International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:10, No:1, 2016

Variable Regularization Parameter Normalized Least Mean Square Adaptive Filter

Young-Seok Choi

Abstract—We present a normalized LMS (NLMS) algorithm with robust regularization. Unlike conventional NLMS with the fixed regularization parameter, the proposed approach dynamically updates the regularization parameter. By exploiting a gradient descent direction, we derive a computationally efficient and robust update scheme for the regularization parameter. In simulation, we demonstrate the proposed algorithm outperforms conventional NLMS algorithms in terms of convergence rate and misadjustment error.

Keywords—Regularization, normalized LMS, system identification, robustness.

I. INTRODUCTION

THE normalized least mean square (NLMS) is one of the widely used adaptive algorithms due to its simplicity. Originally, NLMS incorporates the step-size divided by the squared norm of the input regressor. But in practical applications ϵ-NLMS is widely used where the step-size is divided by the squared norm with a small positive constant ϵ added [1]–[5]. This is for avoiding the situation when the norm of the input regressor is close to zero. The small positive constant is so-called the *regularization parameter*. Besides the numerical stability, it is well-known that the *regularization parameter* plays a critical role in performance of ϵ-NLMS [6]–[8]. The trade-off between convergence rate and misadjustment error exists in ϵ-NLMS with a fixed *regularization parameter*.

To overcome the trade-off in ϵ-NLMS, recently the generalized normalized gradient descent (GNGD) algorithm has been proposed [6]. GNGD uses an adaptive *regularization parameter*, i.e., the *regularization parameter* is suitably updated at every iteration. The updating scheme for the *regularization parameter* works well in initial transient phase but shows limited performance as iteration goes. This is because the adaptation of the *regularization parameter* is not robust to the status of adaptive filters.

In this paper, we propose a NLMS with the robust *regularization parameter*. The proposed approach dynamically updates the *regularization parameter* so that adaptive filters not only converge faster but also have lower misadjustment error. This is achieved by exploiting a state of gradient descent direction at every iteration. To make the *regularization parameter* robust, we introduce a normalized gradient for updating the *regularization parameter*. Also, we show that the proposed method is computationally efficient. In simulation we demonstrate the proposed method outperforms conventional NLMS algorithms in terms of convergence speed and misadjustment error for both linear and nonlinear input signals.

Y.-S. Choi is with the Department of Electronic Engineering, Gangneung-Wonju National University, 7 Jukheon-gil, Gangneung, 210-702 Republic of Korea (phone: +82-33-640-2429; fax: +82-33-646-0740; e-mail: yschoi@gwnu.ac.kr).

II. THE PROPOSED NLMS ALGORITHM

Consider data $d(i)$ that arise from the model

$$d(i) = \mathbf{u}_i \mathbf{w}^\circ + v(i) \tag{1}$$

where $\mathbf{w}^\circ$ is an unknown column vector that we wish to estimate, $v(i)$ is a measurement noise and $\mathbf{u}_i$ denotes a $1 \times M$ input vector,

$$\mathbf{u}_i = [u(i) \;\; u(i-1) \;\; \cdots u(i-M+1)] \tag{2}$$

A. Conventional Regularization for NLMS

Let $\mathbf{w}_i$ be an estimate for $\mathbf{w}^\circ$ at iteration i. Then the *a priori output* estimation error is given by

$$e(i) = d(i) - \mathbf{u}_i \mathbf{w}_{i-1} \tag{3}$$

The ϵ-NLMS algorithm computes $\mathbf{w}_i$ via

$$\mathbf{w}_i \;=\; \mathbf{w}_{i-1} + \frac{\mu}{\|\mathbf{u}_i\|^2 + \epsilon} e(i)\, \mathbf{u}_i{}^* \tag{4}$$

where μ is the step-size ,ϵ is the *regularization parameter* and $(\cdot)^*$ denotes Hermitian transpose. Besides the numerical stabilization of the denominator in (4), the *regularization parameter* ϵ plays a critical role in convergence performance of NLMS. If we choose a large *regularization parameter*, the effective step-size becomes small and thus ϵ-NLMS results in small misadjustment error but shows slow convergence. On the other hand, when a small *regularization parameter* is used, the effective step-size becomes relatively large and thus NLMS converges fast but results in large misadjustment errors. From this point of view we may expect performance improvement by using a variable $\epsilon(i)$ instead of a fixed ϵ.

