A GENERALIZED PROPORTIONATE VARIABLE STEP-SIZE ALGORITHM FOR FAST CHANGING ACOUSTIC ENVIRONMENTS

Osamu Hoshuyama†, Rafik A. Goubran‡, and Akihiko Sugiyama†

†Multimedia Research Laboratories
NEC Corporation, JAPAN
houshu@ccm.cl.nec.co.jp

‡Department of Systems and Computer Engineering
Carleton University, CANADA
goubran@sce.carleton.ca

ABSTRACT

This paper proposes a generalized proportionate variable step-size algorithm based on Affine projection. It controls the step size for each tap individually based on the gradient approximated by the difference between the current coefficient and an averaged filter coefficient with delay. This step-size control is specifically effective for tracking fast changing acoustic environments. It is shown that P-NLMS family and the ES algorithm are special cases of the proposed algorithm in terms of step-size control. Simulation results in the context of echo cancellation in artificial and real environments demonstrate that this step-size control combined with the Affine projection algorithm (APA) reduces the echo by up to 5 dB compared to the standard APA. The best tracking performance, without *a priori* knowledge on the acoustic environment, among the same family of algorithms is obtained.

1. INTRODUCTION

In adaptive filters, it is desirable to achieve fast convergence and fast tracking simultaneously. Fast changing environment is common with adaptive filter applications, e.g. echo cancellers, microphone arrays and noise cancellers. Especially, acoustic echo cancellation for hands-free cellphones is one of the most challenging applications for adaptive filters. Changes in the echo path by hand and face movement are frequent and significant, which require the adaptive algorithm fast-tracking capability.

Although tracking capability is an important characteristic, it has been paid little attention to. Most of the literatures have investigated initial convergence characteristics. However, the nature of tracking is different from that of initial convergence [1]. It means that fast convergence algorithms such as recursive least squares algorithms (RLS) [1] and Affine projection algorithms (APA) [2] provide fast initial convergence, but may not be sufficiently fast in tracking. When adaptation algorithms, which exhibit similar initial convergence, are compared, significant differences in tracking may be observed.

For fast tracking, individual step-size algorithms are effective [3][4][5]. Makino et al. pointed out for echo cancellers that step-size control according to the statistics of the changes in the impulse response, such as standard deviation, speeds up the initial convergence [3]. This exponentially weighted step-size (ES) algorithm, named after the shape of the step-size envelope, has fairly good tracking capability[1] However, the statistics of the changes are different for each acoustic environment and are time-variant in mobile systems like cellphones. Therefore, ES algorithm is not applicable to a wide range of acoustic environment without advance measurement.

Adaptive individual step-size algorithms [4][5] do not have this drawback. Duttweiler's proportionate normalized least mean square algorithm (P-NLMS), originally proposed for network echo cancellation[4], uses the absolute value of the filter coefficient as the step size at each tap. An improved P-NLMS algorithm (IP-NLMS), proposed by Benesty et al., introduces normalization and a constant for the purpose of step-size control. Because a tap with a large filter coefficient tends to vary more significantly than that with a small coefficient, the step size proportional to the filter coefficient provides fast tracking. However, both time-variant and time-invariant components in the actual acoustic impulse responses are taken into account. When the time-invariant components is not negligible compared to the time-variant ones, the time-variant components, which is needed for fast tracking, are contaminated by the other. Therefore, desirable step-size control can not be achieved.

This paper proposes a generalized proportionate variable step-size algorithm that includes P-NLMS, IP-NLMS and ES algorithm as special cases. The proposed algorithm controls the step size at each tap based on the difference between the current coefficient and an averaged filter coefficient with delay. This difference can be considered as an estimate of the gradient, which has no time invariant components, leading to faster tracking speed.

2. IMPROVED PROPORTIONATE AFFINE PROJECTION ALGORITHM

For colored signals such as speech , the Affine projection algorithm (APA) [2], proposed by Ozeki et al., is effective for coefficient adaptation of an adaptive filter. Therefore, an APA with a relatively high order is assumed as the base algorithm. It is also natural that P-NLMS and IP-NLMS are combined with the APA (P-APA and IP-APA) in place of the NLMS algorithm.

