AFFINE PROJECTION ALGORITHM WITH SELECTIVE REGRESSORS

Kyu-Young Hwang and Woo-Jin Song

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

ABSTRACT

Affine projection algorithm, which updates the weight vector based on several previous input vectors, is an useful adaptive filter to improve the convergence speed of LMS-type filter. However, the computational complexity of adaptation algorithm highly depends on the number of input vectors used for update. In this paper, we propose affine projection algorithm with selective regressors whose purpose is to reduce complexity by selecting a subset of input regressors at every iteration. The optimal selection of input regressors is derived by comparing the cost functions based on the principle of minimum disturbance. The new algorithms show good convergence performance as attested to by various experimental results.

1. INTRODUCTION

Adaptive filters with the use of least-mean-square (LMS) adaptation algorithm have been extensively applied to a wide range of diverse fields such as communications, control, acoustics and speech processings due to its computational simplicity and ease of implementation. However, colored input data tend to deteriorate the convergence performance of LMS-type adaptive filter [1][2]. To overcome this problem, Ozeki and Umeda [3] developed the basic form of an affine projection algorithm (APA) that is based on affine subspace projections. APA is a useful family of adaptive filters whose main purpose is to speed the convergence of LMS-type filters, especially for correlated data. Generally the convergence performance of APA becomes improved as the number of previous input vectors increases but the computational complexity can become prohibitively large.

To reduce the computational complexity, a number of selective partial update NLMS and APA have been proposed [4][5]. These algorithms focus on updating a selected subset of filter coefficients at every iteration because the computational complexity is proportional to the number of filter coefficients. In APA, however, the computational complexity of adaptation algorithm also highly depends on the number of input regressors used for update. Therefore, in this paper, we propose the *selective regressor* APA (SR-APA) whose purpose is to reduce complexity by selecting a subset of input regressors at every iteration. The optimal selection of input regressors is derived by comparing the cost functions based on the principle of minimum disturbance and the geometric interpretation. We also develop, as a special case, NLMS with selective regressors.

The paper is organized as follows. In Section II, we derive the conventional APA by posing the adaptation problem as a constraint optimization problem. In Section III, we develop the SR-APA and provide optimal selection method of regressors. Section IV contains experimental results which illustrate the performance of the new adaptive algorithms and Section V presents conclusions.

This work was supported in part by the Brain Korea (BK) 21 Program funded by the Ministry of Education, and in part by HY-SDR Research Center at Hanyang University under the ITRC Program of MIC.

2. AFFINE PROJECTION ALGORITHM

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

$$d(i) = u_i w^o + v(i) \tag{1}$$

where w^o is an unknown column vector that we wish to estimate, $v(i)$ account for measurement noise and u_i denotes $1 \times M$ row input regressor vectors

$$u_i = \begin{bmatrix} u(i) & u(i-1) & \dots & u(i-M+1) \end{bmatrix}.$$

To update w_i, the constrained minimization problem based on the principle of minimum disturbance, which is solved by the affine projection algorithm, can be written as

$$\min_{w_i} \|w_i - w_{i-1}\|^2 \qquad \text{subject to} \qquad d_i = U_i w_i \tag{2}$$

where

$$d_i = \begin{bmatrix} d(i) \\ d(i-1) \\ \vdots \\ d(i-L+1) \end{bmatrix}, U_i = \begin{bmatrix} u_i \\ u_{i-1} \\ \vdots \\ u_{i-L+1} \end{bmatrix}.$$