B. NLMS with Robust Regularization

To achieve this goal, we propose a NLMS with robust regularization which continuously updates the *regularization parameter* so that $J(i) = \frac{1}{2} e^2(i)$ is minimized. In the proposed method, we normalize the gradient, $\nabla_\epsilon J(i)$, by its norm. The proposed parameter $\epsilon(i)$ is recursively updated by

$$\epsilon(i) = \epsilon(i-1) - \rho \frac{\nabla_\epsilon J(i)}{\|\nabla_\epsilon J(i)\|} \tag{5}$$

where

$$\nabla_\epsilon J(i) = \mu \frac{e(i)\, e(i-1)\, \mathbf{u}_i\, \mathbf{u}_{i-1}{}^*}{(\|\mathbf{u}_{i-1}\|^2 + \epsilon(i-1))^2} \tag{6}$$

By introducing the normalized gradient, the proposed parameter $\epsilon(i)$ is less sensitive and robust to $e(i)$. Also, the normalized version of gradient $\nabla_\epsilon J(i)$ with a fixed ρ always make the same stride, neglecting how steep the slope of $J(i)$ is. This properties make the *regularization parameter* $\epsilon(i)$ updated effectively when $\nabla_\epsilon J(i)$ is very small.

In (6), since μ and the denominator are always positive, (5) is rewritten as

$$\epsilon(i) = \epsilon(i-1) - \rho \operatorname{sgn}[e(i)e(i-1)\mathbf{u}_i \mathbf{u}_{i-1}{}^*] \tag{7}$$

where $\operatorname{sgn}(x) = \frac{x}{\|x\|} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$.

Since $\nabla_w J(i) = -e(i)\mathbf{u}_i$ and $\nabla_w J(i-1) = -e(i-1)\mathbf{u}_{i-1}$, the updated equation (7) can be expressed as

$$\epsilon(i) = \epsilon(i-1) - \rho \operatorname{sgn}[\nabla_w J(i) \nabla_w^* J(i-1)] \tag{8}$$

Note that the inner product of two gradient vectors is

$$\begin{aligned} &\nabla_w J(i) \nabla_w^* J(i-1) = \\ &\|\nabla_w J(i)\| \cdot \|\nabla_w J(i-1)\| \cdot \cos\theta \end{aligned} \tag{9}$$

where θ is the acute angle between two vectors. Therefore it holds

$$\operatorname{sgn}[\nabla_w J(i) \nabla_w^* J(i-1)] = \operatorname{sgn}(\cos\theta) \tag{10}$$

since $\|\nabla_w J(i)\| > 0$. Using (10), (8) is compactly rewritten as

$$\epsilon(i) = \epsilon(i-1) - \rho \operatorname{sgn}(\cos\theta) \tag{11}$$

From (11), we know that when the angle between two gradient vectors, $\nabla_w J(i-1)$ and $\nabla_w J(i)$ is from 0° to 90°, $\cos\theta$ becomes positive and thus $\epsilon(i)$ decreases. On the other hand, when the two vectors are diverse such that the acute angle is from 90° to 180°, $\cos\theta$ is negative and thus $\epsilon(i)$ increases. Thus we know that the proposed parameter $\epsilon(i)$ is updated according to a status of gradient descent direction at present and past iterations.

The proposed NLMS algorithm incorporating the robust regularization (7) is summarized as

$$\begin{aligned} \epsilon(i) &= \epsilon(i-1) - \rho \operatorname{sgn}[e(i)e(i-1)\mathbf{u}_i \mathbf{u}_{i-1}{}^*] \\ \mathbf{w}_i &= \mathbf{w}_{i-1} + \tfrac{\mu}{\|\mathbf{u}_i\|^2 + \max(\epsilon(i), \epsilon_{\min})} \, e(i) \, \mathbf{u}_i{}^* \end{aligned} \tag{12}$$

where ϵ_{min} is a minimum allowable value of $\epsilon(i)$.

The preceding algorithm, GNGD [6], has been proposed for the same goal of the proposed method. GNGD updates the *regularization parameter* as

$$\epsilon(i) = \epsilon(i-1) - \rho\,(\nabla_\epsilon J(i)) \tag{13}$$

where ρ is a small positive constant.