An echo canceller is mathematically described as:

$$e(k) = d(k) - y(k), \quad (1)$$
$$y(k) = W^T(k)X(k), \quad (2)$$
$$W(k) \equiv [w_0(k), w_1(k), ..., w_{N-1}(k)]^T, \quad (3)$$
$$X(k) \equiv [x(k), x(k-1), ..., x(k-N+1)]^T, \quad (4)$$

where k is the time index, $e(k)$ is the error, $d(k)$ is the near-end signal obtained at the microphone, $y(k)$ is the output signal of the adaptive filter, $W(k)$ is the coefficient vector consisting of N filter coefficients $w_n(k)$, and $X(k)$ is the input signal vector consisting of the far-end signal samples $x(k-n)$.

[1]Tracking capability of ES algorithm will be presented in Section 4.

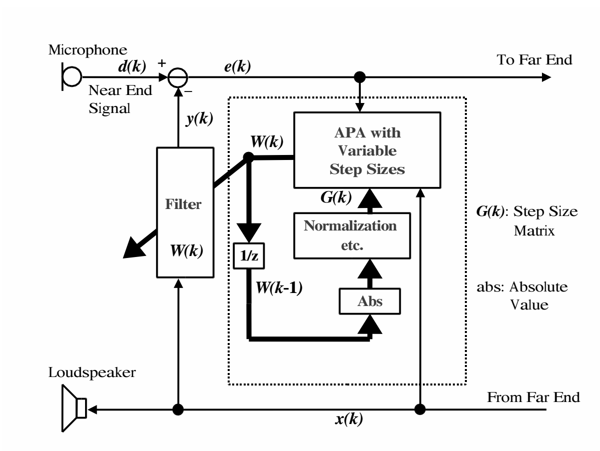

Fig. 1. Structure of Improved Proportionate Algorithm

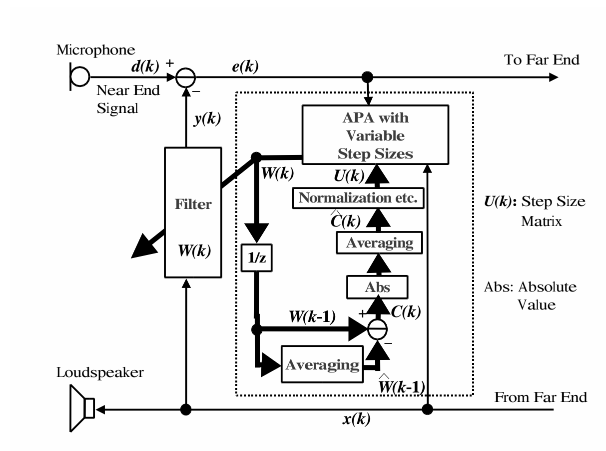

Fig. 2. Structure of Generalizes Gradient Proportionate Algorithm

The p-th order APA with variable individual step sizes is expressed as follows [6]:

$$
\begin{aligned}
W(k+1) &= W(k) \\
&\quad +\alpha\, e(k)\, A(k)\, Q(k)\, [Q_p^T(k) A(k) Q_p(k) + \delta I]^{-1}, && (5)\\
Q_p(k) &= [X(k)X(k-1)...X(k-p+1)], && (6)\\
A(k) &= \text{diag}[a_0(k), a_1(k), ..., a_{N-1}(k)], && (7)
\end{aligned}
$$

where α is the global step size which dominates convergence, tracking speed, and the final error. A small positive number δ, and the identity matrix I are for regularization. diag$[\cdots]$ is the diagonal matrix consisting of the components in the brackets. The diagonal matrix $A(k)$ is a time-varying step-size matrix, which consists of independent step sizes $a_n(k)$ corresponding to each tap.

