\mathrm{tr}\left[K_{VV'}(n)\right] + B \quad (16)$$

where

$$A = 2\mu\sigma_x^2\left\{1+\sqrt{\frac{2}{\pi}}\frac{\Delta}{\mu\sigma_\varepsilon}e^{-\frac{\Delta^2}{8\mu^2\sigma_\varepsilon^2}} - \mathrm{erf}\left(\frac{\Delta}{\sqrt{2}\mu\sigma_\varepsilon}\right)\right\} \quad (17)$$

$$B = N\sigma_x^2\mu^2\sigma_\varepsilon^2\left\{\left(\frac{\Delta}{\mu\sigma_\varepsilon}\right)^2\left[\mathrm{erf}\left(\frac{\Delta}{\sqrt{2}\mu\sigma_\varepsilon}\right) - \mathrm{erf}\left(\frac{\Delta}{2\sqrt{2}\mu\sigma_\varepsilon}\right)\right]\right.$$
$$\left. + 1 + \sqrt{\frac{2}{\pi}}\left(\frac{\Delta}{\mu\sigma_\varepsilon}\right)e^{-\frac{\Delta^2}{2\mu^2\sigma_\varepsilon^2}} - \mathrm{erf}\left(\frac{\Delta}{\sqrt{2}\mu\sigma_\varepsilon}\right)\right\} \quad (18)$$

Equation (16) describes the time evolution of $\mathrm{tr}\left[K_{VV'}(n)\right]$. Both $E[V(n+1)]$ and $\mathrm{tr}\left[K_{VV'}(n+1)\right]$ in (13) and (16), respectively, should be determined using (8) with $\mathrm{tr}\left[K_{VV'}(0)\right] = V^T(0)V(0)$. Note that (13) is not needed to determine the MSE performance.

3: Simulation examples

Fig. 3 depicts a simple system identification problem. Here W^* is the weight vector to be identified. The components of W^* are comprised of the values of 13 equally spaced samples of a time-delayed raised-cosine function. Fig. 4 displays Monte Carlo simulations (100 runs) of $\mathrm{tr}\left[K_{VV'}(n)\right]$ for $N = 13$, $\sigma_x^2 = 1/9$, $\mu = 0.025$, $\sigma_z^2 = E\left[z^2(n)\right] = \xi_0 = 10^{-12}$ and several values of $\Delta = 2^{-b}$. The quantizer shown in Fig. 1 has been used in all simulations. The theoretical curves were determined using (16). The theoretical predictions and the simulation results are in excellent agreement. Fig. 5 presents the time evolution of the MSE for the same parameters. Fig. 6 displays the theoretical MSE curves determined from (8) and (16) and the simulation results obtained using the two-quantizer update equation (2). Clearly, the analytical results derived from (3) can be used to predict the behavior of the quantized LMS algorithm.

A large number of simulations were run for a wide range of all the parameter values. In all cases the same type of agreement was obtained between theoretical and simulation results as shown in Figs. 4, 5 and 6

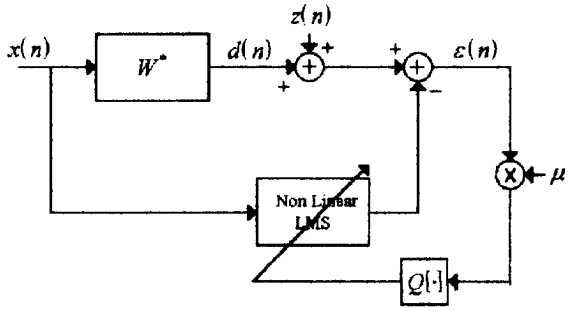

Fig. 3- System identification model

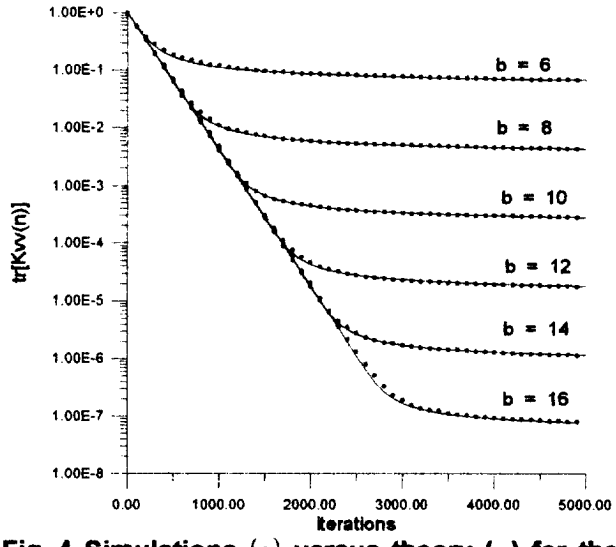

Fig. 4. Simulations (•) versus theory (–) for the time evolution of $\mathrm{tr}\left[K_{VV'}(n)\right]$. Simulations using updating equation (3) and quantizer of Fig. 1 ($\Delta = 2^{-b}$).

