Convex Combination of Affine Projection Filters with Individual Regularization

Kwang-Hoon Kim[1], Young-Seok Choi[2], Seong-Eun Kim[1] and Woo-Jin Song[1]
[1]Dept. of Electrical and Computer Engineering, Phang University of Science and Technology
Pohang, Kyungbuk, 790-784, Korea
[2]Dept. of Biomedical Engineering, Johns Hopkins School of Medicine
Baltimore, MD 21205, USA
E-mail : wjsong@postech.ac.kr

Abstract: We propose a convex combination of affine projection algorithm (APA) filters with individual regularization. Two APA filters with different regularization are adapted independently in order to keep the advantages of both component filters. By using their own error signals, separate filters are independently adapted, while the combination is adjusted by means of a stochastic gradient rule so as to minimize the error of the overall filter. This novel scheme provides improvement of performance in term of the convergence rate and the steady-state error. Experimental results in channel identification show the validity of the proposed algorithm when compared to the APA.

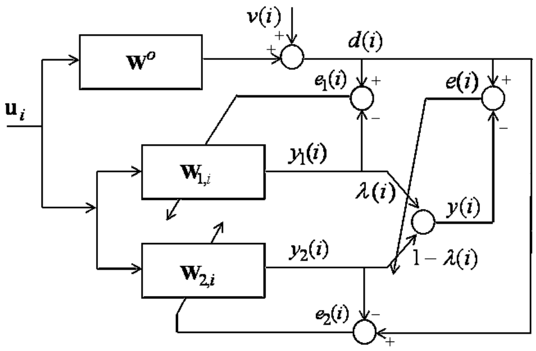

Figure 1. Convex combination of two adaptive filters

1. Introduction

Adaptive filtering algorithms have been widely used in various applications, such as echo cancelation, channel equalization and system identification. The least mean square (LMS) and the normalized LMS (NLMS) algorithm are most commonly used because of simple structure and easiness of implementation. However, its convergence rate is significantly reduced for correlated input signals [1]–[3]. To overcome this problem, the affine projection algorithm (APA) was proposed by Ozeki and Umdea [4]. While the LMS-type filters update the weights based only on the current input vectors, the APA updates the weights on the basis of the last K input vectors [4]–[6].

In the APA, the inversion of the rank-deficient matrix may give rise to the singularity. To avoid this situation, a positive constant δ called the *regularization parameter* is used [1],[2]. We use the regularized APA (R-APA) as opposed to the simple APA in order to emphasize the existence of the regularization parameter δ APA is reserved for the case $\delta = 0$. In addition, it is known that the regularization parameter also plays a critical role in the convergence performance of the APA [7]. The regularization parameter governs both the rate of convergence and the steady-state error. To compromise the trade-off between fast convergence and small steady-state error, the regularization parameter needs to be adjusted. Numerous schemes for controlling the regularization parameter have been proposed [7]–[10]. In [7], the optimal regularization parameter is adjusted so as to minimize the expected norm of the system mismatch vector. An earlier work for the NLMS with a time-varying regularization parameter has been presented [8], [9]. Also, the optimal regularization matrix which is an extension of a scalar regularization parameter is introduced for the R-APA [10]. In [11], a modified update for APA family, which includes an explicit regularization factor was proposed. This scheme optimized regularization to have maximum speed of convergence. Particularly, the explicit regularization factor does not only help in dealing with numerical precision problems, but also allows robust behavior against all possible perturbations.

Recently, novel schemes having different philosophy for solving the trade-off problem in the adaptive filtering have been proposed [12]–[14]. These new approaches are based on convex combination of two adaptive filters as shown in Fig. 1. The combination is executed in a manner that the overall filter keep the advantages of both component filters. The individual filters are independently adapted using their own error signals, while the combination is adapted by means of a stochastic gradient algorithm in order to minimize the error of the overall structure.