It can be solved by using the method of Lagrange multipliers[1]. The cost function to be minimize is

$$J(i) = \|w_i - w_{i-1}\|^2 + 2\text{Re}[\Lambda(d_i - U_i w_i)] \tag{3}$$

where $\Lambda = \begin{bmatrix} \lambda_0 & \lambda_1 & \dots & \lambda_{L-1} \end{bmatrix}$, λ ia a Lagrange multiplier and Re(x) denotes a real part of x. Setting $\partial J(i)/\partial w_i^* = 0$ and $\partial J(i)/\partial \Lambda = 0$, we get

$$w_i - w_{i-1} - U_i^* \Lambda^* = 0 \tag{4a}$$

$$d_i - U_i w_i = 0. \tag{4b}$$

Substituting (4a) into (4b), we get

$$\Lambda^* = (U_i U_i^*)^{-1} e_i \tag{5}$$

where $e_i = d_i - U_i w_{i-1}$. After substituting (5) into (4a) and introducing a small positive stepsize μ, we obtain the following recursion

$$w_i = w_{i-1} + \mu U_i^* (U_i U_i^*)^{-1} e_i. \tag{6}$$

The computational complexity of the APA using L input regressors is $(L^2 + 2L)M + L^3 + L^2$ multiplications per iteration for real data [2].

3. SELECTIVE REGRESSOR AFFINE PROJECTION ALGORITHM (SR-APA)

Our objective is to reduce computational complexity of the original L-order APA by selecting an adequate subset of input regressors at every iteration while minimizing the performance degradation. Let's suppose that we wish to select K input regressors among L input regressors at every iteration. Let $\mathcal{T}_K = \{t_0, t_1, \ldots t_{K-1}\}$ denote a K-subset (subset with K members) of the set $\{0, 1, \ldots, L-1\}$ and let $\mathcal{S}$ be the collection of all K-subsets, i.e., $\mathcal{T}_K \in \mathcal{S}$. We can write a constrained minimization problem for new K-order APA as

$$\min_{w_i} \|w_i - w_{i-1}\|^2 \qquad \text{subject to} \qquad d_{i,\mathcal{T}_K} = U_{i,\mathcal{T}_K} w_i \quad (7)$$

where

$$d_{i,\mathcal{T}_K} = \begin{bmatrix} d(i-t_0) \\ d(i-t_1) \\ \vdots \\ d(i-t_{K-1}) \end{bmatrix}, U_{i,\mathcal{T}_K} = \begin{bmatrix} u_{i-t_0} \\ u_{i-t_1} \\ \vdots \\ u_{i-t_{K-1}} \end{bmatrix}.$$

Then the cost function with selective regressors can be written as

$$J_{\mathcal{T}_K}(i) = \|w_i - w_{i-1}\|^2 + 2\text{Re}[\Lambda_{\mathcal{T}_K}(d_{i,\mathcal{T}_K} - U_{i,\mathcal{T}_K} w_i)]. \quad (8)$$

Similarly, the update equation of APA with selective regressors can be represented by

$$w_i = w_{i-1} + \mu U^*_{i,\mathcal{T}_K}(U_{i,\mathcal{T}_K} U^*_{i,\mathcal{T}_K})^{-1} e_{i,\mathcal{T}_K} \quad (9)$$

where $e_{i,\mathcal{T}_K} = d_{i,\mathcal{T}_K} - U_{i,\mathcal{T}_K} w_{i-1}$.

3.1. Optimal regressor selection

We now turn our attention to how to optimally select the regressors to be used for update at every iteration. Generally, using fewer input regressors in APA causes the performance degradation in convergence speed. Thus the regressor selection should be made by identifying the regressors with the least performance degradation. For this, we should select the regressors which make $J_{\mathcal{J}_K}(i)$ as close as possible to $J(i)$ where $\mathcal{J}_K \in \mathcal{S}$. Assume that the quantity of weight update is small. Assume that the quantity of weight update is small. Then, *a posteriori* error is similar to *a priori* error, i.e., $d_i - U_i w_i \approx e_i$ and $d_{i,\mathcal{J}_K} - U_{i,\mathcal{J}_K} w_i \approx e_{i,\mathcal{J}_K}$. Using this, we find from (3) and (8) that the cost functions can be approximated by

$$\begin{aligned} J(i) &= \|w_i - w_{i-1}\|^2 + 2\text{Re}[\Lambda e_i] && (10a) \\ J_{\mathcal{J}_K}(i) &= \|w_i - w_{i-1}\|^2 + 2\text{Re}[\Lambda_{\mathcal{J}_K} e_{i,\mathcal{J}_K}], && (10b) \end{aligned}$$

