ADAPTIVE FORGETTING FACTOR RECURSIVE LEAST SQUARES ADAPTIVE THRESHOLD NONLINEAR ALGORITHM (AFF-RLS-ATNA) FOR IDENTIFICATION OF NONSTATIONARY SYSTEMS

Shin' ichi Koike, *Member*

NEC Corporation

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

Phone: +81 3 3798 2574 Fax: +81 3 3798 2575 E-mail: koike@bk.jp.nec.com

ABSTRACT

The recursive least squares (RLS) adaptive algorithm is combined with the "adaptive threshold nonlinear algorithm" (ATNA) proposed by the author, to derive RLS-ATNA, resulting in improvement of the convergence rate of the ATNA that offers robust adaptive filters in impulse noise environments. For application of the RLS-ATNA to identification of random-walk modeled nonstationary systems, an adaptive forgetting factor (AFF) control algorithm is proposed that further improves the tracking performance in the steady state. Through analysis and experiments, the effectiveness of the AFF-RLS-ATNA is demonstrated. Fairly good agreement between the simulation and the theoretically calculated convergence validates the analysis.

1. INTRODUCTION

Among the many types of tap weight adaptation algorithm for adaptive filters, the LMS algorithm usually exhibits faster convergence than the other LMS-derived algorithms, when required estimation accuracy is to be realized. The RLS algorithm considerably accelerates the convergence of the LMS algorithm for a *colored* reference input [1, Chap 8].

It is known that the LMS algorithm is quite vulnerable to disturbances, *e.g.*, impulse noise. One of the practical solutions is to use the *sign algorithm.* However, the sign algorithm converges much slower than the LMS algorithm as mentioned above. The author proposed an algorithm called "adaptive threshold nonlinear algorithm" (ATNA) to meet these contradicting requirements, namely, fast convergence and robustness against impulse noise [2].

It is expected that the convergence rate of the ATNA may be improved if it is combined with the RLS algorithm. This paper first derives "recursive least squares adaptive threshold nonlinear algorithm" (RLS-ATNA). Weng *et al*. [3] propose and analyze an RLS algorithm with an amplitude-limiting nonlinear function of the error. Zou *et al*. [4] propose a recursive least M-estimate (RLM) algorithm to mitigate the adverse effects due to impulse noise.

When the RLS-ATNA is used to identify *nonstationary* systems, a fixed forgetting factor for the RLS algorithm may not provide the optimum tracking performance. In fact, for a nonstatonarity that is modeled with a random walk, we find that there exists an optimum value of the forgetting factor that minimizes the steady-state error. This paper further proposes and analyzes a gradient-based adaptive forgetting factor (AFF) control algorithm to be combined with the RLS-ATNA. The new AFF control algorithm gives the nearly optimum forgetting factor to make the steady-state error close to its minimum attainable value. The schematic diagram of the adaptive filtering system equipped with the AFF-RLS-ATNA is illustrated in Fig. 1.

2. ADAPTIVE FORGETTING FACTOR RLS ADAPTIVE THRESHOLD NONLINEAR ALGORITHM

2.1 Formulation of the RLS-ATNA

In this subsection, we formulate "recursive least squares adaptive threshold nonlinear algorithm" (RLS-ATNA) by combining the RLS algorithm with the ATNA proposed by the author [2].

The update equations for the RLS-ATNA are given by the following.

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \alpha_c\, f[e(n)\,;A(n), m]\,\mathbf{g}(n), \quad (1)$$

$$\mathbf{g}(n) = \mathbf{P}(n)\,\mathbf{a}(n) / [\,\lambda + \mathbf{a}^T(n)\,\mathbf{P}(n)\,\mathbf{a}(n)], \quad (2)$$

and

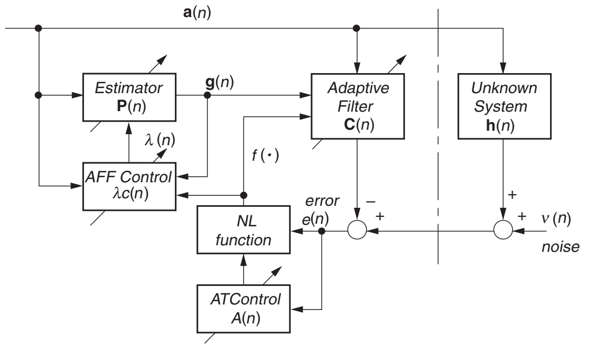

Fig. 1 Schematic diagram for the AFF-RLS-ATNA.