In this paper, we present an improved R-APA with an adaptive combination of one fast (i.e., with a small regularization parameter) and one slow (with a large regularization parameter) AP filter. It is effective for combining fast convergence and low steady-state error. The proposed combination of the APA (C-APA) dynamically adjusts the mixing parameter to combine two AP filters using a stochastic gradient method. The resulting scheme is able to retain the best properties of each component filter, that is, fast convergence and small steady-state error. This paper is organized as follows. In section 2, we introduce the conventional R-APA and convex combination of two R-APA filters. Section 3 illustrates the experimental results and section 4 concludes this paper.

2. Convex Combination of R-APA filters

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

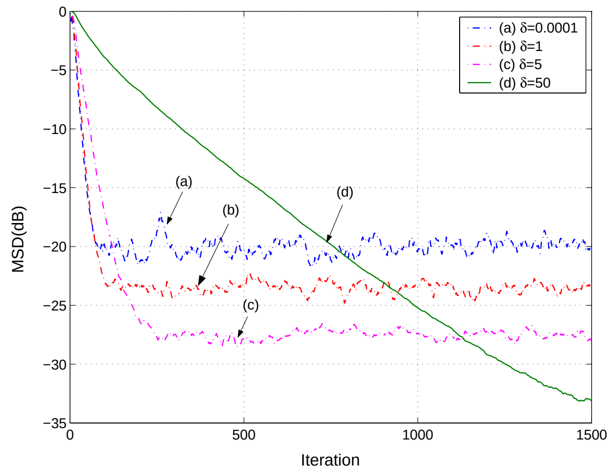

Figure 2. Plots of Mean-Square Deviation (MSD) curves for the R-APA with different regularization parameter

identification model

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

where $\mathbf{w^o}$ 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 a row input vectors as follows:

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

where $\mathbf{u}_i$ and $d(i)$ can be expanded in the form of matrix:

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

2.1 Conventional R-APA

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

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

where $\mathbf{e}_i = \mathbf{d}_i - \mathbf{U}_i \mathbf{w}_{i-1}$, and μ is the step size, δ is the regularization parameter, and * denotes the Hermitian transpose. The regularization parameter is not employed only to avoid the inversion of the rank-deficient matrix $\mathbf{U}_i \mathbf{U}_i^*$, but also to play a critical role in the convergence performance of the R-APA. As shown in Fig. 2, a small regularization parameter will ensure a large effective step-size and thus R-APA converges fast but results in a large steady-state error. On the other hand, a large regularization parameter will yield a small effective step-size and thus the R-APA results in a small steady-state error in the steady-state, but converges slowly. According to the discussions above, we may expect performance improvement through solving this trade-off problem utilizing the convex combination with individual regularization.

2.2 Adaptive combination with Individual Regularization

What we intend to is improving the convergence performance by using an adaptive convex combination of two R-APA filters, the first being a fast filter (i.e., with a small regularization parameter) and the second a slow filter (with a large regularization parameter). The output signals and the output errors of both filters are combined to obtain advantage of each filter: the rapid convergence from the fast filter and the reduced steady-state error from the slow filter. We assume the first filter as a fast filter.

Let's begin with the following adaptive convex combination of two R-APA scheme, which obtains the output of the overall filter as

$$y(i) = \lambda(i) y_1(i) + [1 - \lambda(i)] y_2(i), \tag{4}$$

where $y_1(i)$ and $y_2(i)$ are the outputs of two R-APA filters at time i, $y_k(i) = \mathbf{u}_i \mathbf{w}_{k,i-1}, k = 1, 2$ with $\mathbf{w}_{k,i-1}$ being the adaptive filter coefficient vectors of the k-th filter and $\mathbf{u}_i$ is the input vector and $\lambda(i) \in [0, 1]$ is a mixing parameter. The mixing parameter $\lambda(i)$ is adjusted to retain the advantages of both component filters. For simplicity, we assume that both $\mathbf{w}_{1,i-1}$ and $\mathbf{w}_{2,i-1}$ have the same length, so that the overall filter is obtained by

$$\mathbf{w}_i = \lambda(i) \mathbf{w}_{1,i} + [1 - \lambda(i)] \mathbf{w}_{2,i}. \tag{5}$$