IP-APA, an APA combined with the step-size control method of IP-NLMS, introduces normalization and a constant factor. Thanks to these modifications, improved stability and convergence speed both in the initial convergence and tracking are obtained[2]. The structure of IP-APA is shown in Fig. 1. The step-size matrix $A(k)$ for P-APA is replaced with a new time-varying matrix $G(k)$ in IP-APA. It is described as follows:

$$
\begin{aligned}
A(k) &= G(k) \equiv \text{diag}[g_0(k), g_1(k), ..., g_{N-1}(k)], && (8)\\
g_n(k) &= \frac{(1-\beta)}{2} \frac{\text{abs}\{w_n(k-1)\}}{\sum_{n=0}^{N} \text{abs}\{w_n(k-1)\}} + \frac{\beta}{2N}, && (9)\\
&\quad (n = 0, 1, ..., N-1)
\end{aligned}
$$

where $G(k)$ is the time-varying step-size matrix, and abs$\{\cdot\}$ is an operator to take the absolute value of the argument. Parameter β serves as a constant factor in each step size. When β is 0, IP-APA reduces to P-APA, and for a setting of $\beta = 1$, IP-APA is equivalent to the standard APA.

The step size is basically proportional to the impulse response coefficient, making tracking faster, because a tap with a large coefficient tends to vary more significantly than that with a small coefficient. However, both time-variant and time-invariant components in the actual acoustic impulse responses are taken into account for step-size control. When the time-invariant components are not negligible compared to the time-variant ones, the time-variant components, which are needed for fast tracking, are contaminated by the other. Therefore, desirable step-size control can not be achieved.

[2]Tracking capability of IP-APA will be presented in Section 4.

3. PROPOSED ALGORITHM: GENERALIZED GRADIENT PROPORTIONATE ALGORITHM

The proposed algorithm shown in Fig. 2 controls the step size for each tap individually based on the gradient approximated by the difference between the current coefficients and a delayed filter coefficients. The gradient has no time-invariant components, therefore, it leads to faster tracking. To obtain good and stable estimate of the gradient, long-term average using infinite impulse response filters is used. The absolute value of the estimated gradient is further averaged for stable step-size control. The proposed algorithm is expressed as follows:

$$
\begin{aligned}
A(k) &= U(k) \equiv \text{diag}[u_0(k), u_1(k), ..., u_{N-1}(k)], && (10)\\
u_n(k) &= \frac{(1-\beta)}{2} \frac{\hat{c}_n(k-1)}{\sum_{n=0}^{N} \hat{c}_n(k-1)} + \frac{\beta}{2N}, && (11)\\
\hat{c}_n(k) &= \epsilon\, \hat{c}_n(k-1) + (1-\epsilon)\, \text{abs}\{c_n(k)\}, && (12)\\
c_n(k) &= w_n(k-1) - \gamma\, \hat{w}_n(k-1), && (13)\\
\hat{w}_n(k) &= \eta\, \hat{w}_n(k-1) + (1-\eta)\, w_n(k-1), && (14)\\
&\quad (n = 0, 1, ..., N-1)
\end{aligned}
$$

where $U(k)$ is a step-size matrix corresponding to $A(k)$ and $G(k)$. $c_n(k)$ and $\hat{c}_n(k)$ represent a gradient estimate and its average. $\hat{w}_n(k)$ is the averaged tap coefficient, and γ is a correction factor for the averaged coefficient, $\hat{w}_n(k-1)$. The forgetting factors η and ϵ are for calculation of $\hat{w}_n(k)$ and $\hat{c}_n(k)$, respectively. Due to the group delay introduced in (14), the averaged tap coefficient $\hat{w}_n(k)$ is a delayed version of the current coefficient $w_n(k)$. Therefore, $c_n(k)$, the difference between $\hat{w}_n(k)$ and $w_n(k)$, can be considered as an estimate of the gradient.