$$\mathbf{P}(n+1) = (1/\lambda)[\,\mathbf{P}(n) - \mathbf{g}(n)\,\mathbf{a}^T(n)\,\mathbf{P}(n)\,], \tag{3}$$

where n is the time instant, $\mathbf{c}(n)$ is the tap weight vector (N taps), $\mathbf{a}(n)$ is the filter reference input vector (length N), $e(n)$ is the error signal, $A(n)$ is the threshold parameter, $\mathbf{g}(n)$ is the Kalman gain, $\mathbf{P}(n)$ is the estimator for the inverse covariance matrix of the reference input, λ is the forgetting factor, and $(\cdot)^T$ denotes the *transpose* of a vector or a matrix. The step size α_c is being introduced for generality, though it is usually set to 1. The nonlinear function $f(e\,;A,m)$ is specifically defined as follows.

$$f(e\,;A,m) = e/[\,1+(|e|/A)^m\,],$$

where $A > 0$ and $m = 2, 4, 8, 16, \ldots$. For $m \to \infty$, the function becomes "truncated linear"[2].

The threshold $A(n)$ is updated as follows, using a leaky accumulator having the absolute value of the error as its input.

$$A(n+1) = (1-\rho_A)A(n) + \rho_A M_A\,|\,e(n)\,|, \tag{4}$$

where ρ_A is the leakage factor and M_A is the multiplier.

2.2 Proposed Adaptive Forgetting Factor Control Algorithm

The proposed gradient adaptive control algorithm for the forgetting factor in identification of nonstationary systems is given by the following equations [5].

$$\lambda(n) = 1 - \lambda c(n) \tag{5}$$

and

$$\lambda c(n+1) = \lambda c(n) + \rho_G f[e(n+1)\,;A(n+1),m] \times f[e(n)\,;A(n),m]\,\mathbf{a}^T(n+1)\,\mathbf{g}(n)/\lambda c(n), \tag{6}$$

where $\lambda c(n)$ is the *complementary* forgetting factor at time n and ρ_G is the coefficient for adaptation. The division by $\lambda c(n)$ on the right-hand-side of (6) is for a scaling purpose. Since the amplitude-limiting nonlinear function $f(e\,;A,m)$ is involved, the control of the forgetting factor is also expected to be robust against impulse noise.

2.3 Optimum Forgetting Factor and Minimum Steady-State Excess Mean Square Error

Let us assume that the filter is of an FIR type and is applied to identification of unknown nonstationary systems. The nonstationarity is modeled with an independent and white "random walk" as

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mathbf{w}(n), \tag{7}$$

where $\mathbf{h}(n)$ is the unknown system response vector (length N), $\mathbf{w}(n)$ is the random-walk vector with $E[\mathbf{w}(n)\mathbf{w}^T(n)] = \sigma_w^2\,\mathbf{I}$, σ_w^2 is the variance of each element of the vector $\mathbf{w}(n)$, and $\mathbf{I}$ is the identity matrix.

Although the details are not given due to space limitation, expectation analysis of (1) through (4) for a *fixed forgetting factor* yields a set of difference equations for calculating the transient behavior of the filter convergence. Then, as $n \to \infty$, we can solve the steady-state excess mean square error (EMSE) $\varepsilon(\infty)$ for $N >> 1$ as given by

$$\varepsilon(\infty) \simeq (\alpha_c \lambda c N/2)(H_{2\infty}/H_\infty)\,\sigma_e^2(\infty) + \sigma_a^2\,\sigma_w^2 N/(2\,\alpha_c \lambda c H_\infty), \tag{8}$$

where $H_\infty \triangleq H[(2/\pi)^{1/2}M_A\,;m]$ and $H_{2\infty} \triangleq H_2[(2/\pi)^{1/2}M_A\,;m]$, σ_a^2 is the variance of the reference input, and

$$\sigma_e^2(\infty) \triangleq \varepsilon(\infty) + \sigma_\nu^2, \tag{9}$$

with σ_ν^2 being the additive noise variance. The functions $H(r\,;m)$ and $H_2(r\,;m)$ are defined by

$$H(r\,;m) \triangleq \int_{-\infty}^{\infty} t^2/[\,1+(|t|/r)^m\,] \times p_N(t)\,dt$$

and

$$H_2(r\,;m) \triangleq \int_{-\infty}^{\infty} t^2/[\,1+(|t|/r)^m\,]^2 \times p_N(t)\,dt,$$

respectively, where $p_N(x) \triangleq \exp(-x^2/2)/(2\pi)^{1/2}$ is the *Normal Density Function*.