In the structure of convex combination, two adaptive filters are adapted separately. Therefore, both R-APA filters are operated independently using their own output as

$$\mathbf{w}_{k,i} = \mathbf{w}_{k,i} + \mu \mathbf{U}_i^* \left(\mathbf{U}_i \mathbf{U}_i^* + \delta_k \mathbf{I}\right)^{-1} \mathbf{e}_{k,i}, \tag{6}$$

where $\mathbf{e}_{k,i} = \mathbf{d}_i - \mathbf{U}_i \mathbf{w}_{k,i-1}, \quad k = 1, 2.$

In order to determine the mixing parameter $\lambda(i)$, we incorporated a stochastic gradient rule so that $J(i) = e^2(i)$ is minimized, where $e(i)$=$d(i)$−$\mathbf{u}_i \mathbf{w}_{i-1}$. However, rather than adapting $\lambda(i)$ directly, we will update a new parameter $a(i)$ which defines $\lambda(i)$ through the sigmoidal function as

$$\lambda(i) = sgm[a(i)] = (1 + e^{-a(i)})^{-1}, \tag{7}$$

Accordingly, the update equation for $a(i)$ is given by

$$\begin{aligned} a(i) &= a(i-1) - \frac{\mu_a}{2} \frac{\partial e^2(i)}{\partial a(i-1)} \\ &= a(i-1) + \mu_a e(i) [y_1(i-1) - y_2(i-1)] \\ &\quad \cdot \lambda(i-1) [1 - \lambda(i-1)], \end{aligned} \tag{8}$$

where μ_a is the step size for the adaptation of $a(i)$ and we should choose $a(i)$ to a very high value so that the combination can progress even faster than the fastest component filter. By introducing the sigmoid function, we can easily guarantee $\lambda(i) \in [0, 1]$. In addition, the adaptation rule of $a(i)$ reduces both the stochastic gradient noise and the adaptation speed near $\lambda(i) = 0$ and $\lambda(i) = 1$ when the combination is expected to perform close to on of its component filters without degradation, while allowing a fast adaptation for intermediate values of $\lambda(i)$[14]. In the practical purpose, the adaptation

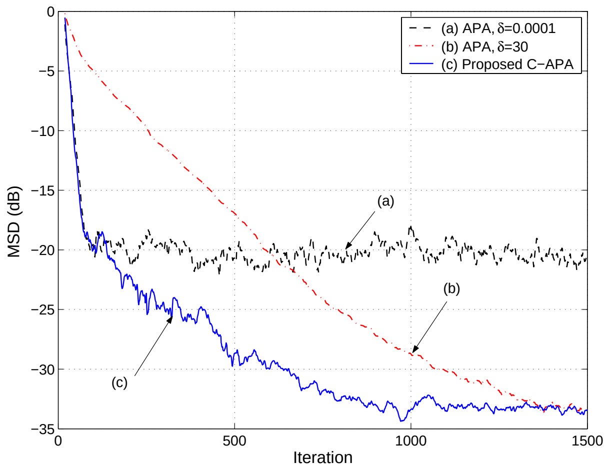

Figure 3. Mean Square Deviation (MSD) curves for the R-APA with different regularization parameters and the proposed C-APA

