A NOVEL ADAPTIVE STEP SIZE CONTROL ALGORITHM FOR ADAPTIVE FILTERS

Shin'ichi Koike

NEC Corporation

7-1, Shiba 5-chome, Minato-ku, Tokyo 108-8001 Japan

Fax : +81 3 3798 9213 E-mail : koike@epc.ho.nec.co.jp

ABSTRACT

A novel adaptive step size control algorithm is proposed, in which the step size is approximated to the theoretically optimum value via leaky accumulators, realizing *quasi-optimal* control. The algorithm is applicable to most of the known tap weight adaptation algorithms. Analysis yields a set of difference equations for theoretically calculating expected filter convergence, and derives residual mean squared error (MSE) after convergence in a formula explicitly solved. Experiment with some examples proves that the proposed algorithm is highly effective in improving the convergence rate. The theoretically calculated convergence is shown to be in good agreement with that obtained through simulations.

I. INTRODUCTION

Many types of tap weight adaptation algorithms have been proposed for use in FIR adaptive filters; Least Mean Square (LMS), Least Mean Fourth (LMF), Sign Algorithm (SA), Signed Regressor Algorithm (SRA), Sign-Sign Algorithm (SSA), *etc.* [1] – [6].

It is well known that, for any algorithm with a fixed value of tap weight adaptation step size, a *trade-off* between the filter convergence rate and the steady-state error does exist. If a larger value of the step size is used, then a faster convergence is attained as long as the filter remains stable. On the other hand, the smaller the step size, the more accurate the estimation in the presence of observation noise.

Consequently, if we adaptively control the step size so that it stay large in the early stages of filter convergence and become smaller as the convergence proceeds, *both* fast convergence and low estimation error could be realized. Along this line of thought, many kinds of adaptive step size control algorithms have been proposed and studied.

Mathews and Xie proposed an adaptive step size control algorithm based on the gradient of "instantaneous" error [7]. [8] proposed a method using correlation between error signal and replica (filter output), where we recognize that the correlation is large in the early stages of convergence but much smaller after convergence. [9] proposed a variable step size LMS algorithm in which the step size is controlled in relation to the error signal power. Aboulnasr and Mayyas improved the previous method using the correlation of error samples at different time instants, thus mitigating the adverse effect of the noise [10]. Recently, [11] has proposed a unique algorithm based on a cost function with variable error power.

Most of the step size control algorithms above are proven effective, to some extent, in the sense that the value of the step size *decreases* as the convergence proceeds. However, none of them, except the stochastic algorithm in [7], assures that the step size is *optimum* at any time instant along the convergence process. This implies that the adaptive step size control algorithms so far proposed are considered to be, more or less, *qualitative.*

Therefore, this paper seeks a *quantitative* approach, namely, we try to develop adaptive algorithm which gives us not just a *decreasing* step size but a step size value as *close to the theoretical optimum* as possible at each time instant.

The paper is organized as follows. In Section II, a novel adaptive step size control algorithm is proposed in a general form. Section III develops analysis, where it is shown why the proposed method gives *quasi-optimum* value of the step size. Difference equations are derived for theoretically calculating the filter convergence process for Least Mean Fourth, Signed Regressor and Sign-Sign Algorithms. Residual Mean Squared Error (MSE) after convergence is further obtained from the difference equations. Section IV gives results of experiment with some practical examples. Section V concludes the paper.

II. PROPOSED ADAPTIVE STEP SIZE CONTROL ALGORITHM

Tap weight update equation for FIR adaptive filters is given by the following general formula.

$$c^{(n+1)} = c^{(n)} + \alpha_c^{(n)} f(e_n + \nu_n) g(a^{(n)}), \qquad (1)$$

where

$c^{(n)} = [\, c_0^{(n)},\ c_1^{(n)},\ \cdots,\ c_{N-1}^{(n)}]^T$

tap weight vector at time n,

$a^{(n)} = [\, a_n,\ a_{n-1},\ \cdots,\ a_{n-N+1}]^T$

reference input vector at time n (length N),

a_n reference input signal (stochastic process, *colored* in general),

e_n error signal,

v_n additive noise,

n time instant,

N number of taps

$\alpha_c^{(n)}$ step size at time n,

$f(\cdot), g(\cdot)$

odd functions, nonlinear in general,

$g(a^{(n)}) = [g(a_n), g(a_{n-1}), \ldots, g(a_{n-N+1})]^T$, and

$(\cdot)^T$ transpose of a vector or a matrix.