As is anticipated, there exists an optimum forgetting factor that minimizes the steady-state EMSE in identification of the random-walk modeled nonstationary systems.

Differentiating (8) with respect to λc, we find the optimum complementary forgetting factor

$$\lambda c_{opt} \simeq \sigma_a \sigma_w / [\,\alpha_c H_{2\infty}^{1/2}\,\sigma_{e\,\min}(\infty)\,] \tag{10}$$

and the minimum steady-state EMSE

$$\varepsilon_{\min}(\infty) \simeq \sigma_a \sigma_w\, N (H_{2\infty}^{1/2}/H_\infty)\,\sigma_{e\,\min}(\infty) \tag{11}$$

with

$$\sigma_{e\,\min}^2(\infty) = \varepsilon_{\min}(\infty) + \sigma_\nu^2. \tag{12}$$

2.4 Analysis of the AFF-RLS-ATNA

We assume that the variance of the forgetting factor is sufficiently small for $\rho_G << 1$. Then, from (6), we can derive a difference equation for the expectation of $\lambda c(n)$, from which we find, as $n \to \infty$, the steady-state complementary forgetting factor as

$$E[\,\lambda c(\infty)\,] \simeq \xi_a^{1/2}\,\sigma_a \sigma_w / [\,\alpha_c H_{2\infty}^{1/2}\,\sigma_{e\,AFF}(\infty)\,], \tag{13}$$

and the steady-state EMSE

$$\varepsilon_{AFF}(\infty) \simeq \gamma_a\,\sigma_a \sigma_w\, N (H_{2\infty}^{1/2}/H_\infty)\,\sigma_{e\,AFF}(\infty) \tag{14}$$

with

$$\sigma_{e\,AFF}^2(\infty) = \varepsilon_{AFF}(\infty) + \sigma_\nu^2, \tag{15}$$

where

$$\gamma_a \triangleq (\,\xi_a^{1/2} + 1/\xi_a^{1/2})/2, \tag{16}$$

and

$$\xi_a = (1+2\eta^2)/(1+2\eta^2/N) \tag{17}$$

for an AR(1) (first-order autoregressive) reference input with regression coefficient η ($0 \le \eta < 1$). Note here that $1 \le \xi_a < 3$ and $1 \le \gamma_a < 2/3^{1/2} = 1.155$.

Comparing (14) with (11) and noting $\gamma_a < 1.155$, we find that the steady-state EMSE is close to the optimum value when the proposed adaptive forgetting factor control algorithm is combined with the RLS-ATNA in identification of nonstationary systems.

3. EXPERIMENTS

In this section, experiments with simulations and theoretical calculations are carried out to examine the performance of the RLS-ATNA and to verify the analysis developed above.

3.1 Convergence Behavior of the RLS-ATNA

Simulation and theoretically calculated results of the filter convergence are compared for the example below. The convergence behavior of the RLS-ATNA is also compared with that of the conventional RLS algorithm. The simulation result in the experiments is an ensemble average of the squared excess error over 1000 independent runs of the filter convergence. In the example, the filter reference input is a colored Gaussian process and is modeled as an AR(1) process with regression coefficient η ($0 \le \eta < 1$).

Example $N = 32$, $\sigma_a^2 = 1$ (0 dB), $\eta = .5$,
$\sigma_\nu^2 = .01$ (−20 dB), $\alpha_c = 1$
$m = 16$, $M_A = 2.5$, $\rho_A = 2^{-8}$

Case 1 $\sigma_w^2 = 0$ (stationary system)
fixed forgetting factor (FFF) $\lambda c = 2^{-11}$
$\rightarrow \varepsilon(\infty) \simeq -41.4$ dB

Case 2 $\sigma_w^2 = 5 \times 10^{-6}$ (nonstationary system)
(a) FFF $\lambda c = 2^{-11}$ $\rightarrow \varepsilon(\infty) \simeq -6.5$ dB
(b) optimum forgetting factor $\lambda c_{opt} = .0193$
$\rightarrow \varepsilon_{min}(\infty) \simeq -19.2$ dB

