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

Affine Projection Adaptive Filter with Variable Regularization

Young-Seok Choi

Abstract—We propose two affine projection algorithms (APA) with variable regularization parameter. The proposed algorithms dynamically update the regularization parameter that is fixed in the conventional regularized APA (R-APA) using a gradient descent based approach. By introducing the normalized gradient, the proposed algorithms give birth to an efficient and a robust update scheme for the regularization parameter. Through experiments we demonstrate that the proposed algorithms outperform conventional R-APA in terms of the convergence rate and the misadjustment error.

Keywords—Affine projection, regularization, gradient descent, system identification.

I. Introduction

THE normalized least mean square (NLMS) is most frequently used adaptive algorithm due to its simplicity and ease of implementation. However, its convergence rate is significantly reduced for colored input signals [1]–[3]. To overcome this problem, the affine projection algorithm (APA) was proposed by Ozeki and Umeda [4]. While the NLMS updates the weights based only on the current input vectors, the APA updates the weights on the basis of last K input vectors [4], [5]. In a noisy environment, the inversion of a rank deficient matrix may give rise to numerical difficulties. To avoid this situation, a positive constant δ called the *regularization parameter* is used. We use the regularized APA (R-APA) as opposed to simply the APA in order to highlight the presence of the regularization parameter δ; the terminology APA is reserved for the case $\delta = 0$. It is well known that the regularization parameter plays a critical role in the performance of the R-APA [6], [7]. In the R-APA, the regularization parameter δ governs the rate of convergence and the misadjustment error. To meet the conflicting requirements of fast convergence and low misadjustment error, the regularization parameter needs to be controlled.

In this paper, we propose two R-APAs with variable regularization parameter. The proposed schemes use a time-varying regularization parameter which is updated using a gradient descent based approach at each instant. We develop an efficient and robust update scheme for the regularization parameter by introducing the *normalized gradient*. We show that the proposed algorithms have low additional complexity compared to the conventional R-APA. Through several experiments we demonstrate that the proposed algorithms outperform the conventional R-APA in terms of the convergence speed and the misadjustment error

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. Proposed R-APAs

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

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

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

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

A. Regularization for R-APA

Let $\mathbf{w}_i$ be an estimate for $\mathbf{w}^\circ$ at iteration i. The R-APA computes $\mathbf{w}_i$ via

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu U_i^*(U_i U_i^* + \delta I)^{-1}\mathbf{e}_i, \tag{3}$$

where

$$U_i = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{u}_{i-1} \\ \vdots \\ \mathbf{u}_{i-K+1} \end{bmatrix} \quad \mathbf{d}_i = \begin{bmatrix} d(i) \\ d(i-1) \\ \vdots \\ d(i-K+1) \end{bmatrix},$$

$\mathbf{e}_i = \mathbf{d}_i - U_i\mathbf{w}_{i-1}$, μ is the step-size, δ is the regularization parameter and * denotes the Hermitian transpose. The obvious effect of the regularization parameter not only employs to avoid the inversion of a rank deficient matrix $U_iU_i^*$, but also plays a critical role in the convergence performance of the R-APA. A large regularization parameter will ensure small effective step-size and thus the R-APA results in small misadjustment error in steady state, but converges slowly. On the other hand, a small regularization parameter will provide large effective step-size and thus the R-APA converges fast but results in large misadjustment error. Along this line of thought we may expect performance improvement by using a variable regularization parameter instead of a fixed δ.

B. Proposed R-APA with Multiple Regularization Parameter

To achieve this purpose, we propose R-APA which continuously updates the regularization parameter so that $J(i) = \frac{1}{2}e^2(i)$ where $e(i) = d(i) - \mathbf{u}_i\mathbf{w}_{i-1}$. To reduce the cost function $J(i)$, the regularization parameter can be modified by a gradient descent algorithm and will be explained by starting with a formulation of the multiple regularization parameter as follows:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu\, U_i^*(U_iU_i^* + \Delta_i)^{-1}\mathbf{e}_i. \tag{4}$$

where Δ_i is a $K \times K$ diagonal matrix defined by

$$\Delta_i = \mathrm{diag}\left[\delta_0(i),\, \delta_1(i), \ldots, \delta_{K-1}(i)\right]. \tag{5}$$

In the proposed method, we update the regularization parameters in (5) iteratively so that $J(i)$ is minimized. To do this, we use stochastic gradient descent algorithm, i.e.,

$$\begin{aligned}\delta_j(i) \;&=\; \delta_j(i-1) - \rho\, \nabla_\delta J(i)\\ &\text{for } j = 0, 1, \ldots, K-1,\end{aligned} \tag{6}$$

where ρ is a small positive learning rate parameter. The gradient of $J(i)$ with respect to $\delta_j(i-1)$, $\nabla_\delta J(i)$, can be shown to be

$$\nabla_\delta J(i) = \frac{\partial J(i)}{\partial e(i)} \cdot \frac{\partial e(i)}{\partial \mathbf{w}_{i-1}} \cdot \frac{\partial \mathbf{w}_{i-1}}{\partial \delta_j(i-1)}. \tag{7}$$