respectively. Using the calculation method of the Lagrange multipliers such as (5), the regressor selection problem can be represented as

$$\begin{aligned} \mathcal{T}_K &= \arg\min_{\mathcal{J}_K \in \mathcal{S}} |J(i) - J_{\mathcal{J}_K}(i)| \\ &= \arg\min_{\mathcal{J}_K \in \mathcal{S}} |e_i^*(U_i U_i^*)^{-1} e_i - e^*_{i,\mathcal{J}_K}(U_{i,\mathcal{J}_K} U^*_{i,\mathcal{J}_K})^{-1} e_{i,\mathcal{J}_K}|. \end{aligned} \quad (11)$$

Since $e^*_{i,\mathcal{J}_K}(U_{i,\mathcal{J}_K} U^*_{i,\mathcal{J}_K})^{-1} e_{i,\mathcal{J}_K}$ is always positive and smaller than $e_i^*(U_i U_i^*)^{-1} e_i$, we can rewrite (11) as

$$\mathcal{T}_K = \arg\max_{\mathcal{J}_K \in \mathcal{S}} e^*_{i,\mathcal{J}_K}(U_{i,\mathcal{J}_K} U^*_{i,\mathcal{J}_K})^{-1} e_{i,\mathcal{J}_K}. \quad (12)$$

Table 1. Computational complexity of conventional APA and SR-APA

	Conventional APA	SR-APA	
		Computations for weight update	Additional computations
Multiplications	$(L^2+2L)M$ $+L^3+L^2$	$(K^2+2K)M$ $+K^3+K^2$	$(L-K)M$ $+L+1$
Divisions	-	-	L
Comparisons	-	-	$L\log_2 K$ $+O(L)$

Therefore the proposed SR-APA is given by

$$\begin{aligned} w_i &= w_{i-1} + \mu U^*_{i,\mathcal{T}_K}(U_{i,\mathcal{T}_K} U^*_{i,\mathcal{T}_K})^{-1} e_{i,\mathcal{T}_K} \\ \mathcal{T}_K &= \arg\max_{\mathcal{J}_K \in \mathcal{S}} e^*_{i,\mathcal{J}_K}(U_{i,\mathcal{J}_K} U^*_{i,\mathcal{J}_K})^{-1} e_{i,\mathcal{J}_K}. \end{aligned} \quad (13)$$

However, the full implementation of (13) can be computationally very expensive because of the high complexity associated with subset selection. Motivated by the relationship between the matrix norms and quadratic forms proposed in [5], we propose an alternative simplified criterion for regressor selection: Rank $e_j^2(i)/\|u_{i-j}\|^2$, $j \in \{0, 1, \ldots, L-1\}$ and select the regressors associated with K largest values for update where $e_j(i) = d(i-j) - u_{i-j} w_{i-1}$. The simplified criterion is formally given by

$$\begin{aligned} \frac{e^2_{t_0}(i)}{\|u_{i-t_0}\|^2} \geq \frac{e^2_{t_1}(i)}{\|u_{i-t_1}\|^2} \geq \cdots \geq \frac{e^2_{t_{K-1}(i)}}{\|u_{i-t_{K-1}}\|^2} \\ \geq \cdots \geq \frac{e_j^2(i)}{\|u_{i-j}\|^2} \end{aligned} \quad (14)$$

where $j \in \{0, 1, \ldots, L-1\}$. Note that the simplified criterion can be derived when we focus only on the diagonal components of $U_{i,\mathcal{J}_K} U^*_{i,\mathcal{J}_K}$. If $U_{i,\mathcal{J}_K} U^*_{i,\mathcal{J}_K}$ is a diagonal matrix, the maximum value in (12) can be rewritten as

