CONVERGENCE ACCELERATION OF LMS ALGORITHM USING SUCCESSIVE DATA ORTHOGONALIZATION

Hyun-Chool Shin and Woo-Jin Song

Division of Electronics and Computer Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Kyungbuk, 790-784, Republic of Korea
e-mail:wjsong@postech.ac.kr

ABSTRACT

It is well-known that the convergence rate gets worse when an input signal to an adaptive filter is correlated. In this paper we propose a new adaptive filtering algorithm that makes the convergence rate highly improved even for highly correlated input signals. By introducing an orthogonal constraint between successive input signal vectors,we overcome the slow convergence problem caused by the correlated input signal. Simulation results show that the proposed algorithm yields highly improved convergence speed and excellent tracking capability under both time-invariant and time-varying environments, while keeping both computation and implementation simple.

1. INTRODUCTION

Adaptive filtering has drawn much attention in recent years due to the potential for estimating and tracking a changing environment. The least-mean square (LMS) algorithm is certainly one of the most referenced adaptive filtering algorithms due to its simplicity [1]. However, the correlated nature of an input signal highly degrades the convergence speed of LMS adaptive filters. In recent years, considerable efforts have been spent in making the convergence rate of LMS improved. Although the recursive least squares (RLS) method shows superior convergence speed over the LMS method, it is computationlly intensive and requires higher storage over the LMS method. Fast RLS algorithms have been developed but they have a tendency to become numerically unstable [2]. These inherent limitations of the RLS method make the use of the method limited.

As a result, many variants of the LMS method have been devised through simple modification or additional filtering to improve the convergence rate [3]. Proakis provided an variant of the LMS method where gradient vectors are linearly filtered [4]. As another attempt for fast convergence, a conjugate gradient (CG) method has been developed [5]–[6]. Although the CG method has convergence properties superior to those of LMS, the CG algorithm still requires much higher computaional cost than the LMS method. Recently the orthogoanl gradient adaptive (OGA) algorithm which filters a gradient vector so that the current gradient vector is orthogonal to the previous one was proposed [7]. Although the OGA algorithm is computationally simple as much as the LMS method, the convergence speed is much slower than the CG algorithm.

In this paper we propose a new adaptive filtering algorithm based on successive input data orthogonaliation. The proposed method shows the fast convegence speed comparable with the CG algorithm while keeping computationally simple as much as the OGA algorithm. The proposed algorithm is motivated by the fact that the orthogonality between the current input vector and the previous one is an important factor for fast convergence. To give better insights we describe this fact from a geometric perspective. The Gram-Schmidt orthogonalization procedure is used to make the orthogonal relation between input vectors. The simulation results show that the proposed algorithm yields highly improved convergence speed and tracking capability for both time-invariant and time-variant enviornments.

Throughout the paper, the following notations are adopted

$\mathbf{x}^T$	Transpose of $\mathbf{x}$
$\|\|\mathbf{x}\|\|$	Euclidean norm of $\mathbf{x}$.

2. GEOMETRIC INTERPRETATION OF LMS

Let a discrete-time signal $x(n)$ be the input to an adaptive transversal filter and $d(n)$ be the desired output. Then the estimation error between the desired signal and the adaptive filter output is given by

$$e(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}(n), \quad (1)$$

where $\mathbf{x}^T(n) = [x(n)\ x(n-1)\ \cdots\ x(n-K+1)]$ is an input vector and $\mathbf{w}^T(n) = [w_0(n)\ w_1(n)\ \cdots\ w_{K-1}(n)]$ is a tap-weight vector. The well-known LMS equation for

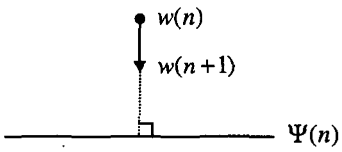

Fig. 1. Geometric description of LMS update

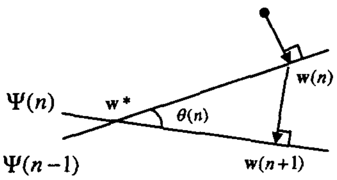

Fig. 2. Relation between convergence rate and acute angle

updating the weight vector is given by

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n), \tag{2}$$

where μ is a positive step-size. The LMS algorithm updates $\mathbf{w}(n)$ so that $e^2(n)$ is minimized.