Each term in (7) is simply derived as

$$\begin{aligned}\frac{\partial J(i)}{\partial e(i)} &= e(i), \qquad \frac{\partial e(i)}{\partial \mathbf{w}_{i-1}} = -\mathbf{u}_i,\\ \frac{\partial \mathbf{w}_{i-1}}{\partial \delta_j(i-1)} &= -\mu U_{i-1}^* \left(U_{i-1}U_{i-1}^* + \Delta_{i-1}\right)^{-2} \Gamma_j\, \mathbf{e}_{i-1},\end{aligned} \tag{8}$$

where we are defining

$$\Gamma_j \equiv \frac{\partial \Delta_{i-1}}{\partial \delta_j(i-1)} = \mathrm{diag}\left[0 \cdots 1 \cdots 0\right], \tag{9}$$

and Γ_j has a value of 1 only at a jth row. Then we have

$$\nabla_\delta J(i) = \mu e(i) \mathbf{u}_i U_{i-1}^* \left(U_{i-1}U_{i-1}^* + \Delta_{i-1}\right)^{-2} \Gamma_j \mathbf{e}_{i-1}. \tag{10}$$

However, we know that $\Delta\delta_j(i) = \delta_j(i) - \delta_j(i-1)$ is proportional to the square order of $e(i)$. So a small $e(i)$ after the initial adaptation results in very small $\Delta\delta_j(i)$ and correspondingly $\delta_j(i)$ undergoes small variation.

In the proposed method, we normalize the gradient, $\nabla_\delta J(i)$, by its norm. The regularization parameter $\delta_j(i)$ is recursively updated by

$$\begin{aligned}\delta_j(i) \;&=\; \delta_j(i-1) - \rho\, \frac{\nabla_\delta J(i)}{\|\nabla_\delta J(i)\|}\\ &\text{for } j = 0, 1, \ldots, K-1,\end{aligned} \tag{11}$$

where ρ is a small positive learning rate parameter and $\|\cdot\|$ denotes the Euclidean norm of a vector. By introducing the normalized gradient, the regularization parameter $\delta_j(i)$ becomes robust to variation of $e(i)$ since the normalized version of gradient $\nabla_\delta J(i)$ with a fixed ρ always makes the same stride, independent of how steep the slope of $J(i)$ is. This property makes the regularization parameter $\delta_j(i)$ relatively stable when $\nabla_\delta J(i)$ is very small.

Then, $\frac{\nabla_\delta J(i)}{\|\nabla_\delta J(i)\|}$ in (11) can be rewritten by

$$\frac{\nabla_\delta J(i)}{\|\nabla_\delta J(i)\|} = \mathrm{sgn}\left(\nabla_\delta J(i)\right), \tag{12}$$

where $\mathrm{sgn}(\cdot)$ is the signum function which takes the sign of variable. From (7), (11) and (12), the proposed R-APA with multiple regularization parameter is given by:

$$\begin{aligned}&\delta_j(i) = \delta_j(i-1) -\\ &\quad -\rho\, \mathrm{sgn}\left(\mu\, e(i)\, \mathbf{u}_i\, U_{i-1}^* \left(U_{i-1}U_{i-1}^* + \Delta_{i-1}\right)^{-2} \Gamma_j \mathbf{e}_{i-1}\right)\\ &\Delta_i = \mathrm{diag}\left(\delta_0(i),\, \delta_1(i), \ldots, \delta_{K-1}(i)\right)\\ &\mathbf{w}_i = \mathbf{w}_{i-1} + \mu\, U_i^*(U_iU_i^* + \Delta_i)^{-1}\mathbf{e}_i.\end{aligned} \tag{13}$$

C. Scalar Regularization Parameter Case

The conventional R-APA update recursion (3) uses the scalar regularization parameter. Recently, a scalar optimum regularization for fast convergence has been started to research as in [6]. In this section, we propose a scalar optimum regularization parameter at each instant by assuming

$$\delta(i) = \delta_0(i) = \cdots = \delta_{K-1}(i). \tag{14}$$

Then the proposed R-APA with a scalar regularization parameter is obtained by

$$\begin{aligned}&\delta(i) = \delta(i-1) -\\ &\quad -\rho\, \mathrm{sgn}\left(\mu\, e(i)\, \mathbf{u}_i\, U_{i-1}^* \left(U_{i-1}U_{i-1}^* + \delta(i)I\right)^{-2} \mathbf{e}_{i-1}\right)\\ &\quad \mathbf{w}_i = \mathbf{w}_{i-1} + \mu\, U_i^*(U_iU_i^* + \delta(i)I)^{-1}\mathbf{e}_i,\end{aligned} \tag{15}$$

since $\frac{\partial \Delta_{i-1}}{\partial \delta_j(i-1)} = \frac{\partial \delta(i-1)I}{\partial \delta(i-1)} = I$.