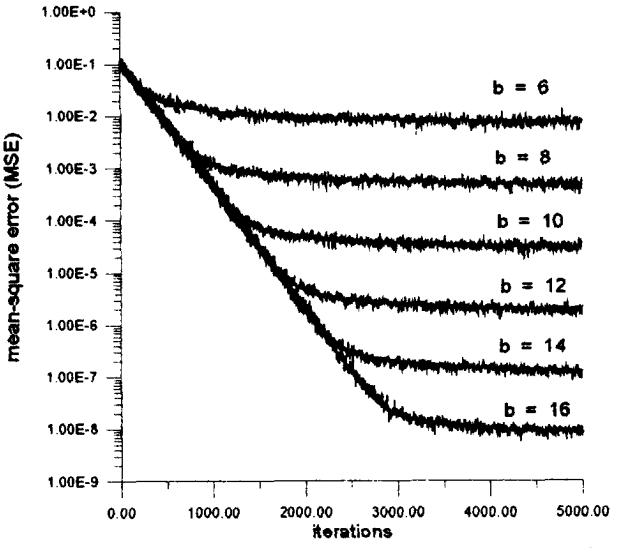

Fig. 5. Simulations (—) versus theory (--) for the mean-square error. Simulations using updating equation (3) and quantizer of Fig. 1 ($\Delta = 2^{-b}$).

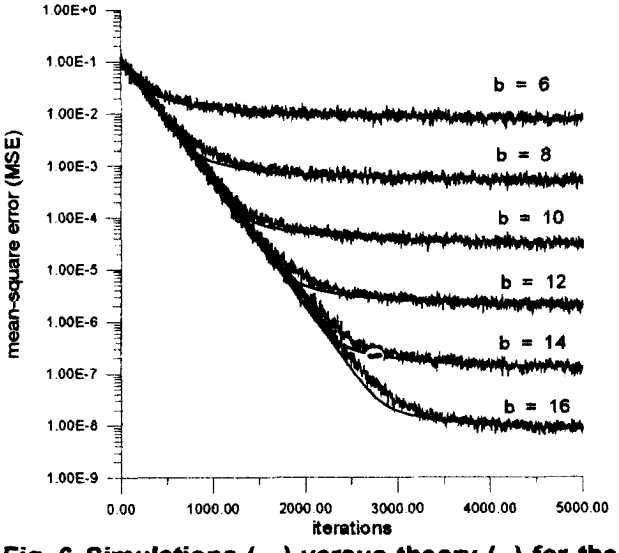

Fig. 6. Simulations (—) versus theory (--) for the mean-square error. Simulations using updating equation (2) and quantizer of Fig. 1 ($\Delta = 2^{-b}$).

4: Conclusions

This paper presented a study of the quantization effects in the finite precision LMS algorithm with arbitrary step sizes. Based on extensive simulation results, a single-quantizer approximation for the finite precision LMS updating equation has been proposed. Conditional moment techniques previously developed for the study of nonlinear versions of the LMS algorithm were extended to the analysis of the quantization effects. Deterministic nonlinear recursions were derived for the mean and second moment matrix of the weight vector about the Wiener weight for white gaussian data models and small algorithm step sizes μ. These recursions were solved numerically and shown to be in very close agreement with the Monte Carlo simulations during all phases of the adaptation process.

Acknowledgement

The authors would like to thank Prof. Rui Seara of Federal University of Santa Catarina for the helpful discussions regarding the digital implementation of the LMS algorithm.

References

[1] R. D. Gitlin, J. E. Mazo and M. G. Taylor, "On the design of gradient algorithms for digitally implemented adaptive filters," *IEEE Trans. on Circuit Theory*, vol. CT-20, pp. 125-136, March 1973.

[2] C. Caraiscos and B. Liu, "A roundoff error analysis of the LMS adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, No. 1, pp. 34-41, February 1984.

[3] S. T. Alexander, "Transient weight misadjustment properties for the finite precision LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-53, No. 9, pp. 1250-1258, September 1987.

[4] N. J. Bershad, "On weight update saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, No. 4, pp. 623-630, April 1990.

[5] D. L. Duttweiller, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, No. 4, pp. 578-586, August 1982.

[6] N. J. Bershad, "On the optimum data nonlinearity in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, No. 1, pp. 69-76, February 1986.

[7] N. J. Bershad, "On error-saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, No. 4, pp. 440-452, April 1988.

[8] J. M. Cioffi, "Limited-precision effects in adaptive filtering," *IEEE Trans. on Circuits and Systems*, vol. CAS-34, No. 7, pp. 821-833, July 1987.

[9] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IEEE Trans. Inform. Theory*, vol. IT-4, 1958.

[10] J. C. M. Bermudez and N. J. Bershad, "A Nonlinear analytical model for the quantized LMS algorithm - the arbitrary step size case," paper submitted to the *IEEE Trans. on Signal Processing*, June 1994.