논문 2008-45SP-2-10

입력 신호의 연속적인 직교화를 통한 LMS 알고리즘의 수렴 속도 향상

(Convergence Acceleration of the LMS Algorithm Using Successive Data Orthogonalization)

신 현 출*

(Hyun-Chool Shin)

요 약

적응 필터의 입력 신호의 상관도 (correlation)가 클 경우 LMS 알고리즘의 수렴 속도는 상당히 느려지게 된다. 본 논문에서는 입력 신호의 상관도가 높은 상황에서 수렴 속도를 향상시킬 수 있는 적응 필터링 알고리즘을 제안한다. 입력 신호에 대하여 직교성을 가지도록 변환을 인위적으로 가하여 LMS 알고리즘의 한계를 극복한다. 제안한 알고리즘의 성능 향상은 시스템 식별 모델을 통하여 그 수렴 속도의 개선을 확인하며 또한 시변 환경 하에서 적응 필터의 시변 추적 능력을 통해 보여 진다.

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 much improved even for highly correlated input signals. By introducing an orthogonal constraint between successive input signal vectors we overcome the slow convergence problem of the LMS algorithm with the correlated input signal. Simulation results show that the proposed algorithm yields fast convergence speed and excellent tracking capability under both time-invariant and time-varying environments, while keeping both computation and implementation simple.

Keywords : *Adaptive filters, LMS, Orthogonalization, convergence speed, Gram-schmidt*

I. Introduction

Adaptive filtering has drawn much attention since its introduction due to the capability to cope with a changing environment. The least-mean square (LMS) algorithm is certainly one of the most frequently used adaptive filtering algorithms due to its simplicity[1]. The correlation of an input signal, however, highly deteriorates the convergence speed of LMS adaptive filters. In recent years, considerable efforts have been made to improve the convergence rate of LMS.

As a result, many variants of the LMS method have been devised with simple modification or additional filtering to improve the convergence rate. Proakis proposed a variant of the LMS method where gradient vectors are linearly filtered[2]. As another attempt for fast convergence, a conjugate gradient (CG) method has been developed[3~5]. Although the CG method has convergence properties superior to those of ordinary LMS, the CG algorithm requires much higher computational complexity than the LMS method. Recently the orthogonal gradient adaptive (OGA) algorithm which filters the gradient vector so that the current gradient vector is orthogonal to the

* 정회원, 숭실대학교 정보통신전자공학과
(Soongsil University)

※ This work was supported by the Soongsil University Research Fund.
접수일자: 2007년12월12일, 수정완료일: 2008년2월27일

previous one was proposed[6]. Although the OGA algorithm is computationally as simple as the LMS method, the convergence speed is much slower than the CG algorithm.

In this letter we propose a new adaptive filtering algorithm based on successive input data orthogonalization. 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. The Gram-Schmidt orthogonalization procedure is used to achieve the orthogonal relation between input vectors. The proposed method indeed shows the fast convergence speed comparable with the CG algorithm while keeping computationally as simple as the OGA algorithm. Throughout the letter, the following notations are adopted:

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

Ⅱ. 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 error between the desired signal and the adaptive filter output is given by

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

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 algorithm for 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 small positive constant for the step-size. The LMS algorithm updates $\mathbf{w}(n)$ so that $e^2(n)$ is minimized.

To observe 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.,

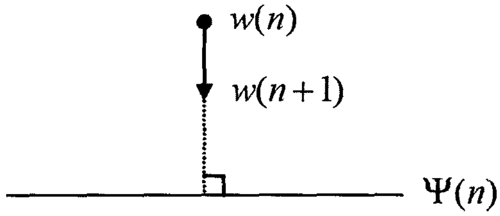


그림 1. LMS 알고리즘의 기하학적 해석
Fig. 1. Geometric interpretation of LMS update.

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

Then the LMS algorithm moves $\mathbf{w}(n)$ toward the hyperplane $\Psi(n)$ since $e^2(n)$ becomes smaller as $\mathbf{w}(n)$ gets closer to $\Psi(n)$. From the linear algebraic theory[7], we know that all the perpendicular vectors to the hyperplane $\Psi(n)$ are parallel to $\mathbf{x}(n)$. The modification vector $\mu e(n)\mathbf{x}(n)$ in (2) is parallel to $\mathbf{x}(n)$ since $\mu e(n)$ is a scalar quantity. This means that the modification vector is perpendicular to $\Psi(n)$. Thus from a geometric perspective $\mathbf{w}(n)$ gets updated toward and perpendicular to the hyperplane $\Psi(n)$. This interpretation is visually described in Fig. 1.