D. Stability

To guarantee the stability of the proposed algorithms, we need to set $\delta_{\min}$ such as

$$\delta_j(i) = \max(\delta_{\min},\, \delta_j(i)), \tag{16}$$

i.e., if $\delta_j(i)$ gets less than $\delta_{\min}$, $\delta_j(i)$ is replaced by $\delta_{\min}$ at each instant. Also, it is known that the convergence in the mean of R-APA is guaranteed for any μ satisfying [3]

$$0 < \mu < 2. \tag{17}$$

Let us define the *a posteriori* estimation error as

$$\mathbf{r}_i = \mathbf{d}_i - U_i \mathbf{w}_i, \tag{18}$$

i.e., the error in estimating $\mathbf{d}_i$ by using the new weight estimate. Since $U_i\mathbf{w}_i$ will be a better estimate for $\mathbf{d}_i$ than $U_i\mathbf{w}_{i-1}$, the property $\|\mathbf{r}_i\|^2 \le \|\mathbf{e}_i\|^2$ (with equality only when $\mathbf{e}_i = 0$) should be satisfied. Assuming a scalar regularization parameter, it holds that

$$\mathbf{r}_i = \left(I - \mu\, U_i\, U_i^*(U_iU_i^* + \delta(i)I)^{-1}\right) \mathbf{e}_i. \tag{19}$$

and

$$\|\mathbf{r}_i\|^2 = \mathbf{e}_i^*\, A^* A\, \mathbf{e}_i \le \mathbf{e}_i^*\, I\, \mathbf{e}_i = \|\mathbf{e}_i\|^2, \tag{20}$$

where we are defining

$$A = \left(I - \mu\, U_i\, U_i^*(U_iU_i^* + \delta(i)I)^{-1}\right).$$

Therefore, $\|\mathbf{r}_i\|^2 \le \|\mathbf{e}_i\|^2$, if and only if the matrix $I - A^*A$ is positive-definite by (20). In addition, let $U_iU_i^* = V_i\Lambda_iV_i^*$ denotes the eigen-decomposition of the matrix $U_iU_i^*$. Then

$$U_iU_i^* + \delta(i)I = V_i(\Lambda_i + \delta(i)I)V_i^* \tag{21}$$

TABLE I
COMPUTATIONAL COMPLEXITY

Algorithm	multiplications	additions
R-APA [3]	$(K^2+2K)M$ $+K^3+K$	$(K^2+2K)M$ $+K^3+K^2$
Proposed (11)	$KM+K^2+3K+1$	$KM+K^2-K$
Proposed (15)	$KM+K^2+K+3$	$KM+K^2-K$

and

$$(U_iU_i^*+\delta(i)I)^{-1}=V_i(\Lambda_i+\delta(i)I)^{-1}V_i^*. \quad (22)$$

Using the eigen-decomposition of $U_iU_i^*$ and (22), the following holds:

$$\begin{aligned}A^*A &= \left(I-\mu V_i\Lambda_i^{'}V_i^*\right)^*\left(I-\mu V_i\Lambda_i^{'}V_i^*\right)\\ &= I-2\mu V_i\Lambda_i^{'}V_i^*+\mu^2V_i\Lambda_i^{'2}V_i^* \quad (23)\end{aligned}$$

where $\Lambda_i^{'}=\Lambda_i(\Lambda_i+\delta(i)I)^{-1}$. So, it holds that

$$I-A^*A=\mu V_i\Lambda_i^{'}(2-\mu\Lambda_i^{'})V_i^{'}. \quad (24)$$

To satisfy $(I-A^*A)$ is positive-definite, we find

$$\begin{aligned}&2-\mu\Lambda_i^{'}=2-\mu\Lambda_i(\Lambda_i+\delta(i)I)^{-1}\\ &=2-\mu\,\mathrm{diag}\left(\frac{\lambda_o(i)}{\lambda_o(i)+\delta(i)},\cdots,\frac{\lambda_{K-1}(i)}{\lambda_{K-1}(i)+\delta(i)}\right)>0.(25)\end{aligned}$$

Then, we get the lower bound of the regularization parameter for the stability of the proposed algorithms as

$$\delta_{\min}>\lambda_{\max}(i)\left(\frac{\mu}{2}-1\right), \quad (26)$$

where $\lambda_{\max}(i)$ is a maximum value of $\lambda_k(i)$ with $1\le k\le K-1$.