The proposed adaptive step size control algorithm is described as follows. First the step size at time n is calculated as

$$\alpha_c^{(n)} = q_0^{(n)T} q^{(n)} / \tau^{(n)}, \qquad (2)$$

where $q_0^{(n)T} q^{(n)}$ denotes inner product of vectors $q_0^{(n)}$ and $q^{(n)}$ each of length N, and $\tau^{(n)}$ is a scalar. $q_0^{(n)}$, $q^{(n)}$ and $\tau^{(n)}$ are updated by leaky accumulators as

$$q_0^{(n+1)} = (1-\rho) q_0^{(n)} + \rho (e_n + v_n) a^{(n)}, \qquad (3)$$

$$q^{(n+1)} = (1-\rho) q^{(n)} + \rho f(e_n + v_n) g(a^{(n)}), \qquad (4)$$

and

$$\tau^{(n+1)} = (1-\rho_\tau) \tau^{(n)} + \rho_\tau \{ f(e_n + v_n) g(a^{(n)})^T a^{(n-Ld)} \}^2 . \qquad (5)$$

Here, ρ and ρ_τ are leakage factors, and the delay $Ld (\geq 1)$ is chosen sufficiently large so that $E[a_n a_{n-Ld}] \cong 0$ holds for a *colored* reference input.

It will be shown in the next section why the step size given by (2) combined with (3) through (5) is close to the theoretical optimum or *quasi-optimum*.

Assume that the leakage factors are given in an integer power of 2. To update $q_0^{(n)}$ in (3), N *Multiplications* are necessary. If the vector $f(e_n + v_n) g(a^{(n)})$ in (1) is calculated and stored for reuse in (4) and (5), $N+1$ *Multiplications* are needed to update $\tau^{(n)}$. Therefore, to calculate the step size at each time n, $3N+1$ *Multiplications* and 1 *Division* in total are required.

III. ANALYSIS

Suppose that the adaptive filter is applied to identification of an unknown time-invariant system whose response vector is $h = [h_0, h_1, \ldots, h_{N-1}]^T$. Defining "tap weight error" vector $\theta^{(n)} = h - c^{(n)}$, we find from (1)

$$\theta^{(n+1)} = \theta^{(n)} - \alpha_c^{(n)} f(e_n + v_n) g(a^{(n)}) \qquad (6)$$

and

$$e_n = a^{(n)T} \theta^{(n)}. \qquad (7)$$

Now, let us assume that the reference input is a *stationary Gaussian* process, *colored* in general, with covariance matrix $\mathbf{R}_a = E[a^{(n)} a^{(n)T}]$ and variance σ_a^2, the additive noise is independent stationary *Gaussian* with variance σ_v^2, and the tap weights are statistically independent of the reference input (*Independence Assumption*).

Then, from (6), one can derive the following difference equations for the mean and the 2nd order moment of the "tap weight error."

$$m^{(n+1)} = m^{(n)} - E[\alpha_c^{(n)}] p^{(n)} \qquad (8)$$

$$\mathbf{K}^{(n+1)} = \mathbf{K}^{(n)} - E[\alpha_c^{(n)}](\mathbf{V}^{(n)} + \mathbf{V}^{(n)T}) + E[(\alpha_c^{(n)})^2] \mathbf{T}^{(n)}, \quad (9)$$

where

$$m^{(n)} = E[\theta^{(n)}], \quad \mathbf{K}^{(n)} = E[\theta^{(n)} \theta^{(n)T}],$$

$$p^{(n)} = E[f(e_n + v_n) g(a^{(n)})], \qquad (10)$$

$$\mathbf{V}^{(n)} = E[f(e_n + v_n) g(a^{(n)}) \theta^{(n)T}], \text{ and} \qquad (11)$$

$$\mathbf{T}^{(n)} = E[f^2(e_n + v_n) g(a^{(n)}) g(a^{(n)})^T]. \qquad (12)$$

For most of the known tap weight adaptation algorithms, it is possible to express

$$E_a[f(e_n + v_n) g(a^{(n)})] = \mathbf{W}^{(n)} \theta^{(n)}, \qquad (13)$$

where $E_a[\cdot]$ denotes expectation with respect to the reference input $a^{(n)}$, and $\mathbf{W}^{(n)}$ is a matrix inherent to the algorithm being used.