Case 3 $\sigma_w^2 = 5 \times 10^{-6}$ (nonstationary system)
adaptive forgetting factor (AFF)
$\lambda c(0) = 2^{-11}$, $\rho_G = 2^{-12}$
AFF switched ON at $n = 10000$
$\rightarrow \varepsilon(\infty) \simeq -19.1$ dB
$\lambda c(\infty) \simeq .022$

The results of the experiment for Case 1 are shown in Fig. 2, where the RLS-ATNA converges as fast as the conventional RLS algorithm. Fig. 3 depicts the results for Case 2 in which the system to be identified is nonstationary. The steady-state EMSE for the optimum forgetting factor attains the minimum value that is much lower than that for the small forgetting factor as used in Case 1.

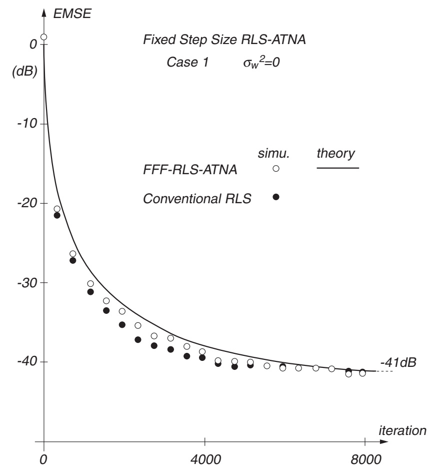

Fig. 2 Adaptive filter convergence; FFF-RLS-ATNA versus conventional RLS (Case 1).
$N = 32$, $\sigma_a^2 = 1$ (0 dB), $\eta = .5$, $\sigma_\nu^2 = .01$ (−20 dB)
$m = 16$, $M_A = 2.5$, $\rho_A = 2^{-8}$, $\alpha_c = 1$, $\sigma_w^2 = 0$,
$\lambda c = 2^{-11}$

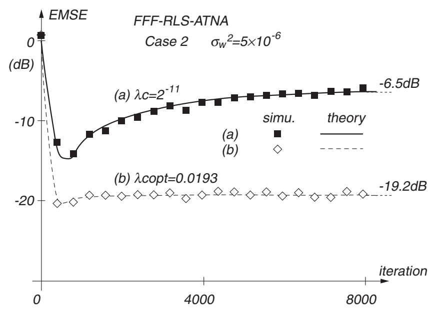

Fig. 3 Adaptive filter convergence; FFF-RLS-ATNA (Case 2).
$N = 32$, $\sigma_a^2 = 1$ (0 dB), $\eta = .5$, $\sigma_\nu^2 = .01$ (−20 dB)
$m = 16$, $M_A = 2.5$, $\rho_A = 2^{-8}$, $\alpha_c = 1$,
$\sigma_w^2 = 5 \times 10^{-6}$
(a) $\lambda c = 2^{-11}$ (b) $\lambda c_{opt} = .0193$

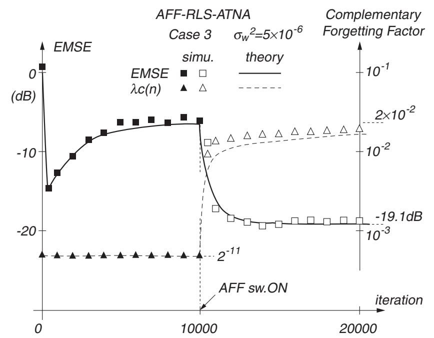

Fig. 4 Adaptive filter convergence; AFF-RLS-ATNA (Case 3).
$N = 32$, $\sigma_a^2 = 1$ (0 dB), $\eta = .5$, $\sigma_\nu^2 = .01$ (−20 dB)
$m = 16$, $M_A = 2.5$, $\rho_A = 2^{-8}$, $\alpha_c = 1$,
$\sigma_w^2 = 5 \times 10^{-6}$
$\lambda c(0) = 2^{-11}$, $\rho_G = 2^{-12}$
AFF control switched ON at $n = 10000$