To see the behavior of the LMS method from a geometric perspective we define a hyperplane which consists of all vectors $\mathbf{w}$ such that $e(n) = 0$, i.e.,

$$\Psi(n) = \{\mathbf{w}|\mathbf{x}^T(n)\mathbf{w} = d(n)\}.$$

Then the LMS algorithm moves $\mathbf{w}(n)$ toward the hyperplane $\Psi(n)$ since $e^2(n)$ is smaller as $\mathbf{w}(n)$ is nearer to $\Psi(n)$. From linear algebraic theory [8], we know that all the perpendicular vectors to the hyperplane $\Psi(n)$ are parallel to $\mathbf{x}(n)$. In addition, the direction vector for updating the weight vector in (2) is parallel to $\mathbf{x}(n)$ since $\mu e(n)$ is a scalar quantity. Thus from a geometric perspective $\mathbf{w}(n)$ is updated toward and perpendicular to the hyperplane $\Psi(n)$. This interpretation is visually described in Fig. 1.

Based on the above geometric description, we explain the relation between the convergence rate and the input vectors. The key is in an acute angle between the current and the previous hyperplane. To best visualize the result, consider the case when $K = 2$. Then two hyperplanes are defined as

$$\Psi(n-1) = \{(w_0, w_1)|x(n-1)w_0 + x(n-2)w_1 = d(n-1)\}$$

and

$$\Psi(n) = \{(w_0, w_1)|x(n)w_0 + x(n-1)w_1 = d(n)\}.$$

In this case two hyperplanes are two straight lines. We assume that the step-size is chosen so that the convergence speed is maximized. Let $\theta(n)$ be the acute angle between two hyperplanes $\Psi(n-1)$ and $\Psi(n)$ and $\mathbf{w}^*$ be the desired solution, as illustrated in Fig. 2. Then it is easily derived that

$$\cos\theta(n) = \frac{\|\mathbf{w}(n+1) - \mathbf{w}^*\|}{\|\mathbf{w}(n) - \mathbf{w}^*\|}. \tag{3}$$

Consider the case $\theta_1(n) > \theta_2(n)$ where

$$\cos\theta_i(n) = \frac{\|\mathbf{w}_i(n+1) - \mathbf{w}^*\|}{\|\mathbf{w}(n) - \mathbf{w}^*\|} \qquad i = 1, 2.$$

Then since $\cos\theta_1(n) < \cos\theta_2(n)$, it follow that

$$\|\mathbf{w}_1(n+1) - \mathbf{w}^*\| < \|\mathbf{w}_2(n+1) - \mathbf{w}^*\|. \tag{4}$$

(4) means that $\mathbf{w}_1(n+1)$ is nearer to $\mathbf{w}^*$ than $\mathbf{w}_2(n+1)$. So we know that $\mathbf{w}(n+1)$ is closer to $\mathbf{w}^*$ as $\theta(n)$ increases from 0° to 90°. So, for fast convergence it is desired that $\theta(n)$ is close to 90°.

Note that the acute angle $\theta(n)$ is equal to the angle between two vectors, $\mathbf{x}(n)$ and $\mathbf{x}(n-1)$ which are perpendicular to $\Psi(n)$ and $\Psi(n-1)$, respectively. So, the angle can be expressed as

$$\cos\theta(n) = \frac{\mathbf{x}^T(n-1)\mathbf{x}(n)}{\|\mathbf{x}(n-1)\| \cdot \|\mathbf{x}(n)\|} \tag{5}$$

using the inner product property of $\mathbf{x}(n)$ and $\mathbf{x}(n-1)$. This result holds for a higher dimensional vector space, i.e., $K > 2$. As can be seen in (5), the angle between two hyperplanes is determined by the inner product between two input vectors at adjacent time. When $\mathbf{x}(n)$ is orthogonal to $\mathbf{x}(n-1)$, $\theta(n)$ becomes 90° and thus fast convergence is achieved.