Mean Squared Error (MSE) is given by

$$\varepsilon^{(n)} = E[e_n^2]$$
$$= tr(\mathbf{R}_a \mathbf{K}^{(n)}), \qquad (14)$$

where $tr(\cdot)$ denotes trace of a matrix.

Using (13) in (10) and (11),

$$p^{(n)} = \mathbf{W}^{(n)} m^{(n)}, \qquad (15)$$

and

$$\mathbf{V}^{(n)} = \mathbf{W}^{(n)} \mathbf{K}^{(n)} \qquad (16)$$

result.

Now, from (9), (14) and (16), setting the partial derivative of the MSE at time $n+1$ with respect to the step size at time n, $\partial \varepsilon^{(n+1)} / \partial \alpha_c^{(n)}$, equal to 0, one can solve the *theoretically optimum* step size at time n as

$$\alpha_c^{(n)}{}_{opt} = tr(\mathbf{R}_a \mathbf{W}^{(n)} \mathbf{K}^{(n)}) / tr(\mathbf{R}_a \mathbf{T}^{(n)}). \qquad (17)$$

Noting (13) and $E_a[(e_n + v_n) a^{(n)}] = \mathbf{R}_a \theta^{(n)}$ from (3), we find, with some averaging delay in the leaky accumulator,

$$E[q_0^{(n)T} q^{(n)}] \approx tr(\mathbf{R}_a \mathbf{W}^{(n)} \mathbf{K}^{(n)}), \qquad (18)$$

Also from (5), with some delay,

$$E[\tau^{(n)}] \approx tr(\mathbf{R}_a \mathbf{T}^{(n)}). \qquad (19)$$

Then, (2), (18) and (19) yield, *approximately*,

$$E[\alpha_c^{(n)}] \approx \alpha_c^{(n)}{}_{opt} \qquad (20)$$

(20) means that the step size given by (2) "tracks" the theoretically optimum step size via leaky accumulators, thus realizing *quasi-optimal* adaptive step size control, particularly in the early stages of filter convergence.

Referring to (2),

$$E[\alpha_c^{(n)}] \cong E[q_0^{(n)T} q^{(n)}] / E[\tau^{(n)}] \qquad (21)$$

and

$$E[(\alpha_c^{(n)})^2] \cong E[(q_0^{(n)T} q^{(n)})^2] / E[(\tau^{(n)})^2], \qquad (22)$$

where $E[(q_0^{(n)T} q^{(n)})^2]$, $E[q_0^{(n)T} q^{(n)}]$, $E[(\tau^{(n)})^2]$ and $E[\tau^{(n)}]$ are iteratively updated by a set of difference equations which can be derived from (3), (4) and (5) (but not given here).

Next, let us solve theoretical MSE after convergence.

Assume that the filter converges as $n \to \infty$. Then, from (8),

$$p^{(\infty)} = 0$$

which implies

$$m^{(\infty)} = E[q_0^{(\infty)}] = E[q^{(\infty)}] = 0 .$$

And, for a sufficiently small ρ,

$$E[q_0^{(\infty)T} q^{(\infty)}] \cong (\rho/2)\, tr(\mathbf{S}^{(\infty)}), \text{ and} \quad (23)$$

$$E[(q_0^{(\infty)T} q^{(\infty)})^2] \cong (\rho/2)^2 \{ tr^2(\mathbf{S}^{(\infty)}) + tr(\mathbf{T}_0^{(\infty)}\mathbf{T}^{(\infty)}) + tr((\mathbf{S}^{(\infty)})^2)\}, \quad (24)$$

where $\mathbf{T}_0^{(n)} = E[(e_n + \nu_n)^2 a^{(n)} a^{(n)T}]$ and $\mathbf{S}^{(n)} = E[(e_n + \nu_n) f(e_n + \nu_n) a^{(n)} g(a^{(n)})^T]$.