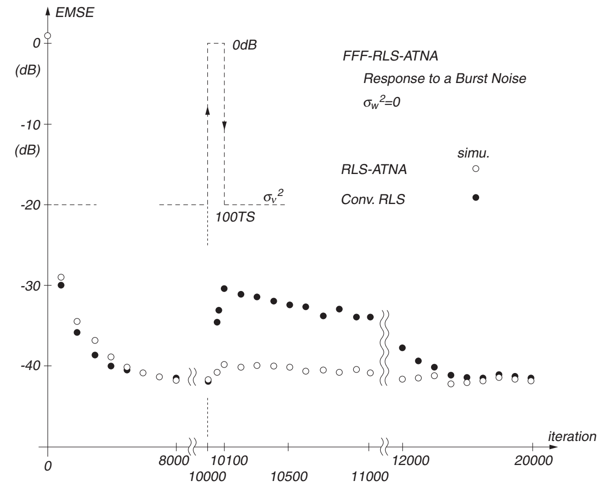

Fig. 5 Response of the filter to a burst noise (+20 dB, 100 time slots); RLS-ATNA versus conventional RLS.
$N = 32$, $\sigma_a^2 = 1$ (0 dB), $\eta = .5$,
$\sigma_\nu^2 = .01$ (−20 dB) → 1 (0 dB) → .01 (−20 dB)
for n =10000 −10100.
$m = 16$, $M_A = 2.5$, $\rho_A = 2^{-8}$, $\alpha_c = 1$, $\sigma_w^2 = 0$,
$\lambda_c = 2^{-11}$

In Fig. 4, the results for Case 3 are shown, where the proposed adaptive forgetting factor control is enabled after the filter has converged with the small initial forgetting factor in identification of nonstationary systems. Clearly, we observe that the AFF is effective in controlling the forgetting factor to make the EMSE close to its minimum.

In the figures above, the theoretical convergence curves exhibit fairly good agreement with those obtained through simulations.

3.2 Response to a Burst Noise

To demonstrate the robustness of the RLS-ATNA against impulse noise, filter response to a burst noise is examined. In this experiment, a burst increase of +20 dB for a short duration (100 time slots) in the noise variance, simulating impulse noise, occurs after the filter has converged.

The parameters of the filter are the same as above, but the noise variance varies as

$$\sigma_\nu^2 = .01\ (-20\ \text{dB}) \rightarrow 1\ (0\ \text{dB}) \rightarrow .01\ (-20\ \text{dB}) \quad \text{for } n = 10000 - 10100.$$

The filter responses to the burst noise for the RLS-ATNA and the conventional RLS algorithm are shown in Fig. 5. The RLS-ATNA sufficiently suppresses the influence of the burst noise, allowing only a few dB increase in the EMSE. However, for the conventional RLS algorithm the increase in the EMSE becomes greater than 10 dB and decays very slowly.

4. CONCLUSION

"Adaptive forgetting factor recursive least squares adaptive threshold nonlinear algorithm" (AFF-RLS-ATNA) has been proposed which is basically the ATNA combined with recursive estimation of the inverse covariance matrix of the filter reference input and with a gradient-based adaptive forgetting factor control algorithm. The proposed AFF-RLS-ATNA exhibits robustness against impulse noise, and offers nearly optimum tracking performance in identification of nonstationary systems.

The transient behavior and the steady-state performance of the AFF-RLS-ATNA have been analyzed to yield a set of difference equations for theoretically calculating the filter convergence.

The results of the experiments for a practical example have shown that the AFF-RLS-ATNA gives the steady-state EMSE close to its minimum attainable value, while preserving the robustness against impulse noise. Fairly good agreement between the simulation and the theoretically calculated convergence has proven the validity of the analysis.

In the paper, the number of the tap weights is assumed sufficiently large. Analysis for a small N is to be further developed.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*. New Jersey: Prentice-Hall, 1996.

[2] S.Koike, "Adaptive threshold nonlinear algorithm for adaptive filters with robustness against impulse noise," *IEEE Trans. Signal Processing*, vol.45, pp.2391-2395, Sept. 1997.

[3] J. F. Weng and S. H. Leung, "Adaptive nonlinear RLS algorithm for robust filtering in impulse noise," in *Proc. IEEE ISCAS 97*, vol.4, June 1997, pp.2337-2340.

[4] Y. Zou, S. C. Chan, and T. S. Ng, "A recursive least M-estimate (RLM) adaptive filter for robust filtering in impulse noise," *IEEE Signal Processing Letters*, vol.7, pp.324-326 , Nov. 2000.

[5] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Processing*, viol.41, pp.2075-2087, June 1993.