Convergence Acceleration of the LMS Algorithm Using Successive Data Orthogonalization

Hyun-Chool Shin
Division of Electronics and Computer Engineering
University of Science and Technology (POSTECH)
San 31 Hyoja Dong, Pohang,
Kyungbuk, 790-784, Rep. of Korea
e-mail: shinhe@postech.ac.kr

Woo-Jin Song
Division of Electronics and Computer Engineering
University of Science and Technology (POSTECH)
San 31 Hyoja Dong, Pohang,
Kyungbuk, 790-784, Rep. of Korea
e-mail: wjsong@postech.ac.kr

ABSTRACT
We propose a new adaptive filtering algorithm whose convergence rate is very fast even for a highly correlated input signal. It is well-known that convergence rate gets worse when the input signal to an adaptive filter is correlated. Introducing an orthogonal constraint between successive input signal vectors makes us overcome the slow convergence caused by the correlated input signal. It is shown that the proposed algorithm yields highly improved convergence speed and tracking capability for both time invariant and time varying environments, while being very simple both in computation and implementation.

KEY WORDS
Convergence rate, LMS algorithm, adaptive filters.

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. However, the correlated nature of an input signal highly degrades the convergence speed of LMS adaptive filters. To improve the convergence and tracking property, many variants [1]–[3] of the LMS algorithm have been devised. Among them, the linearly filtered gradient adaptive (FGA) algorithm was the first attempt as a reduced form of the conjugate gradient algorithm [4]. It is also shown that the momentum LMS (MLMS) algorithm can be viewed as a linearly filtered gradient algorithm [5].

Recently, the orthogonal gradient adaptive (OGA) algorithm [6] based on orthogonal projection of gradient vectors was proposed to improve the convergence properties of the FGA algorithm. The orthogonal gradient vector is obtained by filtering a gradient vector through a single pole filter of which the coefficient is chosen so that the current gradient vector is orthogonal to the previous gradient vector. In the OGA algorithm, fast convergence is achieved from the orthogonal update direction. In this case, however, the length of gradient vectors changes inaccurately due to the orthogonal projection of gradient vectors. This makes the performance of the OGA algorithm limited. Note that gradient vectors are scalar-weighted vectors of the input data vectors. So, if we make the current input data vector be orthogonal to the previous one, we can get orthogonal gradient vectors without changing the length of gradient vectors, i.e., no orthogonal projection process for gradient vectors is required.

In this letter, we propose a new adaptive filtering algorithm based on successive input data orthogonalization. We show from a geometric perspective that the orthogonality between the current input vector and the previous input vector is an important condition for fast convergence. Based on the fact, a new algorithm is derived using the Gram-Schmidt orthogonalization process [7]. Since successive input vectors become orthogonal, the merits of the OGA algorithm are inherited and shortcomings are overcome. It is shown that the proposed algorithm yields highly improved convergence speed and tracking capability for both time-invariant and time-varying environments compared with the OGA algorithm.

Throughout the letter, the following notations are adopted:

\( x^T \)  Transpose of \( x \)
\( \| x \| \)  Euclidean norm of \( x \).

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

\[
e(n) = d(n) - x^T(n)w(n),
\]

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

\[
w(n+1) = w(n) + \mu e(n)x(n),
\]

where \( \mu \) is a positive step-size. The LMS algorithm updates \( w(n) \) so that \( e^2(n) \) is minimized.
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

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

(3)

From (3) and Fig. 2, we know that $\mathbf{w}(n + 1)$ is closer to $\mathbf{w}^*$ as $\theta(n)$ increases from $0^\circ$ and $90^\circ$. So, for fast convergence it is desired that $\theta(n)$ is close to $90^\circ$.

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. The angle between $\mathbf{x}(n)$ and $\mathbf{x}(n - 1)$ can be obtained using the inner product property. Consequently, it follows that

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

(4)

This result holds for a higher dimensional vector space, i.e., $K > 2$. As can be seen in (4), the angle between two hyperplanes is mainly determined by the correlation between two input vectors at adjacent time. The more highly the input signal is correlated, the smaller the angle between two hyperplanes becomes and thus the slower the convergence speed is.