3. CONVERGENCE ACCELERATION USING SUCCESSIVE DATA ORTHOGONALIZATION

From Sec. 2 we know that the desired condition for fast convergence is that $\mathbf{x}(n)$ is orthogonal to $\mathbf{x}(n-1)$, i.e., $\mathbf{x}^T(n)\mathbf{x}(n-1) = 0$. To meet the desired condition we construct new orthogonal input signal vectors by using the Gram-Schmidt orthogonalization procedure, which is a step-by-step procedure for constructing an orthogonal basis from an existing non-orthogonal basis [8].

Assume that the previous hyperplane and the current hyperplane are defined by

$$\Psi'(n-1) = \{\mathbf{w}|\mathbf{x}'^T(n-1)\mathbf{w} = d'(n-1)\}$$

and

$$\Psi(n) = \{\mathbf{w}|\mathbf{x}^T(n)\mathbf{w} = d(n)\},$$

respectively. According to the Gram-Schmidt procedure, a new orthogonal input vector to $\mathbf{x}'(n-1)$ is obtained by

$$\mathbf{x}'(n) = \mathbf{x}(n) - \frac{\mathbf{x}'^T(n-1)\mathbf{x}(n)}{\|\mathbf{x}'(n-1)\|^2}\mathbf{x}'(n-1), \tag{6}$$

where $\mathbf{x}'(0) = \mathbf{x}(0)$. It can be easily seen that $\mathbf{x}'(n)$ is orthogonal to $\mathbf{x}'(n-1)$.

Consider two linear equations related with $\Psi'(n-1)$ and $\Psi(n)$:

$$\mathbf{x}'^T(n-1)\mathbf{w} = d'(n-1) \quad (7)$$

and

$$\mathbf{x}^T(n)\mathbf{w} = d(n). \quad (8)$$

Let $\frac{\mathbf{x}'^T(n-1)\mathbf{x}(n)}{\|\mathbf{x}'(n-1)\|^2}$ be $\alpha(n)$. Multiplying $\alpha(n)$ to (7) and substracting from (8) lead to

$$\mathbf{x}^T(n)\mathbf{w} - \alpha(n)\mathbf{x}'^T(n-1)\mathbf{w} = d(n) - \alpha(n)d'(n-1), \quad (9)$$

Using (6), (9) reduces to

$$\mathbf{x}'^T(n)\mathbf{w} = d(n) - \alpha(n)d'(n-1). \quad (10)$$

From (10) a new desired output $d'(n)$ corresponding to $\mathbf{x}'(n)$ is given by

$$d'(n) = d(n) - \alpha(n)d'(n-1), \quad (11)$$

where $d'(0) = d(0)$. Therefore a new hyperplane $\Psi'(n)$ is established:

$$\Psi'(n) = \{\mathbf{w} | \mathbf{x}'^T(n)\mathbf{w} = d'(n)\}.$$

With a new input vector $\mathbf{x}'(n)$ and a new desired output $d'(n)$, a new error is defined as

$$e'(n) = d'(n) - \mathbf{x}'^T(n)\mathbf{w}(n). \quad (12)$$

Using (6), (11), and (12), the proposed update equation to minimize $e'^2(n)$ is given by

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e'(n)\mathbf{x}'(n). \quad (13)$$

From a geimetric perspective, the above Gram-Schmidt procedure forms a new hyperplane $\Psi'(n)$ from $\Psi(n)$ and the proposed algorithm in (13) updates $\mathbf{w}(n)$ toward and perpendicular to a newly defined hyperplane $\Psi'(n)$.