Comparing (11)-(14) with (9), the absolute coefficient in (9) is replaced with the averaged estimate of the gradient. The averaged estimate, $\hat{c}_n(k)$, of the gradient is calculated from the difference between the current coefficient $w_n(k-1)$, and its modified average $\gamma\, \hat{w}_n(k-1)$. Because $\hat{w}_n(k-1)$ represents the time-invariant component of the coefficient, which approximates the coefficient itself, $\hat{c}_n(k)$ is the remaining time-variant component. Therefore, the proposed step-size control method can be viewed as an IP-APA taking mainly time-variant components into account. This term dominated by time-variant components enables fast tracking capability in fast changing environments. It is reasonable that a step size proportional to the absolute gradient improves tracking capability.

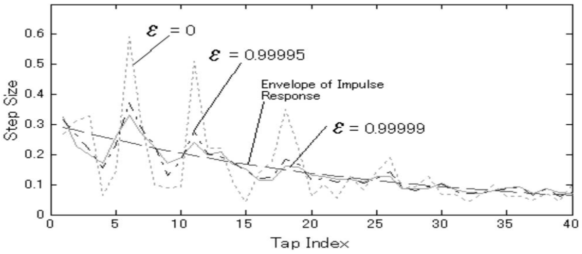

Fig. 3. Step Sizes for Different Averaging Coefficients. (Magnified)

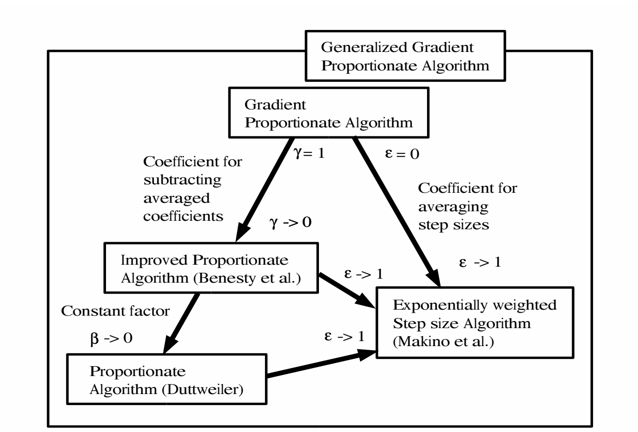

Fig. 4. Family Tree of Generalized Gradient Proportionate Algorithm.

3.1. Interpretation of the Proposed Algorithm

The proposed algorithm, IP-APA, and P-APA are all similar in the step-size control in (9) and (11). It is worthwhile to investigate the relationships among these algorithms. As is discussed later, the proposed algorithm expressed by (10)-(14) includes IP-APA, P-APA, and also ES algorithm. Because it controls the step size in proportion to the gradient and is a generalized version of the other algorithms, it will be called hereinafter a generalized gradient proportionate Affine projection algorithm (GGP-APA).

When the forgetting factor ϵ is set to 0 in GGP-APA, it is a subset of the generalized version. Let us call it a gradient proportionate APA (GP-APA). When γ is equal to 0, GP-APA reduces to IP-APA. Therefore IP-APA can be considered as a simplified GP-APA more suitable for fast initial convergence than for tracking. Further setting of β to 0 clearly leads to P-APA.

As ϵ approaches to 1, the absolute difference between $\hat{w}_n(k)$ and $w_n(k)$ is accumulated as $\hat{c}_n(k)$ for a long time and the accumulated difference becomes a statistics value of the filter coefficient as in Fig. 3. If the difference $c_n(k)$ is considered as a statistics value of impulse response changes, GGP-APA becomes ES algorithm.

As a result, the GGP-APA represents the whole family consisting of GP-APA, IP-APA, P-APA, and ES algorithm. A family tree of these algorithms is illustrated in Fig. 4

3.2. Required Computations

Additional computations by GGP-APA are not significant compared to those by the high-order APA implemented in its original form. Different from the original APA, the high-order APA with variable and individual step-size control has no fast computation algorithm [6]. Therefore, GGP-APA can be implemented with a similar number of computations as those for IP-APA that is different only in the step-size control.

