A Variable Step-Size LMS Algorithm

BOŽO KRSTAJIĆ, LJUBIŠA STANKOVIĆ, ZDRAVKO USKOKOVIĆ Department of Electrical Engineering University of Montengro Cetinjski put bb., 81000 Podgorica, Montenegro YUGOSLAVIA

Abstract: - The paper proposes a new adaptive VS LMS algorithm, obtained by combining LMS algorithms with different step sizes without calculating their weighting coefficients. As a criterion for choosing the VS LMS algorithm step size, we take the ratio between the weighting coefficients' bias and variance. Identification of an unknown system in nonstationary noisy environment is performed and simulations with the proposed and other VS LMS algorithms are presented. Simulation results confirm the favorable properties of the proposed algorithm in nonstationary environment with abrupt changes of unknown system parameters.

Key-Words: -Adaptive LMS algorithm, variable step size, bias and variance of weighting coefficients. System identification

1 Introduction

There is a number of adaptive algorithms, [1,2,3,4,6,8], derived from the conventional LMS algorithm. Variable step-size methods [4, 5, 6] aim to improve the convergence of the LMS algorithm, while preserving the steady-state performance. There are several criteria for varying the LMS algorithm step-size: sign changes of successive samples of the gradient [4], squared instantaneous error and square of a time-averaging estimate of the successive error autocorrelation [5]. These criteria lead to the corresponding types of VS LMS algorithms, which are more efficient than the LMS algorithm for coefficient's tracking in a nonstationary environment.

In [8] we proposed one more possible approach for the LMS-based adaptive algorithm performance improvement. The idea is to combine various adaptive LMS-based algorithms in order to obtain suitable adaptive system that represents the best combination of the algorithms. The solution proposed in [8] produced favorable results, but it was associated with a parallel calculation of weighting coefficients for several LMSbased adaptive algorithms. Elements for estimating the quality of an adaptive algorithm are recognized in a specific statistical approach based on the comparison of bias and variance, [7], of the adaptive system's weighting coefficients, viewed as random variables.

Based on the idea of combining several LMS algorithms, we here propose a new way of choosing the step size for the VS LMS algorithm.

2 LMS Algorithm

Let us define the input signal vector:

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

and the vector of weighting coefficients:

 $\overline{W}(k) = [W_0(k) \ W_1(k) \ W_2(k) \dots \ W_{N-1}(k)]^{\mathrm{T}}$

of the adaptive filter at an instant k, [1,2,3]. The output error is: $e(k) = d(k) - \overline{W}^T(k)\overline{X}(k)$, where d(k) is the reference signal. The filter output signal is given by $v(k) = \overline{W}^T(k)\overline{X}(k)$.

The weighting coefficients in the LMS algorithm are obtained from the following expression:

$$W(k+1) = W(k) + 2\mu e(k)X(k),$$
(1)

where μ is the algorithm step size.

The considered adaptive system identification problem consists in trying to adjust a set of weighting coefficients so that the system output tracks a reference signal, $d(k) = \overline{W}^{*T}(k)\overline{X}(k) + n(k)$, where n(k) is a zeromean Gaussian independent noise with the variance σ_n^2 , and $\overline{W}^*(k)$ is the optimal weighting vector (Wiener vector). In a nonstationary environment $\overline{W}^*(k)$ is timevarying, and is described as, [3,5,9]:

$$\overline{W}^{*}(k+1) = \overline{W}^{*}(k) + \overline{Z}(k), \qquad (2)$$

where $\overline{Z}(k)$ is a zero-mean sequence, independent on $\overline{X}(k)$ and n(k), with covariance matrix $Q = \sigma_Z^2 \overline{I}$.

An important performance measure for the adaptive filters is its steady state mean square deviation (MSD). For the adaptive LMS filters in nonstationary environment, with standard assumptions as in [3,6], the MSD is given by:

$$MSD = \lim_{k \to \infty} E[\overline{V^T}(k)\overline{V}(k)] = \mu \sigma_n^2 N + \frac{tr(QR^{-1})}{\mu}, \quad (3)$$

where $\overline{V}(k) = \overline{W}(k) - \overline{W}^*(k)$ is the weighting vector deviation and *R* is the autocorrelation matrix of the input signal vector $\overline{X}(k)$. The first term in MSD, directly proportional to the algorithm step size μ , is a consequence of the gradient noise. The second one, inversely proportional to μ , stems from the optimal filter variations, (2). Note that the analysis for the stationary case directly follows from the above relations, for *Q* null matrix.