Note that the OGA algorithm uses $e(n)$ instead of $e'(n)$ in (13). The computational increase over the OGA algorithm is related only with $d'(n)$ in (11). So one more multiplication and one more addition are required. We can obtain a normalized version of the proposed algorithm by employing $\mu = 1/\|\mathbf{x}'(n)\|^2$.

4. SIMULATION RESULTS

To evaluate the convergence properties of the proposed method computer simulations are carried out in the system identification and the channel equalization problem. For the performance comparison the normalized OGA method and the CG method are selected:

$$\begin{aligned} \text{Normalized OGA[7]}: &\quad \mathbf{w}(n+1) = \mathbf{w}(n) + \mu\frac{e(n)\mathbf{x}'(n)}{\|\mathbf{x}(n)\|^2} \\ \text{CG[6]}: &\quad \text{refer to [6]} \\ \text{Proposed}: &\quad \mathbf{w}(n+1) = \mathbf{w}(n) + \mu\frac{e'(n)\mathbf{x}'(n)}{\|\mathbf{x}'(n)\|^2}. \end{aligned}$$

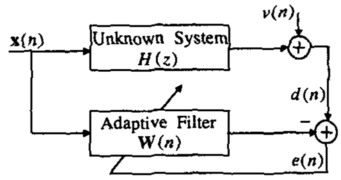

Fig. 3. System identification configuration

4.1. System Identification

The system identification problem is to estimate the impulse response of a unknown system. The system identification configuration is shwon in Fig. 3. The unknown system $H(z)$ is represented by a moving average (MA) model

$$H(z) = \sum_{k=0}^{K-1} h_k z^{-k},$$

where

$$\mathbf{h}^T = [h_0(n) \ -1.0 \ 0.5 \ 0.5 \ -0.5 \ -0.8 \ 0.3 \ 0.1 \ -0.5].$$

To check both convergent and tracking capability a time-varying component $h_0(n)$ is given by

$$h_0(n) = \begin{cases} 1 & \text{if } n < 3000 \\ 1 + 0.5\sin(2\pi n/3000) & \text{otherwise.} \end{cases}$$

The unknown system $H(z)$ is driven by a correlated zero mean signal $x(n)$. The input signal $x(n)$ is generated by filtering Gaussian zero-mean white noise through autoregressive (AR) filter such that the eigenspread could be set to 1600 and 1700. Also the Gaussian zero-mean white noise $v(n)$ with the variance of σ_v^2 is added to the output of the unknown system. Then the desired signal $d(n)$ is given by

$$d(n) = \sum_{k=0}^{K-1} h_k x(n-k) + v(n).$$

For simulations, we assume that $K = 9$, $\sigma_v^2 = 10^{-4}$, and $\mu = 0.1$. Each simulation is carried out 50 times and averaged. Fig. 4 shows the learning curves of the normed squared parameter errors. From the results, we can see the proposed algorithm outperforms the normalized OGA (NOGA) in terms of convergence speed and tracking capability and is comparable with the CG method.

4.2. Channel Equalization

We evaluate the performance of the proposed algorithm in the simulation of a transversal adaptive equalizer for equalizing the distortion introduced in a band-limited channel. The channel equalization configuration is shown in Fig. 5

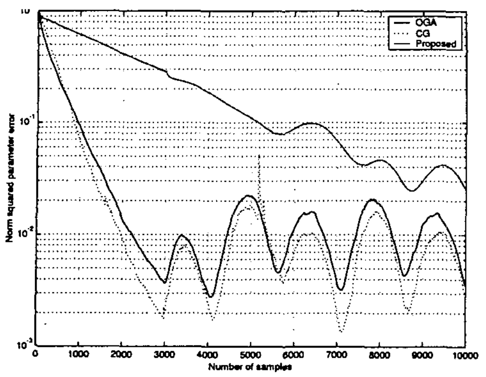

Fig. 4. Comparison of the norm squared parameter errors (system identification)