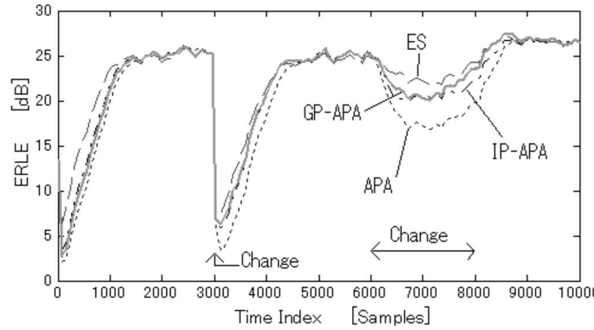

Fig. 5. Averaged ERLE for White Noise in Artificially Generated Changing Environment.

4. SIMULATIONS

Simulations to evaluate some family members of GGP-APA were performed assuming acoustic echo cancellation for hands-free cellphones, using data in artificial environment and those recorded with a hands-free cellphone in a fast-changing real environment.

4.1. Artificial Environment

Data in changing environment, artificially generated by computers, were used to see how the algorithms behave under controlled environment. The standard APA, IP-APA, ES-APA, and GP-APA were compared. A 128-tap impulse response of the echo path was changed from the initial value to another, both of which had been generated from white Gaussian signals multiplied by an exponential envelope. The two responses were interpolated linearly during the changing period. This interpolation imitates a gradual change in the echo path. The far end signal was generated from white Gaussian noise.

The number of taps of the adaptive filters N was 128. Signal to noise ratio was about 28dB for all over the sequence. For GP-APA, $\eta = 0.9999$, $\gamma = 1.0$, and $\epsilon = 0$. For all the algorithms, $\alpha = 1.0$, $\delta = 0.00001$, and $\beta = 0.5$. The order of APA, P was 8 because it showed the fastest tracking for the speech in real environments.

Figure 5 shows ERLEs (echo return loss enhancement) for a white noise sequence. In the initial part of the data sequence (from 0-th sample to 1000-th sample), all the algorithms converged very fast and there is almost no difference except for ES algorithm. ES algorithm showed the fastest convergence, because the envelope of impulse responses perfectly matched the step-size matrix of ES algorithm. At 3000-th sample, there was an abrupt change in the echo path. IP-APA and GP-APA controlled their step sizes properly, which resulted fast tracking. The differences in the algorithms are not significant for such an extreme environment change.

From 6000-th sample to 8000-th sample, there was a gradual change in the echo path. This condition implements an actual change in the acoustic environment when 8 kHz sampling is assumed. The standard APA is not good at tracking and the ERLE is degraded by 9 dB in the echo path change. The worst ERLE

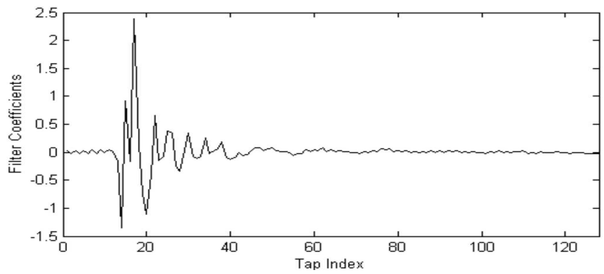

Fig. 6. Typical Impulse Response of the Echo Path.

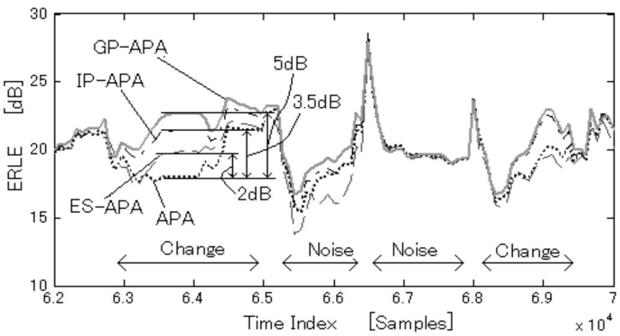

Fig. 7. ERLE for Speech in Real Environment

during the change is an important factor for the user's subjective impression. With IP-APA and GP-APA, the degradation of ERLE is significantly smaller than that with the standard APA. Over 4 dB improvement was observed at the worst ERLE for each change. Though the echo paths before and after the change do not contain time-invariant factors, GP-APA has 1 dB advantage over IP-APA.