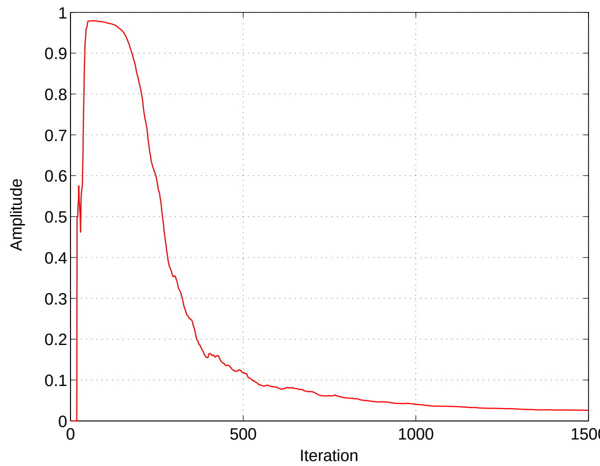

Figure 4. Plot of the mixing parameter $\lambda(i)$ in the proposed C-APA

parameter $a(i)$ is limited to the interval $[-4, 4]$, to prevent the algorithm to stop when $\lambda(i)$ or $1-\lambda(i)$ is too close 0 or 1 [13].

To improve the performance of the original convex combination algorithm, two modification schemes have been proposed [13]. One of the modification is to take advantage of the fast filter to speed up the convergence of the slow filter. This can be done by transferring a portion of weights in fast filter $\mathbf{w}_1$ to weights in slow filter $\mathbf{w}_2$ and is given by

$$\mathbf{w}_2(i) = \alpha \mathbf{w}_2(i) + (1-\alpha)\mathbf{w}_1(i), \tag{9}$$

where α is a parameter close to 1.

Another modification is to improve the convergence of the mixing coefficient $a(i)$. It is obvious that when the difference between $y_1(i-1)$ and $y_2(i-1)$ is very small, the adaptation performance of $a(i)$ is deteriorated. We can alleviate this problem by inserting the momentum term as the follows

$$\begin{aligned} a(i) = a(i-1) &+ \mu_a e(i)[y_1(i-1) - y_2(i-1)] \\ &\cdot \lambda(i-1)[1-\lambda(i-1)] + \rho[a(i) - a(i-1)], \end{aligned} \tag{10}$$

where α is a positive constant. In general, 0.5 is a good choice for speeding up the adaptation of the mixing coefficient. Simulations will be performed to support performance improvement by these two modification schemes.

The key point of the adaptive convex combination of adaptive filter is to adjust the overall filter by the mixing parameter $\lambda(i)$ in accordance with the performance of the two individual filter. The issue of how to optimally combine the individual filters is a challenging one. In summary, the proposed C-APA with individual regularization is operated by following mechanism. In situations where a fast convergence speed would be desirable, the mixing coefficient $a(i)$ makes $\lambda(i)$ evolve towards 1. On the contrary, in stationary period, the slow APA filter shows better performance, which results in $\lambda(i)$ get close to 0.

3. Experimental Results

We demonstrate the performance of the proposed algorithm by carrying out computer experiments 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. The input signal u_i is obtained by filtering a white, zero-mean,gaussian random sequence through a first-order system

$$G(z) = \frac{1}{(1-0.9z^{-1})}. \tag{11}$$

The signal-to-noise ratio (SNR) is calculated by

$$\mathbf{SNR} = 10\mathbf{log}_{10}\left(\frac{E[y^2(i)]}{E[v^2(i)]}\right), \tag{12}$$

where $y(i) = \mathbf{u}_i\mathbf{w}^\circ$. The measurement noise $v(i)$ is added to $y(i)$ such that SNR $= 30$dB. We use the mean square deviation (MSD) between the real channel and the adaptive filter to measure the performance of the proposed algorithm. The MSD, $E\|\mathbf{w}^\circ - \mathbf{w}_i\|^2$, is taken and averaged over 100 independent trials. The order of the APA is set to $K = 4$. The step-size is set as $\mu = 0.5$. To adapt the combination, we use $\mu_a = 100$. And $\alpha = 0.9$ and $\rho = 0.5$ are used. Finally, the mixing coefficient $a(i)$ has been initialized to zero.