From (4), (6) and (13) we know that $\Delta\epsilon(i) = \epsilon(i) - \epsilon(i-1)$ is proportional to the square order of $e(i)$ while $\Delta\mathbf{w}_i = \mathbf{w}_i - \mathbf{w}_{i-1}$ is proportional to $e(i)$. So a small $e(i)$ after the initial adaptation results in very small $\Delta\epsilon(i)$ and correspondingly $\epsilon(i)$ undergoes small variation. This is a reason why GNGD shows limited performance as iteration goes although it works well during initial transient period. Comparing the proposed method with GNGD, we know that the merits of GNGD are inherited and the shortcomings of it are overcome in the proposed method.

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY

Algorithms	multiplications	divisions
Proposed	$2M+7$	1
GNGD	$2M+10$	2
ϵ-NLMS	$2M+3$	1

C. $\epsilon_{\min}$ *and Stability*

To guarantee the stability of the proposed algorithm, we need to set ϵ_{min}. Let us define the *a posteriori output* estimation error

$$r(i) = d(i) - \mathbf{u}_i \mathbf{w}_i \tag{14}$$

Then from (3) the following holds

$$r(i) = \left(1 - \frac{\mu\|\mathbf{u}_i\|^2}{\epsilon + \|\mathbf{u}_i\|^2}\right) e(i) \tag{15}$$

Since the magnitude of the *a posteriori output* error will not exceed that of the *a priori output* error, i.e,. $|r(i)| < |e(i)|$, $\epsilon_{\min}$ must satisfy

$$\left|1 - \frac{\mu\|\mathbf{u}_i\|^2}{\epsilon_{\min} + \|\mathbf{u}_i\|^2}\right| < 1 \tag{16}$$

Therefore we get

$$\epsilon_{\min} > \|\mathbf{u}_i\|^2 \left(\frac{\mu}{2} - 1\right) \tag{17}$$

The inequality (17) can be written in terms of μ as

$$0 < \mu < \frac{2(\epsilon_{\min} + \|\mathbf{u}_i\|^2)}{\|\mathbf{u}_i\|^2} \tag{18}$$

When $\epsilon_{\min} = 0$, the stability bound in (18) reduces to $0 < \mu < 2$ which is that of conventional NLMS.

D. *Computational Complexity*

To show that the proposed method is computationally efficient, Table I compares computational cost in terms of multiplication and division for various algorithms. As can be seen in Table I, the proposed method has a computational complexity less than GNGD and similar to that of ϵ-NLMS. Computational savings are obtained by using the normalized gradient in the proposed method.

III. SIMULATION RESULTS

We illustrate the performance of the proposed NLMS algorithm by carrying out computer simulations in the channel estimation scenario. The unknown channel $H(z)$ is represented by FIR structure with 10 taps. The adaptive filter and the unknown channel are assumed to have the same order. The performance of the proposed algorithm is compared with that of GNGD and ϵ-NLMS. For simulations, a linear and a nonlinear signal are used as a input signal. The linear input signal is obtained by filtering a white, zero-mean, Gaussian random sequence $x(i)$ through an AR filter

$$u(i) = 0.9u(i-1) + x(i) \tag{19}$$

International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:10, No:1, 2016

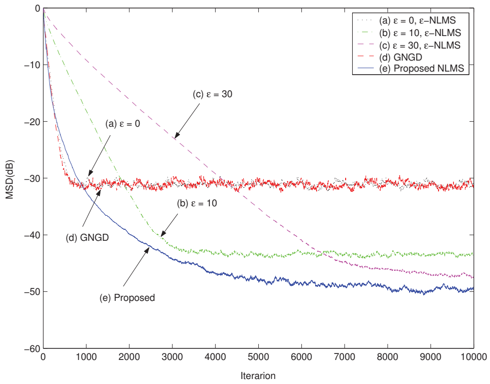

Fig. 1 Comparison of ϵ-NLMS, GNGD and the proposed method for the linear input signal (19)