$$\begin{aligned} \max_{\mathcal{J}_K \in \mathcal{S}} e^*_{i,\mathcal{J}_K}\left(U_{i,\mathcal{J}_K} U^*_{i,\mathcal{J}_K}\right)^{-1} e_{i,\mathcal{J}_K} \approx \frac{e^2_{t_0}(i)}{\|u_{i-t_0}\|^2} + \frac{e^2_{t_1}(i)}{\|u_{i-t_1}\|^2} \\ + \ldots + \frac{e^2_{t_{K-1}}(i)}{\|u_{i-t_{K-1}}\|^2}. \end{aligned} \quad (15)$$

which makes (14) be the solution of (12). Although the simplified criterion is not exactly equivalent to (13), it has satisfactory convergence performance, while keeping the computational complexity low. For every input sample, the additional computational complexity for (14) is $(L-K)M + L + 1$ multiplications and L divisions for calculation, and $O(L) + L\log_2 K$ comparisons for regressor selection by the heapsort algorithm [6]. Table 1 shows the computational complexity of conventional APA and the proposed SR-APA. The additional computational complexity is relatively small compared with that for the weight update.

3.2. NLMS with selective regressors

A special case of SR-APA is NLMS with selective regressors obtained by setting $K = 1$. From (13), NLMS with selective regressors

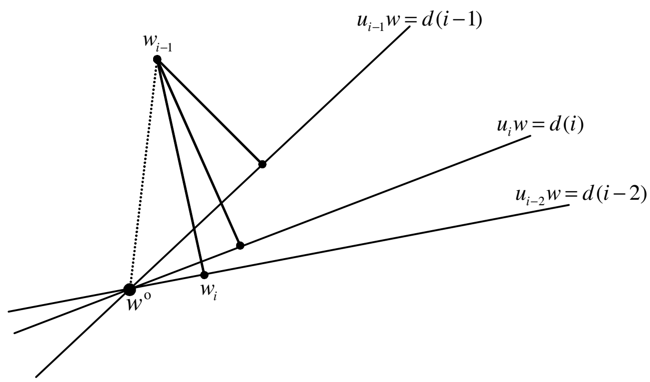

Fig. 1. The weight update example of the proposed algorithm for L=3, K=1 and M=2.

can be represented as

$$w_i = w_{i-1} + \mu \frac{u_{i-t}^*}{\|u_{i-t}\|^2} e_t(i)$$
$$t = \arg \max_{j \in \{0,1,\ldots,L-1\}} \frac{e_j^2(i)}{\|u_{i-j}\|^2} \quad (16)$$

in which the simplified criterion is no more required, that is, the simplified criterion is exactly equivalent to (16). Note that NLMS with selective regressors is the approximated version of L-order APA, which means that NLMS with selective regressors has better convergence performance than conventional NLMS.

3.3. Geometric interpretation

In this section, we will investigate the geometric interpretation of (13). Assume that $\mu = 1$ for convenience. Then, w_i is obtained by projecting the given weight vector, w_{i-1} onto the intersection of the affine subspace defined by $\{d(i-t), u_{i-t}\}$ where $t \in \mathcal{T}_K$. By the Pythagorean theorem, we can write

$$\|w_i - w^o\|^2 = \|w_{i-1} - w^o\|^2 - \|w_i - w_{i-1}\|^2 \quad (17)$$

for given w_{i-1} and w^o. Using the update equation in (13), we can also write

$$\|w_i - w_{i-1}\|^2 = e_{i,\mathcal{T}_K}^* (U_{i,\mathcal{T}_K} U_{i,\mathcal{T}_K}^*)^{-1} e_{i,\mathcal{T}_K}. \quad (18)$$

Substituting (18) into (17), we get

$$\|w_i - w^o\|^2 = \|w_{i-1} - w^o\|^2 - e_{i,\mathcal{T}_K}^* (U_{i,\mathcal{T}_K} U_{i,\mathcal{T}_K}^*)^{-1} e_{i,\mathcal{T}_K} \quad (19)$$

Since $e_{i,\mathcal{T}_K}^* (U_{i,\mathcal{T}_K} U_{i,\mathcal{T}_K}^*)^{-1} e_{i,\mathcal{T}_K}$ is the maximum value for all K-subsets, $\|w_i - w^o\|^2$, the norm square of a posteriori weight error vector is minimized. In other words, the proposed algorithm updates the weight vector based on the combinations of input regressors which have the best convergence speed of all possible combinations. Fig 1. shows the weight update example for $L = 3$, $K = 1$ and $M = 2$. To minimize the norm square of a posteriori weight error vector, u_{i-2} is selected for update.