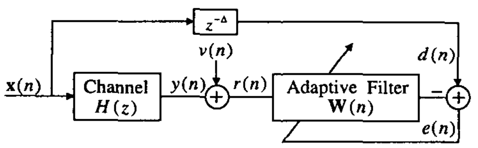

Fig. 5. Channel equalization configuration

The discrete-time channel model for the simulation is given by

$$H(z) = \sum_{0}^{1} h_k z^{-k} = 1 + 0.9z^{-1}.$$

The input to the channel is a random sequence $x(n)$ with values $\{\pm 1\}$. Then the input-output relation for the channel has the form

$$y(n) = \sum_{k=0}^{K-1} h_k x(n-k)$$

and the observed input to the equalizer is given by

$$r(n) = y(n) + v(n),$$

where $v(n)$ is white measurement noise with variance σ_v^2 and is independent of the data $x(n)$. Then the error signal $e(n)$ is given by

$$e(n) = d(n) - \sum_{k=0}^{K-1} w_k(n) r(n-k).$$

For the simulations we assume that $K = 50$, $\sigma_v^2 = 10^{-4}$, $\Delta = 1$, and $\mu = 0.1$. Each simulation is carried out 50 times and averaged. As can be seen in Fig. 6, the proposed algorithm outperforms the NOGA method and the CG method in terms of the convergence rate.

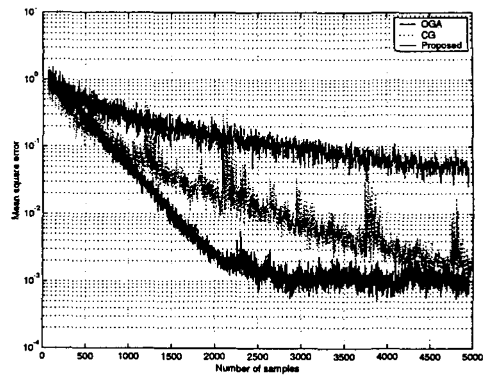

Fig. 6. Comparison of the mean sqaure errors (channel equalization)

5. CONCULSIONS

We have presented a new adaptive filtering algorithm in a simple manner that makes the convergence rate highly improved even for a highly correlated input signal. By keeping the input signal vectors orthogonal at adjacent time, the fast convergence is achieved. The proposed algorithm has been derived in a very general framework. So it can be easily applicable to various applications such as channel equalization, echo cancellation, and so on.

6. REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1985.

[2] S. Haykin, *Adaptive Filter Theory*, NJ: Prentice Hall, second edition, 1991.

[3] G. O. Glentis, K. Berberidis, and S. Theodoridis, "Efficient least squares adaptive algorithms for FIR transversal filtering," *IEEE Signal Processing Mag.*, vol. 16, pp. 13–41, Apr. 1999.

[4] J. G. Proakis, "Channel identification for high speed digital communications," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 916–922, Dec. 1974.

[5] G. K. Boray and M. D. Srinath, "Conjugate gradient techniques for adaptive filtering," *IEEE Trans. Circuits Syst. I*, vol. CAS-39, pp. 1–10, Jan. 1992.

[6] P. S. Chang and A. N. Willson Jr., "Adaptive filtering using modified conjugate gradient," *Proceedings of the 38th Midwest Symposium on Circuits and Systems*, vol. 1, pp. 243–246, 1996.

[7] J. S. Lim, "New adaptive filtering algorithms based on an orthogonal projection of gradient vectors," *IEEE Signal Processing Lett.*, vol. 7, pp. 314–316, Nov. 2000.

[8] G. Strang, *Linear algebra and its applications*, Florida: Harcourt Brace Jovanovich, 1988.

Acknowledgments

This work was supported in part by the Brain Korea (BK) 21 Program funded by the Ministry of Education, Republic of Korea.