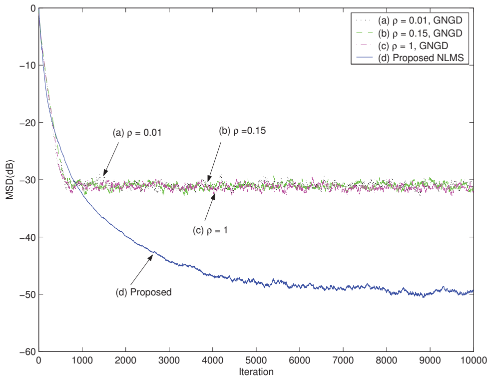

Fig. 2 Comparison of GNGD and the proposed method for the linear input signal (19)

and the nonlinear signal is formed by

$$u(i) = \frac{u(i-1)}{1+u(i-1)^2} + x(i-1)^3 \tag{20}$$

The sinal-to-nose ratio(SNR) is calculated by

$$\mathrm{SNR} = 10\log_{10}(E[y(i)^2]/E[v(i)^2]) \tag{21}$$

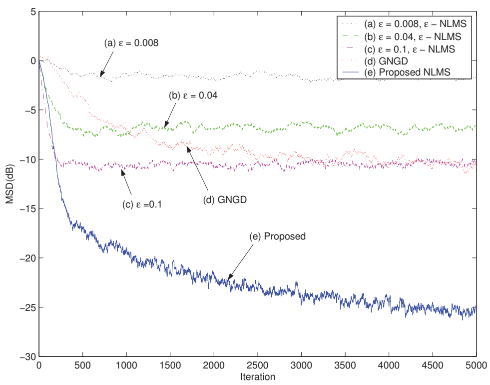

Fig. 3 Comparison of ϵ-NLMS, GNGD and the proposed method for the nonlinear input signal (20)

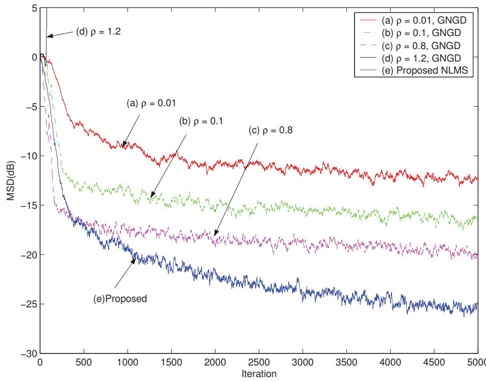

Fig. 4 Comparison of GNGD and the proposed method for the nonlinear input signal (20)

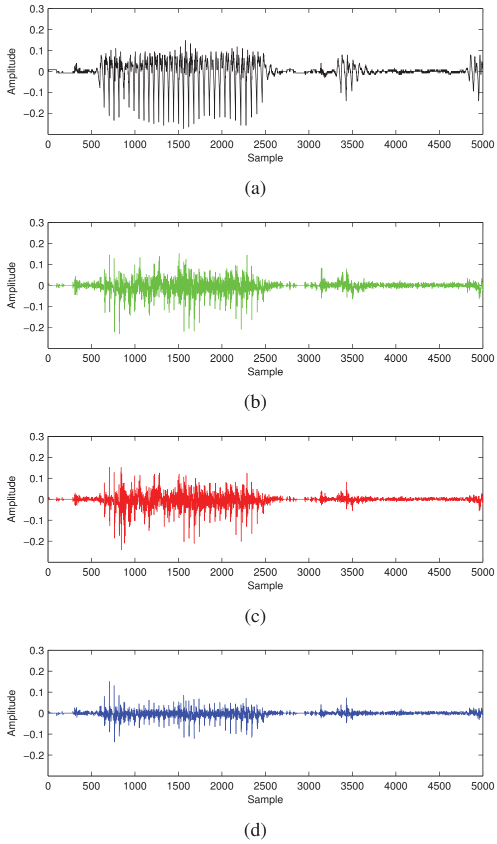

Fig. 5 (a) Reference speech (b) Prediction error of the ϵ-NLMS (ϵ = 0.0001) (c) Prediction error of the GNGD (d) Prediction error of the proposed method

where $y(i) = \mathbf{u}_i\mathbf{w}^o$.