4. EXPERIMENTAL RESULTS

We illustrate the performance of the proposed algorithm by carrying out computer experiments in a channel estimation in which the unknown channel is randomly generated. The adaptive filter and the unknown channel are assumed to have the same number of taps. The input signal $u(i)$ is obtained by filtering a white, zero-mean, Gaussian random sequence through a first-order autoregressive system

$$G(z) = \frac{1}{1 - 0.9z^{-1}}.$$

As a result, a highly correlated Gaussian signal is generated. The signal-to-noise ratio (SNR) is calculated by

$$\text{SNR} = 10 \log_{10} \left(E\left[y^2(i)\right] / E\left[v^2(i)\right] \right)$$

where $y(i) = u_i w^o$ and the measurement noise $v(i)$ is added to $y(i)$. The step-size is set to $\mu = 0.5$. For better convergence performance, the data $\{d_i, U_i\}$ are taken as

$$d_i = \begin{bmatrix} d(i) \\ d(i-D) \\ \vdots \\ d(i-(L-1)D) \end{bmatrix}, U_i = \begin{bmatrix} u_i \\ u_{i-D} \\ \vdots \\ u_{i-(L-1)D} \end{bmatrix}$$

where $D = 8$. The experimental results are obtained by ensemble average over 200 independents trials.

First, we compare the convergence performance of the conventional APA, the SR-APA and the SR-APA using simplified criterion. Fig.2 shows a plot of the MSE learning curve versus iteration number for the three APA algorithms with SNR = 30dB. The adaptive filter length is set to $M = 32$. The order of conventional APA is set to 16 and the number of selective regressors set to 12 ($K = 12$) out of 16 ($L = 16$). The convergence speed of the proposed 12-order APA with selective regressors is similar to the conventional 16-order APA. Moreover, no significant difference is observed between the convergence speeds of the two proposed algorithms while the simplified version has a better computational merit.

Fig.3 shows the learning curves of NLMS and NLMS with selective regressors. The adaptive filter length is set to $M = 32$ and SNR is set to 30dB. To compare the convergence performance of the proposed algorithms with that of the conventional NLMS, the number of regressor candidates, L is set to 4, 2 while the number of selective regressors is set to $K = 1$. As can be seen, the proposed NLMS with selective regressors converges faster than the conventional NLMS as the number of regressor candidates increases.

The convergence curves for the conventional APA and the proposed algorithm with similar computational complexity are shown in Fig.4. The adaptive filter length is set to $M = 128$ and SNR is set to 50dB. To make the computational complexity similar, the order of the conventional APA is set to 34 (197132 multiplications) and the number of selective regressors is set to $K = 33$ out of $L = 64$ (188643 multiplications, 64 divisions and 332+$O(64)$ comparisons). As can be seen, the proposed algorithm has faster convergence speed than the conventional APA when the computational complexity is set to be similar.

5. CONCLUSIONS

We have proposed the SR-APA whose purpose is to reduce computational complexity by selecting a subset of input regressors. The optimal selection of input regressors has been derived by the comparison