The misalignment, as we show, is due to both the effects of gradient noise (weighting coefficients variations around the average value) and the weighting vector lag (difference between the average and the optimal value), [3,6]. It can be written, for the *i*-th weighting coefficient:

$$V_{i}(k) = (E(W_{i}(k)) - W_{i}^{*}(k)) + (W_{i}(k) - E(W_{i}(k)))$$

= bias_{i}(k) + p_{i}(k), (4)

where $p_i(k)$ is a zero-mean random variable with the variance $\sigma_i^2(k)$. Now, from (3) and (4) we can express bias and variance for the *i*-th weighting coefficient, in steady state, as:

$$bias_i^2 = \frac{\sigma_z^2}{\mu \sigma_x^2}, \quad \sigma_i^2 = \mu \sigma_n^2, \quad (5)$$

where σ_x^2 is the variance of white zero mean input signal $\overline{X}(k)$.

As shown in [3,6], from (3) it is possible to obtain the optimal algorithm step size for each coefficient:

$$\mu_i^{opt} = \frac{\sigma_z}{\sigma_n \sigma_x},\tag{6}$$

that minimizes the value of the MSD.

The above relation is not of much use in practice, since in each iteration it requires availability of data concerning optimal filter variations (2), as well as of the input signal variance. If the optimal step size (6) is substituted into (5), the equality of bias and variance easily follows, i.e.

$$bias_i \Big|_{\mu_i^{opt}} = \sigma_i \Big|_{\mu_i^{opt}}.$$
(7)

3 Criterion for Comparison of LMS Algorithms

Basic idea of the combined adaptive system lies in parallel and independent adaptation of two or more LMS adaptive algorithms, with the choice of the best weighting coefficient value in each iteration [8].

Criterion for choosing the best weighting coefficient value is based on comparing the values of

weighting coefficients' bias and variance for each of the parallel algorithms. The weighting coefficients $W_i(k)$, for any parallel adaptive algorithm, assume random values around the optimal values $W_i^*(k)$, with the bias $bias(W_i(k)) = bias_i$ and the variance $\sigma_{W_i(k)} = \sigma_i$. These two quantities are related as [7,8]:

$$\left|W_{i}(k) - W_{i}^{*}(k) - bias_{i}\right| \leq \kappa \sigma_{i}.$$
(8)

The above inequality holds with a probability $P(\kappa)$, depending on the value of parameter κ . For example, with $\kappa = 2$ and a Gaussian distribution of random variable $W_i(k)$, inequality (8) is satisfied with 95% probability, since P(2) is 0.95.

Let us now define the confidence intervals for random variables $W_i(k)$, [7,8]:

$$D_i(k) = [W_i(k) - (\kappa + \Delta \kappa)\sigma_i, W_i(k) + (\kappa + \Delta \kappa)\sigma_i], (9)$$

where $\Delta \kappa$ is the parameter which takes into account the bias of $W_i(k)$. Then, from (8) and (9) we can conclude that, as long as the bias is small, i.e. $|bias_i| < \Delta \kappa \sigma_i$, the optimal value $W_i^*(k)$ belongs to the confidence interval $D_i(k)$, independently on parameters of the LMS algorithm. It means that, for small bias, the confidence intervals (9) for different LMS algorithms intersect. When, on the other hand, the bias becomes large, then the central positions of the intervals are far apart for different algorithms, and they do not intersect.

Since we do not have apriori information about the $bias_i$, we will use a specific statistical approach [7,8]. Taking, as a criterion for the bias/variance trade-off, the condition that the bias and variance are of the same order of magnitude [7], i.e.,

$$|bias_i| \cong \Delta \kappa \sigma_i,$$
 (10)

we get the criterion for the choice of LMS adaptive algorithm. Comparison of (7) with (10) indicates that $\Delta \kappa = 1$ should be used in our application. This criterion is in accord with (7), the analytical expression for the optimal step size.

Let us consider two LMS adaptive algorithms whose coefficients are calculated with different step sizes p and $q \mu_{\text{max}} \ge p > q$. We check if the confidence intervals (9), for both algorithms in each iteration, intersect. According to (8), (9) and (10), this reduces to the check of the following inequality

$$\left|W_{i}^{p}(k) - W_{i}^{q}(k)\right| < (\kappa + 1)(\sigma_{i}^{p} + \sigma_{i}^{q}).$$
(11)

If the intervals intersect (i.e., (11) is satisfied), it means that the bias is already small (corresponding to the steady state) so we should choose the weighting coefficient with smaller variance or step size, i.e. $W_i^q(k)$. If the intervals do not intersect (bias is large) we choose the weighting coefficient with the smallest bias (larger step size), $W_i^p(k)$, i.e. with best tracking of optimal filter rapid variations.