Also for $\rho_\tau << 1$,

$$E[\tau^{(\infty)}] = tr(\mathbf{R}_a \mathbf{T}^{(\infty)}), \quad (25)$$

and

$$E[(\tau^{(\infty)})^2] \cong (E[\tau^{(\infty)}])^2 \quad (26)$$

result. (21), (23) and (25) calculate

$$E[\alpha_c^{(\infty)}] \cong (\rho/2)\, tr(\mathbf{S}^{(\infty)}) / tr(\mathbf{R}_a \mathbf{T}^{(\infty)}), \quad (27)$$

while (22), (24) and (26) yield

$$E[(\alpha_c^{(\infty)})^2] \cong (\rho/2)^2\, tr^2(\mathbf{S}^{(\infty)})\{1 + (tr(\mathbf{T}_0^{(\infty)}\mathbf{T}^{(\infty)}) + tr((\mathbf{S}^{(\infty)})^2)) / tr^2(\mathbf{S}^{(\infty)})\} / tr^2(\mathbf{R}_a \mathbf{T}^{(\infty)}). \quad (28)$$

Next, from (9) and (16), as $n \to \infty$,

$$\mathbf{V}^{(\infty)} + \mathbf{V}^{(\infty)T} = \mathbf{W}^{(\infty)}\mathbf{K}^{(\infty)} + \mathbf{K}^{(\infty)}\mathbf{W}^{(\infty)} = \mu_c^{(\infty)}\mathbf{T}^{(\infty)} \quad (29)$$

holds, where "equivalent" step size $\mu_c^{(\infty)}$ is given by

$$\mu_c^{(\infty)} = E[(\alpha_c^{(\infty)})^2] / E[\alpha_c^{(\infty)}] \cong (\rho/2)\, tr(\mathbf{S}^{(\infty)})\{1 + (tr(\mathbf{T}_0^{(\infty)}\mathbf{T}^{(\infty)}) + tr((\mathbf{S}^{(\infty)})^2)) / tr^2(\mathbf{S}^{(\infty)})\} / tr(\mathbf{R}_a \mathbf{T}^{(\infty)}). \quad (30)$$

(29) and (30) are combined to solve $\mathbf{K}^{(\infty)}$ and hence the residual MSE $\varepsilon^{(\infty)}$. Note that, for a fixed step size ($= \alpha_c$), we simply set $\alpha_c^{(n)} = \alpha_c$ and $\mu_c^{(\infty)} = \alpha_c$.

Following is the result for some examples of adaptive step size tap weight adaptation algorithm. The theoretical value of the residual MSE is given by $\varepsilon^{(\infty)} = \delta\sigma_v^2/(1-\delta)$, where

LMF : $f(x) = x^3, \quad g(a) = a$

$$\delta \cong (\rho/4) N\{1 + (8/3/N)\, tr(\mathbf{R}_a^2)/(\sigma_a^4 N)\} \times \sigma_a^4 N / tr(\mathbf{R}_a^2), \quad (31)$$

SRA : $f(x) = x, \quad g(a) = sgn(a)$

$$\delta \cong (\rho/4) N\{1 + (1/N)(tr(\mathbf{R}_a^2)/(\sigma_a^4 N) + (\pi/2) tr(\mathbf{R}_a \mathbf{T}_a)/(\sigma_a^2 N))\}\, \sigma_a^2 N / tr(\mathbf{R}_a \mathbf{T}_a), \quad (32)$$

and

SSA : $f(x) = sgn(x), \quad g(a) = sgn(a)$

$$\delta \cong (\rho/4) N\{1 + (1/N)(tr(\mathbf{R}_a^2)/(\sigma_a^4 N) + (\pi/2)^2 tr(\mathbf{R}_a \mathbf{T}_a)/(\sigma_a^2 N))\}\, \sigma_a^2 N / tr(\mathbf{R}_a \mathbf{T}_a), \quad (33)$$

with $\mathbf{T}_a = (2/\pi) \arcsin(\mathbf{R}_a \sigma_a^2)$.

Note that for the above algorithms the residual MSE $\varepsilon^{(\infty)}$ has a finite "floor" which is proportional to σ_v^2, ρ and N in theory.

IV. EXPERIMENT

Simulations and theoretical calculations are performed for the following three examples with different types of tap weight adaptation algorithm, where filter convergence with the proposed adaptive step size control algorithm is compared to that with a fixed step size.