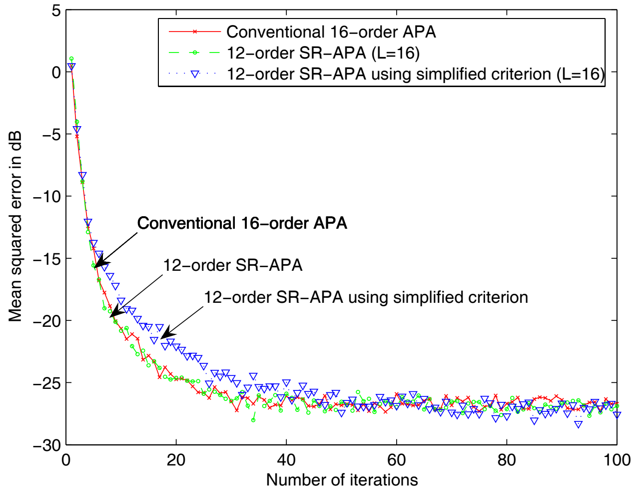

Fig. 2. Convergence curves for the conventional APA, the SR-APA and the SR-APA using simplified criterion.

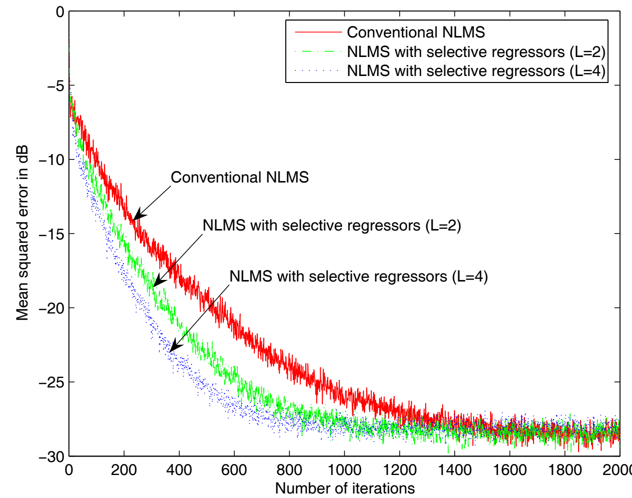

Fig. 3. Convergence curves for the conventional NLMS and NLMS with selective regressors (L=4,2).

of the cost functions based on the principle of minimum disturbance. Moreover, a simplified approximation has been proposed to alleviate the large computational complexity of selection criterion. The simplified criterion has been shown to be capable of maintaining a good convergence performance.

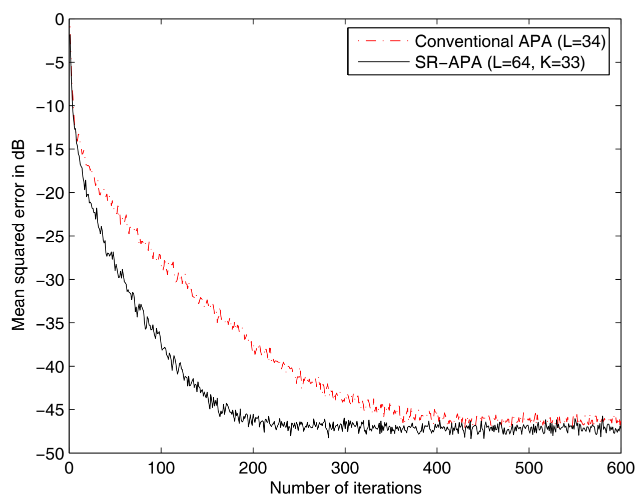

Fig. 4. Convergence curves for the conventional APA and the SR-APA with similar computational complexity.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, New York: Wiley, 2003.

[3] K. Ozeki and T. Umeda, “An adaptive filtering algorithm using an orthogonal projection to an affine subsapce and its properties,” *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19-27, 1984.

[4] S. C. Douglas, “Adaptive filters employing partial updates,” *IEEE Trans. Circuits and Systems-II*, vol. 44, no. 3, pp. 209-216, Mar. 1997.

[5] K. Doğançay and O. Tanrikulu, “Adaptive filtering algorithms with selective partial updates,” *IEEE Trans. Circuits and Systems-II*, vol. 48, no. 8, pp. 762-769, Aug. 2001.

[6] D. E. Knuth, *Sorting and Searching vol.3 of The Art of Computer Programming*, 2nd ed. MA: Addison-Wesley, 1973.