We take the value of the variance $(\sigma_i^{\mu})^2$ at an instant k, for the LMS with the step size μ , to be equal to the variance value in the steady state. Note that the estimate of the variances $(\sigma_i^{\mu})^2$ obtained by using (5), depend on the particular problem. It should be performed before the confidence intervals check. Details of this estimation will be given later.

4 New VS LMS Adaptive Algorithm

The VS LMS algorithm has the same form as the standard LMS algorithm, but in the course of adaptation the algorithm step size μ is changed, [5,6]. By changing the algorithm step size one changes the convergence speed, as well as the amount of variation of the weighting coefficients around the optimal value.

Formally, the VS LMS algorithm uses coefficients update of the form:

$$\overline{W}(k+1) = \overline{W}(k) + 2e(k)\overline{M}(k)\overline{X}(k), \quad (12)$$

where $\overline{M}(k) = diag(\mu_0(k), \mu_1(k), \mu_2(k), \dots, \mu_{N-1}(k))$ is a diagonal matrix of step sizes for individual weighting coefficients at the instant *k*.

In order to derive our VS LMS algorithm, let us apply the presented combining method to two standard LMS adaptive algorithms with different step sizes. Let the first one have the maximal step size value μ_{max} which does not violate the algorithm convergence condition [1,2,3], while the second one is characterized in each iteration by the variable step size $\mu_i(k)$.

The analysis from the previous section may now be applied to these two algorithms. After choosing the better algorithm, based on the proposed criterion, both algorithms will, in each iteration, take the set of better values of the weighting coefficients as a starting point for the next iteration.

Thus, according to (12), denote the *i*-th weighting coefficient at an instant *k* by $W_i^p(k)$, $W_i^q(k)$, for the first and the second LMS algorithm, respectively. Weighting coefficients for these algorithms would be calculated, in each iteration, according to the relations

$$W_i^{p}(k+1) = W_b + 2\mu_{\max} e(k)x(k-i), \quad (13)$$

$$W_i^q(k+1) = W_b + 2\mu_i(k)e(k)x(k-i), \quad (14)$$

where W_b is the coefficient value selected as the best choice from the previous iteration.

In order to define the criterion for choosing better

weighting coefficient and the step size value, at an instant k+1, let us substitute (5), (13) and (14) into (11). We get the inequality:

$$a(k)(\sqrt{\mu_{\max}} - \sqrt{\mu_i(k)}) < \frac{(\kappa+1)}{2}\sigma_n, \quad (15)$$

where a(k) = |e(k)x(k-1)| represents half of the estimated *i*-th coordinate of the performance criterion gradient for the LMS algorithm, [1,2,3].

The best bias-to-variance ratio is obtained for the particular step size that turns (15) into an equality. By solving inequality (15), we arrive at the relation for the step size calculation for the *i*-th weighting coefficient in the *k*-th iteration:

$$\sqrt{\mu_i(k)} = \begin{cases} \sqrt{\mu_{\max}} - \frac{(\kappa + 1)\sigma_n}{2a(k)} & \text{if } \mu_i(k) > \mu_{\min} \\ \sqrt{\mu_{\min}} & \text{if } \mu_i(k) \le \mu_{\min} \end{cases}$$
(16)

Analysis of (16) leads to the idea of avoiding advance calculation of the weighting coefficients, i.e. the parallel LMS algorithms. Instead, based on the calculation of the parameter a(k) and taking into account (16), we determine the value $\mu_i(k)$, i.e. a more appropriate step size of the LMS algorithm for each weighting coefficient at each instant of time. Obviously, this is just the idea of a VS LMS algorithm.



Figure 1: MSD for considered algorithms

In our simulations we have estimated the value of σ_n^2 by using relation (5), where the variance σ_i^2 is obtained by, [7]:

$$\sigma_i = \frac{median(|W_i(k) - W_i(k-1)|)}{0.6745\sqrt{2}},$$
 (17)

for k=1,2,...,L and $\sigma_z^2 << \sigma_i^2$. The above relation produces good estimates for all stationary cases, as well as for the indicated nonstationary ones, including abrupt

coefficient changes.

Note that any other estimation of σ_n^2 is valid for the algorithm. Namely, although it somewhat affects the step size choice precision, this error does not significantly affect the overall algorithm performance, as shown in [7].