The measurement noise $v(i)$ is added to $y(i)$ such that SNR $=$ 30dB. The mean square deviation (MSD), $E\|\mathbf{w}^o - \mathbf{w}_i\|^2$, is taken and averaged over 50 independent trials. The initial value $\epsilon(0)$ and $\epsilon_{\min}$ is set to zero for all simulations. We use the input signals (19) and (20) for Figs. 1, 2 and Figs. 3, 4, respectively. Fig. 1 compares the MSD curves of ϵ-NLMS, GNGD and the proposed method for $\mu = 0.5$. Dashed lines indicate learning curves of ϵ-NLMS with fixed ϵ where we choose $\epsilon = 0, 10$ and 30. For GNGD, ρ is set to 0.15 and for the proposed method ρ is set to 0.075. As can be seen, the proposed method shows faster convergence rate and lower misadjustment error than ϵ-NLMS and GNGD. Fig. 2 illustrates the performance comparison between the proposed method and GNGD for various ρ. The proposed method outperforms GNGD for various ρ. Fig. 3 shows a performance comparison of ϵ-NLMS, GNGD and the proposed method for the nonlinear input signal (20). We choose $\rho = 0.005$ and $\mu = 1.99$ for GNGD and $\rho = 0.01$ and $\mu = 1.99$ for the proposed method. Fig. 4 compares the performance of GNGD with $\rho = 0.01, 0.1$, 0.8 and 1.2 and that of the proposed method. As in the case of the linear input signal, the proposed method shows better performance than conventional NLMS algorithms even for the nonlinear input signal.

In this experiment, we consider a prediction of a speech as a nonstationary signal. The order of the adaptive filter is 10 and a one-step predictor is used. Fig. 5(a) shows the speech signal that is used as the reference signal. The speech prepared in the experiment is sampled by 8KHz. The initial value $\epsilon(0)$ and $\epsilon_{\min}$ are set to zero. The step-size μ is set to 1.5. In the ϵ-NLMS, $\epsilon = 0.0001$ is chosen. We use $\rho = 0.001$ for the both the GNGD and the proposed method. Figs. 5(b)–5(d) show prediction error $e(i)$ of the ϵ-NLMS, the GNGD, and the proposed method, respectively. We see that the prediction error of the proposed method is smaller than those of the ϵ-NLMS and the GNGD. Consequently, the proposed method outperforms the ϵ-NLMS and the GNGD for a nonstationary signal such as speech.

IV. Conclusions

We have proposed a novel NLMS algorithm with robust regularization. The performance improvement of NLMS has been achieved by developing an efficient and robust adaptation scheme for the variable *regularization parameter* $\epsilon(i)$ using the normalized gradient. In simulation we have illustrated the proposed method outperforms conventional NLMS algorithm in terms of convergence rate and misadjustment error with an efficient computation.

References

[1] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, 2002.
[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Englewood Cliffs, NJ: Prentice Hall, 2003.
[3] A. Benveniste, M. Metivier and P. Priouret, *Adapive Algorithms and Stochstic Approximation*, New York: Springer-Verlag, 1990
[4] J. Homer, “Detection guided NLMS estimation of sparsely parametrized channels,” *IEEE Trans. Circuits Syst. II*, vol. 47, No. 12, pp. 1437–1442, Dec. 2000.
[5] S. J. Park, C. G. Cho, C. Lee and D. H. Youn, “Integrated echo and noise canceler for hands-free applications,” *IEEE Trans. Circuits Syst. II*, vol. 49, No. 3, pp. 188–195, March 2002.
[6] D. P. Mandic, “A generalized normalized gradient descent algorithm,” *IEEE Signal Processing Lett.*, vol. 11, No. 3, pp. 115–118, Feb. 2004.
[7] A. I. Hanna, I. Yates and D. P. Mandic, “Analysis of the class of complex-valued error adaptive normalized nonlinear gradient descent algorithms,” in *Proc. IEEE Int. Conf. on Accoustics, Speech, and Signal Processing, ICASSP’03*, Hong Kong, Apr. 2003, pp. II-705–708.
[8] V. Myllylä and G. Schmidt, “Pseudo-optimal regulariztion for affine projection algorithms,” in *Proc. IEEE Int. Conf. on Accoustics, Speech, and Signal Processing, ICASSP’02*, Orlando, Florida, May 2002, pp. 1917–1920.