Fig. 3 shows the MSD curves of the R-APA and the proposed C-APA. Dashed lines indicate the results of the R-APA with $\delta = 0.0001$ and $\delta = 30$, respectively. As can be seen, the proposed C-APA inherits the best properties of each component filter, having the fast convergence speed and small steady-state error comparable to R-APA with $\delta = 0.0001$ and $\delta = 30$, respectively. With the proposed scheme, the convergence speed and the steady-state error always dictated by the R-APA with small regularization parameter and R-APA with large regularization parameter, respectively.

In Fig. 4, the evolution of the mixing parameter $\lambda(i)$ is shown. This curve is also obtained over 100 independent trials. From this result,we can see that the parameter $\lambda(i)$

increases toward unity initially, to provide the fast convergence speed for the overall filter and then converges to a small value to obtain a small steady-state error from the slow filter. The evolution of mixing parameter evidently matches the demands of the convex combination.

4. Conclusion

Adaptive convex combination scheme is an interesting approach to solve the trade-off problem in the area of the adaptive filtering. As a way to break the convergence speed vs steady-state error compromise inherent to R-APA filters, we have presented a convex combination of the R-APA filters with individual regularization. The proposed algorithm carries out the update by adjusting the mixing parameter dynamically to get an overall output of improved quality. Experimental results show the improved convergence performance with a fast convergence speed and a small steady-state error compared to the conventional R-APA.

5. Acknowledgement

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

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] E. Eweda, "Transient performance degradation of the LMS, RLS, sign, signed regressor, and sign-sign algorithms with data correlation," *IEEE Trans. Cricuits Syst. II*, Analog Digit. Signal Process, vol. 46, no. 8, pp. 1055–1062, Aug. 1999.

[4] 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.

[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] S. J. M. de Almeida, J. C. M. Bermudez, N. J. Bershad, and M. H. Costa, "A statistical analysis of the affine projection algorithm for unity step size and autoregressive inputs," *IEEE Trans. Circuits Syst. I*, Reg. Papers, vol. 52, no. 7, pp. 1394–1405, Jul. 2005.

[7] V. Myllylaä and G. Schmidt, "Pseudo-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.

[8] D. P. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Lett.*, vol. 11, No. 2, pp. 115–118, Feb. 2004.

[9] Y.-S. Choi, H.-C. Shin and W.-J. Song, "Robust regularization for normalized LMS algorithms," *IEEE Trans. Circuits Syst. II*, vol. 53, pp.627–631, Aug. 2006.

[10] Y.-S. Choi, H.-C. Shin and W.-J. Song, "Adaptive regularization matrix for affine projection algorithm," *IEEE Trans. Circuits Syst. II*, vol. 54, pp.1087–1091, Dec. 2007.

[11] H. Rey, L. R. Vega, S. Tressens, J. Benesty, "Variable explicit regularization in affine projection algorithm: robustness issues and optimal choice," *IEEE Trans. Signal Processing*, vol. 55, No. 5, pp. 2096–2109, May. 2007.

[12] M. Martínez-Ramón, J. Arenas-García, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "An adaptive combination of adaptive filters for plant identification," in *Proc. 14th Int.Conf. Digital Signal Processing*, Santorini, Greece, pp.1195–1198, 2002.

[13] J. Arenas-García, V. Gómez-Verdejo, and A. R. Figueiras-Vidal, "New algorithms for improved adaptive convex combination of LMS transversal filters," *IEEE Trans. Instrum. Meas.*, vol. 54, No. 6, pp.2239–2249, Dec. 2005.

[14] J. Arenas-García, V. Gómez-Verdejo, and A. R. Figueiras-Vidal, A. H. Sayed, "Mean-Square Performance of a Convex Combination of Two Adaptive Filters," *IEEE Trans. Signal Processing*, vol. 54, No. 3, pp.1078–1090, Mar. 2006.