4.2. Real Environment

To evaluate the tracking capability in the real environment, a cellphone mockup equipped with a loudspeaker was used to collect data. A typical impulse response of the echo path is depicted in Fig. 6. As the far-end signal, a male speech was used. During the data recording, a hand was moving in front of the loudspeaker at about 1 Hz to imitate actual echo path change by talker's behaviors.

Sampling rate was 8 kHz and the number of taps of filters N was 128. All the parameters for the algorithms are the same as in Section 4.1.

Figure 7 shows ERLEs for the speech. Due to nonlinearity of the echo path and ambient noise, the highest ERLE was limited to less than 28 dB. Improvement by ES algorithm is unstable because the envelope of the impulse response was not always exponential. From 6.33×10^4-th to 6.5×10^4-th samples, IP-APA and GP-APA show much faster tracking. Especially, GP-APA is the fastest during the path change, when only tone modulation in the echo was audible. Please note that there was almost no change in the echo level. GP-APA exhibits 5 dB higher ERLE than that of the standard APA in the path change.

Figure 8 shows the value of the step sizes at the 6.33×10^4-th sample in Fig. 7 for GP-APA, IP-APA and ES algorithm. It also depicts the absolute coefficient differences between the points at 6.33×10^4-th and 6.5×10^4-th samples. The step sizes are normalized so that they are almost the same among the three algorithms from 0-th to 20-th tap indexes on average. The step sizes for GP-APA models the actual difference of the filter coefficients in the best way. This fact leads to faster tracking capability.

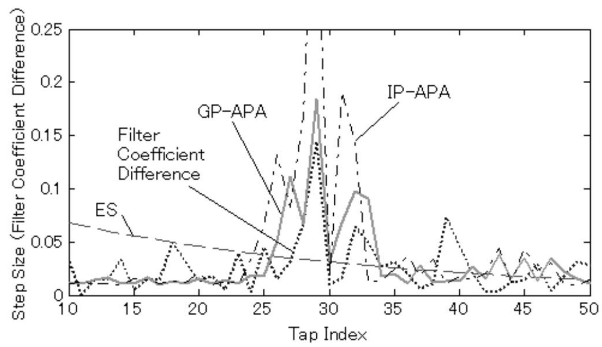

Fig. 8. Step Sizes and Impulse Response Difference. (Magnified)

5. CONCLUSION

A generalized proportionate variable step-size algorithm based on APA has been proposed. It controls the step size for each tap individually based on the gradient approximated by the difference between the current coefficient and an averaged filter coefficient with delay. This step-size control is specifically effective for tracking fast changing acoustic environments. It has been shown that P-NLMS family and the ES algorithm are special cases of the proposed algorithm in terms of step-size control. Simulation results in the context of echo cancellation in artificial and real environments have demonstrated that this step-size control combined with the APA reduces the echo by up to 5 dB compared to the standard APA. The best tracking performance among the same family of algorithms was obtained without *a priori* knowledge on the acoustic environment. Reduction of computations is left for a future work.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Prentice Hall, 1996.

[2] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an Affine subspace and its property," Electronics and Communications in Japan, vol.67-A, pp.19–27, 1984.

[3] S. Makino, et al. "Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response," IEEE Trans. on Speech and Audio Processing, vol. 1, pp. 101-108, Jan. 1993.

[4] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," IEEE Trans. on Speech and Audio Processing, vol. 8, pp. 508-518, Sept. 2000.

[5] J. Benesty and S.L. Gay, "An improved PNLMS algorithm," IEEE ICASSP-02, vol.2, pp. 1881-1884, May 2002.

[6] J. Benesty et al. *Advances in Network and Acoustic Echo Cancellation*, Springer, 2001.