Example #1 *LMF Algorithm*
- $N = 4$, $h = [.05,\ .994,\ .01,\ -.1]^T$
- *AR1 modelled input* $\sigma_a^2 = 1$,
- regression coefficient $\eta = .5$
- $\sigma_v^2 = 1\ (0\ dB)$
- *fixed step size* : $\alpha_c = 2^{-13}$
- *adaptive step size* : $\rho = \rho_\tau = 2^{-10}$

Example #2 *Signed Regressor Algorithm (SRA)*
- $N = 4$, $h = [.05,\ .994,\ .01,\ -.1]^T$
- *AR1 modelled input* $\sigma_a^2 = 1$,
- regression coefficient $\eta = .75$
- $\sigma_v^2 = .01\ (-20dB)$
- *fixed step size* : $\alpha_c = 2^{-8}$
- *adaptive step size* : $\rho = \rho_\tau = 2^{-8}$

Example #3 *Sign-Sign Algorithm (SSA)*
- $N = 8$
- $h = [.01,\ .811,\ .05,\ -.572, -.1, -.05, -.02, -.01]^T$
- *White & Gaussian input* $\sigma_a^2 = 1$,
- $\sigma_v^2 = .0001\ (-40dB)$
- *fixed step size* : $\alpha_c = 2^{-13}$
- *adaptive step size* : $\rho = \rho_\tau = 2^{-6}$

Figs. 1 through 3 show results of the experiment for *Examples #1* through *#3*, respectively. The simulation results clearly indicate that the proposed adaptive step size control algorithm makes the filter convergence considerably faster, and prove its effectiveness for different algorithms and various values of filter parameters.

It is also observed that the theoretically calculated convergence curves exhibit fairly good agreement with those of simulations. Note that in each of the figures convergence with *theoretically optimum* step size is plotted for comparison.

V. CONCLUSION

A novel adaptive step size control algorithm for adaptive filters has been proposed, in which the step size is approximated to (or *tracks*) the theoretically optimum value, thus realizing *quasi-optimal* control.

The step size at each time instant is calculated as an inner product of two vectors divided by a scalar. Those vectors and scalar are updated via leaky accumulators. Basically, the algorithm can be applied to virtually any type of tap weight adaptation algorithm.

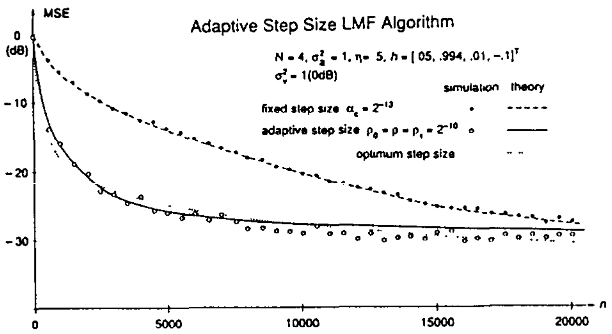

Fig.1 Convergence of adaptive filters
(Example #1; LMF, N = 4).

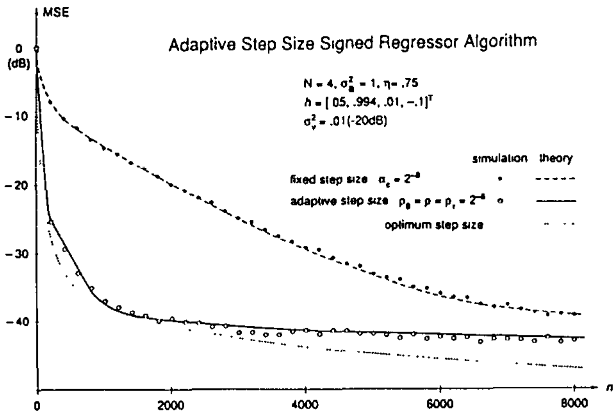

Fig.2 Convergence of adaptive filters
(Example #2; SRA, N = 4).