E. Computational Complexity

Table I lists the number of multiplications, additions of the conventional R-APA and the computation of the variable regularization parameter at each instant. We know that the additional costs required to obtain variable regularization parameter are low compared to overall complexity of the conventional R-APA.

III. EXPERIMENTAL RESULTS

We illustrate the performance of the proposed algorithms by carrying out computer simulations in a channel identification scenario. The unknown channel $H(z)$ has 16 taps and is randomly generated. The adaptive filter and the unknown channel are assumed to have the same number of taps. A Gaussian distributed signal is used for the input signal. The input signal is obtained by filtering a white, zero-mean. Gaussian random sequence through a first-order system $G(z)=1/(1-0.9z^{-1})$. The signal-to-noise ratio (SNR) is calculated by

$$\mathrm{SNR}=10\log_{10}\left(\frac{E[y^2(i)]}{E[v^2(i)]}\right), \quad (27)$$

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

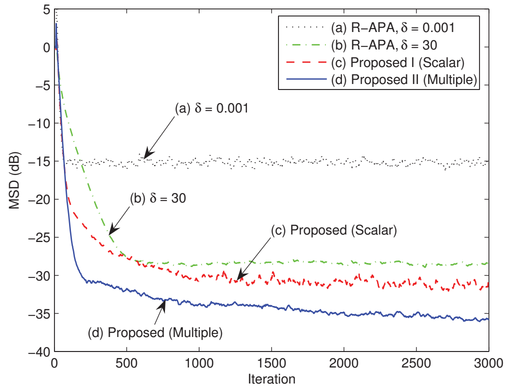

Fig. 1 Plot of MSD curves of the proposed APAs and R-APA (K=8)

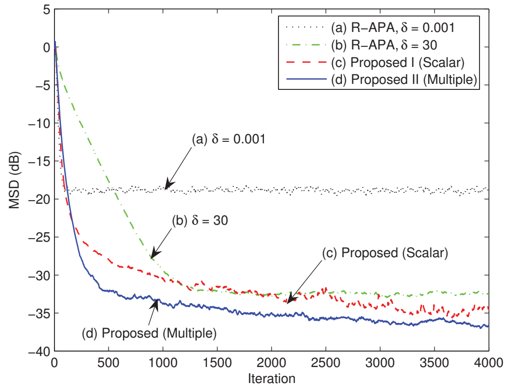

Fig. 2 Plot of MSD curves of the proposed APAs and R-APA (K=4)

The measurement noise $v(i)$ is added to $y(i)$ such that SNR $=30$dB. The mean square deviation (MSD), $E\|\mathbf{w}^\circ-\mathbf{w}_i\|^2$, is taken and averaged over 100 independent trials. The initial value $\delta_j(0)$ is set to 0.001 and $\delta_{\min}$ is chosen to 0.0001 for all experiments.

In Fig. 1, we show the MSD curves for $K=8$, $\mu=0.5$, and $\rho=1.0$. Dashed lines indicate the results of R-APA with fixed regularization parameters where we choose $\delta=0.001$ and 30. As can be seen, the proposed R-APAs have the faster convergence and the lower misadjustment error. In addition, the proposed R-APA with multiple regularization parameter has a improved performance than with common regularization parameter as expected. In Fig. 2, we choose $K=4$, $\mu=0.5$, and $\rho=0.9$. A similar result of Fig. 1 is observed in Fig. 2.

IV. CONCLUSION

We have proposed two R-APAs with variable regularization parameter. The regularization parameter is modified to minimized the given cost function using a gradient descent based approach. By optimizing the regularization parameter at each instant, the proposed algorithms achieve the performance improvement compared to the conventional R-APA. In addition, the proposed method gives birth to an efficient and robust update scheme for the regularization parameter by introducing the normalized gradient. Through experiments we show that the proposed algorithms lead to faster convergence rate and lower misadjustment error.

REFERENCES

[1] B. Widrow and S. D. Sterns, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1985.
[2] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, 2002.
[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Englewood Cliffs, NJ: Prentice Hall, 2003.
[4] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electro. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19–27, 1984.
[5] H.-C. Shin and A. H. Sayed, "Mean-square peformance of a family of affine projection algorithms," *IEEE Trans. Signal Processing*, vol. 52, pp. 90–102, Jan. 2004.
[6] V. Myllylä and G. Schmidt, "Psedo-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.
[7] S. C. Douglas, "Generalized gradient adaptive step sizes for stochastic gradient adaptive filters," in *Proc. IEEE Int. Conf. on Accoustics, Speech, and Signal Processing, ICASSP'95*, vol. 2, pp. 1396–1399, 1995.