Figure 2: Step size variations for the considered algorithms

5 Simulation Results

The proposed new variable step size LMS (NVSS) algorithm is implemented for nonstationary environments in a system identification setup. The performance of the algorithm is compared with the Harris variable step size LMS algorithm (HVSS), [5], and the robust variable step size LMS (RVSS) algorithm, [5], which has been shown to have better performance than other known variable step size algorithms. Parameters of this algorithm are taken from [5].

In all simulations presented here, the reference signal d(k) is corrupted by a zero-mean uncorrelated Gaussian noise with variance σ_n^2 . The unknown system has four time-invariant coefficients, and the FIR adaptive filters are of the same order. Both considered filters are excited by a zero-mean, white Gaussian signal of unity variance. The optimal weighting vector is nonstationary and generated according to the random walk model (2), with $\sigma_Z^2 = 0.0001$. However, we consider a more complex case with an additional abrupt change of optimal weighting vector. The abrupt change is generated by multiplying all the system coefficients by -1 at the very middle of the adaptive procedure. Results are obtained by averaging over 500 independent runs, as in [5].

As it may be observed from the presented results,

in the first 30 iterations the noise power was estimated according to the proposed solution and the relation (17). Maximal and minimal step size of each algorithms are $\mu_{\text{max}} = 0.11$ and $\mu_{\text{min}} = \mu_{\text{max}} / 1000$ The value of parameter κ was $\kappa = 1.75$.

The RVSS and HVSS have the following parameters: α =0.97, β =0.8, γ =0.0027, m_0 =3 and m_1 =3, with *SNR*=20*dB*. Figure 1 shows the MSD characteristics for each algorithm. In order to clearly compare the obtained results, for each simulation we calculated the average MSD (*MSD*_a). The corresponding values of the average MSD are *MSD*_a=0.60684, *MSD*_a=0.68004 and *MSD*_a=1.55603, for the NVSS the RVSS and the HVSS, respectively. Thus, the NVSS has the best overall performance. Figure 2 shows the average step sizes for RVSS, HVSS and NVSS, respectively. As it can be observed, the advantage of the proposed algorithm stems from more appropriate changes of its step size values.

6 Conclusion

The proposed VS LMS algorithm differs from the other VS LMS algorithms only in the criterion for the step size change. The new criterion is based on the equality of bias and variance of the weighting coefficients. This way we aim to improve the step size as much as possible, with respect to the minimal mean square deviation of weighting coefficients. This choice of the criterion for the step size variation contributed to considerable improvements in the adaptive algorithm performance, since the current step size value is calculated independently on its previous values. Presented simulation results and analysis show advantages of the proposed solution with respect to other known algorithms for various values of system and algorithm parameters.

7 Acknowledgment

This work is supported by the Volkswagen Stiftung, Federal Republic of Germany.

References:

- [1] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, N.J. 07632.X1. 1993.
- [2] P. S. R. Diniz, Adaptive Filtering: Algorithms and Practical Implementation, Kluwer Academic Publishers, Norwell, 1999.
- [3] B. Widrow, J. M. McCool, M. G. Larimore and C. R. Johnson, *Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter*, Proc. of the IEEE, Vol. 64, pp. 1151-1161, 1976..

- [4] R.W. Harris, D.M. Chabries and F. A. Bishop, A Variable Step (VS) Adaptive Filter Algorithm, IEEE Trans. on Acoustics, Speech and Sig. Processing, vol. ASSP 34, no.2, pp.309-316, April 1986.
- [5] T. Aboulnasr and K. Mayyas, A Robust Variable Step-Size LMS-Type Algorithm: Analysis and Simulations, IEEE Trans. on Signal Proc., vol. 45, No. 3, pp.631-639, March 1997.
- [6] E. Eweda, Comparison of RLS, LMS and Sign Algorithms for Tracking Randomly Time-Varying Channels, IEEE Trans. on Signal Proc., vol. 42, No. 11, pp. 2937-2944, November, 1994.
- [7] LJ. Stanković, V. Katkovnik, Algorithm for the instantaneous frequency estimation using timefrequency distributions with variable window width, IEEE Sign. Proc. Letters, vol. 5, No. 9, pp. 224-227 Sept. 1998.
- [8] B. Krstajić, LJ. Stanković, Z. Uskoković, I. Djurović, Combined adaptive system for identification of unknown systems with varying parameters in a noisy environment, 6th ICECS, Conf. Proceedings; vol. II, 745-748, Pafos, Cyprus, 1999.
- [9] N. Yousef and A. H. Sayed, A Unified Approach to the Steady-State and Tracking Analyses of Adaptive Filters, IEEE Transac. On Sig. Proc. vol. 49, No. 2, pp. 314-324, February 2001.