Based on the above geometric framework, we can examine the relation between the convergence rate and the characteristics of the input vectors. We will pay attention to the acute angle between the current and the previous hyperplane. Let $\theta(n)$ be the acute angle. To best visualize the situation, consider the case when $K = 2$. Then the previous and the current hyperplanes are defined as

$$\Psi(n-1) = \mathbf{w} | \mathrm{x}(n-1)\mathrm{w}_0 + \mathrm{x}(n-2)\mathrm{w}_1 = \mathrm{d}(n-1)$$

and

$$\Psi(n) = (\mathrm{w}_0, \mathrm{w}_1) | \mathrm{x}(n)\mathrm{w}_0 + \mathrm{x}(n-1)\mathrm{w}_1 = \mathrm{d}(n),$$

respectively. In this case the two hyperplanes are two straight lines. Then the solution to the two linear equations defining the two hyperplanes is the crossing point $\mathbf{w}^*$ of the straight lines. We assume that the step-size is chosen to maximize the convergence speed. This situation is illustrated in Fig. 2.

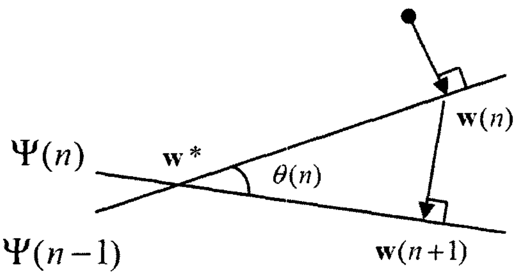


그림 2. 수렴 속도와 입력 신호와의 관계
Fig. 2. Geometric illustration of convergence rate and input signal.

Then it is easily derived that

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

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

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

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

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

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

Note that the acute angle $\theta(n)$ is equal to the angle between two vectors, $\mathrm{x}(n)$ and $\mathrm{x}(n-1)$ which are perpendicular to $\Psi(n)$ and $\Psi(n-1)$, respectively. By using the inner product property of $\mathrm{x}(n)$ and $\mathrm{x}(n-1)$, the angle can be expressed by

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

It is obvious that 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 of adjacent time. When $\mathrm{x}(n)$ is orthogonal to $\mathrm{x}(n-1)$, $\theta(n)$ becomes 90° and thus fastest convergence is achieved.

Ⅲ. Convergence Acceleration Using successive data orthogonalization

From Sec.Ⅱ we know that the desired condition for fast convergence is that $\mathrm{x}(n)$ is orthogonal to $\mathrm{x}(n-1)$, i.e., $\mathrm{x}^T(n)\mathrm{x}(n-1) = 0$. 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[7].

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

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

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

Assume that the desired outputs when the input vectors are $\mathrm{x}(n)$ and $\mathrm{x}'(n-1)$ are $d(n)$ and $d'(n-1)$, respectively. Then from the superposition theory the output $d'(n)$ corresponding to the input vector $\mathrm{x}'(n)$ in (6) is given by

$$d'(n) = d(n) - \alpha(n)d'(n-1) \tag{7}$$

where $\alpha(n) = \dfrac{\mathrm{x}'^T(n-1)\mathrm{x}(n)}{\| \mathrm{x}'(n-1) \|^2}$ and $d'(0) = d(0)$.