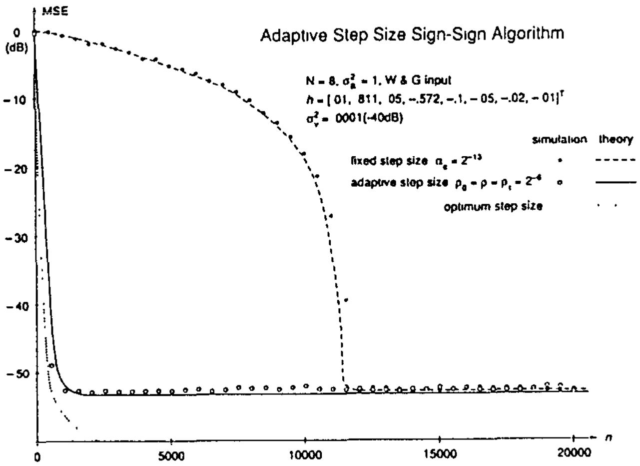

Fig.3 Convergence of adaptive filters
(Example #3; SSA, N = 8).

Analysis has yielded a set of difference equations to describe the filter convergence in the transient phase, and has shown that the residual MSE after convergence, explicitly given in a formula, has a finite floor that depends on the number of taps, leakage factor, additive noise power, *etc.* These theoretical results may contribute to the practical design of adaptive filters.

Experiment with some examples has proven effectiveness of the proposed algorithm in improving the filter convergence rate for a wide range of filter parameters. The results of the experiment also exhibit good agreement between the theoretically calculated convergence and that obtained through simulations.

The proposed algorithm requires *division* to calculate the step size (see (2)). The following "dividerless" approach using another leaky accumulator may be suggested for further study.

$$\alpha_c^{(n+1)} = (1 - \rho_a \tau^{(n)})\,\alpha_c^{(n)} + \rho_a q_0^{(n)T} q^{(n)}.$$

Further study will also be required to simplify the proposed step size control algorithm for a specific tap weight adaptation algorithm.

REFERENCES

[1] D. L. Duttweiler, "Adaptive Filter Performance with Nonlinearities in the Correlation Multipliers," *IEEE Trans. on ASSP*, vol. 30, no. 4, pp. 578-586, August 1982.

[2] W. A. Sethares, "Adaptive Algorithms with Nonlinear Data and Error Functions," *IEEE Trans. on SP*, vol. 40, no. 9, pp. 2199-2206, September 1986.

[3] E. Walach and B. Widrow, "The Least Mean Fourth (LMF) Adaptive Algorithm and Its Family, " *IEEE Trans. on IT*, vol. 30, no. 2, February 1984.

[4] V. J. Mathews and S. H. Cho, "Improved Convergence Analysis of Stochastic Gradient Adaptive Filters Using the Sign Algorithm," *IEEE Trans. on ASSP*, vol. 35, no. 4, pp. 450-454, April 1987.

[5] Eweda Eweda, "Analysis and Design of a Signed Regressor LMS Algorithm for Stationary and Nonstationary Adaptive Filtering with Correlated Gaussian Data," *IEEE Trans. on CS*, vol. 37, no. 11, pp. 1367-1374, November 1990.

[6] S. Koike, "Analysis of the Sign-Sign Algorithm Based on Gaussian Distributed Tap Weights," *Proceedings IEEE ICASSP98*, vol. III, pp. 1673-1676, Seattle, WA, May 1998.

[7] V. J. Mathews and Z. Xie, "A Stochastic Gradient Adaptive Filter with Gradient Adaptive Step Size," *IEEE Trans. on SP*, vol. 41, no. 6, pp. 2075-2087, June 1993.

[8] A. Kanemasa and K. Niwa, "An Adaptive-Step Sign Algorithm for Fast Convergence of a Data Echo Canceller," *IEEE Trans. on Commun.*, vol.35, no.10, pp. 1102-1108, October 1987.

[9] R. H. Kwong and E. W. Johnston, "A Variable Step Size LMS Algorithm," *IEEE Trans. on SP*, vol. 40, no. 7, pp. 1633-1642, July 1992.

[10] T. Aboulnasr and K. Mayyas, "A Robust Variable Step-Size LMS-Type Algorithm: Analysis and Simulations," *IEEE Trans. on SP*, vol. 45, no. 3, pp. 631-639, March 1997.

[11] C. Cowan and C. Rusu, "Novel Cost Function Adaptation Algorithm for Echo Cancellation," *Proceedings IEEE ICASSP98*, vol. III, pp. 1501-1504, Seattle, WA, May 1998