3. Convergence Acceleration Using Successive Data Orthogonalization

The desired condition for fast convergence is that $\mathbf{x}(n)$ is orthogonal to $\mathbf{x}(n - 1)$, i.e., $\mathbf{x}^T(n)\mathbf{x}(n - 1) = 0$. This condition implies $\theta(n) = 90^\circ$. To meet the desired condition we construct new orthogonal input signal vectors by using the Gram-Schmidt orthogonalization process, which is a step-by-step procedure for constructing an orthogonal basis from an existing non-orthogonal basis.

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 process, a new input vector is obtained by

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

(5)

where $\mathbf{x}'(0) = \mathbf{x}(0)$. It can be easily seen that $\mathbf{x}'(n)$ is orthogonal to $\mathbf{x}'(n - 1)$. Correspondingly, a new desired output is obtained by

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

(6)

where $d'(0) = d(0)$. Then we can establish new hyperplane $\Psi'(n)$ which is perpendicular to $\mathbf{x}'(n)$:

$$\Psi'(n) = \{ \mathbf{w}|\mathbf{x}'^T(n)\mathbf{w} = d'(n) \}.$$
By updating \( w(n) \) toward \( \Psi'(n) \) instead of \( \Psi(n) \), the convergence can be accelerated.

The resulting update equation using the orthogonal input vectors is given by
\[
w(n+1) = w(n) + \mu e'(n)x'(n),
\]
where \( e'(n) = d'(n) - x'^T(n)w(n) \). Note that the OGA algorithm uses \( e(n) \) instead of \( e'(n) \) in (7). The computational increase over the OGA algorithm is related only with \( d'(n) \) in (6). So one more multiplication and one more addition are required. We can obtain a normalized version of the proposed algorithm by employing \( \mu = 1/\|x'(n)\|^2 \).

4. Simulation Results

Computer simulations are carried out in the system identification configuration to illustrate the convergence property of the proposed algorithm. The unknown system \( H(z) \) is represented by a moving average (MA) model
\[
H(z) = \sum_{k=0}^{K-1} h_k z^{-k}.
\]

Two types of the unknown system are used for the simulation. One is time-invariant, i.e.,
\[
h_1^T = [1.0 -1.0 0.5 2.5 -2.5 -0.8 0.3 0.1 -2.5].
\]
The other is a time-varying system given by
\[
h_2(n) = (1.0 + 0.6 \sin(2\pi n/500))h_1.
\]
The correlated input signal to the unknown system is generated by
\[
x(n) = \alpha u(n) + 1.8 x(n-1) - 0.81 x(n-2),
\]
where \( u(n) \) is zero-mean Gaussian random sequence with unit variance and the gain \( \alpha \) is chosen so that \( x(n) \) has unit power. For the simulations, we have \( \alpha = 0.6 \). Also, the Gaussian zero-mean white noise \( v(n) \) with the variance of \( \sigma_v^2 \) 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 the simulations we assume that \( K = 9 \) and \( \sigma_v^2 = 0.001 \). Each simulation is carried out 10 times and averaged.

The system identification problem is to estimate \( h_k \) given \( x(n) \) and \( d(n) \). The squared norm of the parameter error is defined by \( \sum_{k=0}^{K-1} (h_k - w_k(n))^2 \). In the simulation, normalized algorithms are used:

- Normalized LMS[8]: \( w(n+1) = w(n) + \frac{e(n)x(n)}{\|x(n)\|^2} \)
- Normalized OGA[6]: \( w(n+1) = w(n) + \frac{e(n)x(n)}{\|x(n)\|^2} \)
- Proposed: \( w(n+1) = w(n) + \frac{e'(n)x'(n)}{\|x'(n)\|^2} \).

Fig. 3 shows the learning curves of the normed squared parameter errors when the unknown system is time-invariant. From the results, we can see the proposed algorithm outperforms the normalized LMS (NLMS) and the normalized OGA (NOGA) in terms of convergence speed. However, it shows larger fluctuation in steady-state. This problem can be overcome by using smaller step-size in steady-state.

To evaluate the effect of nonstationary environments the simulation is performed under the time varying unknown system \( h_2(n) \). As can be seen in Fig. 4, the proposed algorithm outperforms the NOGA algorithm in terms of tracking property as well as convergence rate.

5. Conclusions

We have presented a new adaptive filtering algorithm whose convergence rate is very fast even for a highly corre-
lated input signal. With the Gram-Schmidt orthogonalization process, the input signal vectors at adjacent time become orthogonal and thus the fast convergence is realized. 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.

Acknowledgments

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

References