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

$$e'(n) = d'(n) - \mathrm{x}'^T(n)\mathrm{w}(n) \tag{8}$$

Using (6), (7), and (8), the proposed update equation to minimize $e'^2(n)$ is given by

$$\mathrm{w}(n+1) = \mathrm{w}(n) + \mu e'(n)\mathrm{x}'(n). \tag{9}$$

The proposed update equation in (9) can be geometrically interpreted by establishing a hyperplane:

$$\Psi'(n) = \mathrm{w} | \mathrm{x}'(\mathrm{n})\mathrm{w} = \mathrm{d}'(\mathrm{n}) \tag{10}$$

The above Gram-Schmidt procedure forms a new hyperplane $\Psi'(n)$ out of $\Psi(n)$ and the proposed algorithm in (9) updates $\mathrm{w}(n)$ toward and perpendicular to a newly defined hyperplane $\Psi'(n)$ instead of $\Psi(n)$. The angle between $\Psi'(n)$ and $\Psi'(n-1)$ is equal to 90° and thus fast convergence is achieved. Also we can obtain a normalized version of the proposed algorithm by simply setting $\mu=\gamma/\parallel \mathbf{w}'(n)\parallel^2 \quad (0<\gamma\le 1)$.

Ⅳ. Simulation Results

To evaluate the convergence properties of the proposed method computer simulations are carried out in the system identification configuration. For comparison the normalized OGA[6] method and the CG method[4] are selected.

The system identification problem is to estimate the impulse response of a unknown system. 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

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

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 an autoregressive (AR) filter such that the eigenspread has values between 1600 and 1700. Also 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 $\alpha=0.1$. Each simulation is carried out 50 times and averaged. Fig. 3 shows the learning curves of the norm-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. It is almost as good as the CG method whose computational load is prohibitive as shown in Table I.

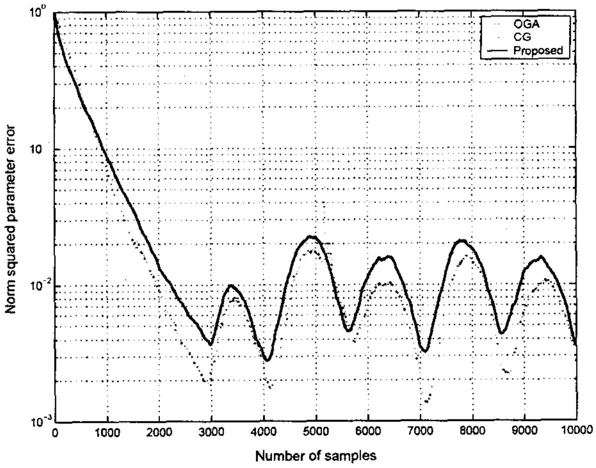


그림 3. 제안한 알고리즘의 수렴 성능의 비교
Fig. 3. Performance comparison of the norm squared parameter error.

표 1. 제안한 알고리즘의 계산량 비교
Table 1. Computational complexity of the proposed algorithm.

	OGA [4]	CG[4]	Proposed
Number of multiplications	5K+2	$2K^2$+10K+3	5K+3

Ⅴ. Conclusions

We have presented a new adaptive filtering algorithm that makes the convergence rate much improved even for a highly correlated input signal. As we see in Fig. 3 and Table I, the fast convergence of the CG method requires an extremely large amount of computation, multiplications of $O(K^2)$. On the other hand, the OGA method requires multiplications of $O(K)$ but its convergence speed is very slow. The proposed method overcomes these shortcomings of CG and OGA, resulting in much improved convergence speed which is comparable to CG, and computational cost as low as that of OGA.

참 고 문 헌

[1] B. Widrow and S. D. Sterns, *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.

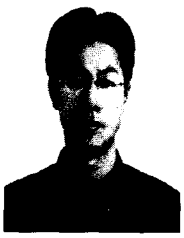
[6] P. S. Chang and A. N. Willson Jr., "Adaptive filtering using modified conjugate gradient," *Proceedings of the 38th Midwest Symposium on Circuits and Systems*, vol. 1, pp.243-246, 1996.

[7] P. S. Chang and A. N. Willson Jr., "Analysis of conjugate gradient algorithms for adaptive filtering," *IEEE Trans. Signal Processing*, vol. 48, pp.409-418, Feb. 2000.

[8] 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.

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

저 자 소 개



신 현 출(정회원)

1997년 포항공과대학교 전자전기 공학과 학사

1999년 포항공과대학교 전자전기 공학과 석사

2004년 포항공과대학교 전자전기 공학과 박사

2004년～2007년 존스홉킨스의대 신경공학연구소 박사후연구원

2007년～현재 숭실대학교 IT대학 정보통신전자 공학과 전임강사

<주관심분야 : 적응신호처리, 생체신호처리, 뇌신경공학>