CONVERGENCE ANALYSIS RESULTS FOR THE CLASS OF AFFINE PROJECTION ALGORITHMS

Sundar G. Sankaran and A. A. (Louis) Beex

Systems Group – DSP Research Laboratory
The Bradley Department of Electrical and Computer Engineering
Virginia Tech, Blacksburg VA 24061 – 0111.

Abstract

Over the last decade, a class of equivalent algorithms called the affine projection class of algorithms, which accelerate the convergence of the normalized LMS (NLMS) algorithm, has been discovered independently. The APA algorithms update weight estimates on the basis of multiple input signal vectors. In this paper, we present the results of the convergence analysis of the APA class of algorithms using a simple model for the input signal vectors. Conditions for convergence of the algorithms are presented. The convergence rate of APA is exponential, and it improves as the number of input signal vectors used for adaptation is increased. However, the rate of improvement in performance (time-to-steady-state) diminishes as the number of input signal vectors increases. For a given convergence rate, APA algorithms exhibit less misadjustment (steady state error) than NLMS. Simulation results are provided to corroborate the analytical results.

I. Introduction

Over the last decade, a class of equivalent algorithms such as the Affine Projection Algorithm (APA), the Partial Rank Algorithm (PRA), the Generalized Optimal Block Algorithm (GOBA), and NLMS with Orthogonal Correction Factors (NLMS-OCF) has been developed to accelerate the convergence of the normalized LMS (NLMS) algorithm [2, 3], especially for colored inputs. The distinguishing characteristic of these algorithms, developed independently from different perspectives, is that they update the weights on the basis of multiple input signal vectors, while the NLMS algorithm updates the weights on the basis of a single input vector. In the sequel, we will refer to this entire class of algorithms as affine projection algorithms, since APA is the earliest and most popular algorithm in this class that inherits its name. However, the convergence results presented here are applicable to the entire class of affine projection algorithms.

In this discussion, we present the results of the convergence analysis of the affine projection algorithms using a simple model for the input signal vector. In addition to the usual independence assumption [4], the angular orientation of the input vectors is assumed to be discrete. This assumption makes the convergence analysis tractable.

The weight update equation of APA is presented in Section II. Section III begins with a list of the assumptions that are used. While the details of the analysis are provided elsewhere [1], the main results of the analysis are summarized in Section III. Section IV corroborates the summarized results by comparing the analytical results with the results obtained from simulations. Concluding remarks are provided in Section V.

II. The Class of Affine Projection Algorithms

In system identification problems the system input x_n and corresponding measured output d_n, possibly contaminated with measurement noise ε_n, are known. The objective is to estimate an N dimensional weight vector $\mathbf{w}_n$, such that the estimated output $\hat{d}_n = \mathbf{w}_n^H \mathbf{x}_n$, where $\mathbf{x}_n = (x_n, x_{n-1}, \ldots, x_{n-N+1})^t$ is the input vector at the nth instant, is as close as possible to the measured output d_n in mean-squared error sense. The affine projection algorithms are iterative procedures to estimate these weights.

The APA class, as mentioned earlier, updates the weights on the basis of multiple input vectors. We use the weight update equation of the NLMS-OCF algorithm [3] for our discussions, since it is more general than in the other algorithms of this family and since the NLMS-OCF update equation is conducive to the analysis. The adaptive filter weights are updated by APA as shown below:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_0 \mathbf{x}_n + \mu_1 \mathbf{x}_n^1 + \ldots + \mu_M \mathbf{x}_n^M \tag{1}$$

where $M+1$ is the number of input vectors used for adaptation, $\mathbf{x}_n$ is the input vector at the nth instant, $\mathbf{x}_n^k$ (for $k = 1,2,\ldots,M$) is the component of $\mathbf{x}_{n-kD}$ that is orthogonal to $\mathbf{x}_n, \mathbf{x}_{n-D}, \mathbf{x}_{n-2D}, \ldots, \mathbf{x}_{n-(k-1)D}$, and μ_k (for $k = 0,1,\ldots,M$) is chosen as in (2).

$$\mu_k = \begin{cases} \dfrac{\overline{\mu} e_k^*}{\mathbf{x}_n^H \mathbf{x}_n} & for\ \ k = 0,\ if\ \|\mathbf{x}_n\| \neq 0 \\ \dfrac{\overline{\mu} e_n^{k*}}{\mathbf{x}_n^{kH} \mathbf{x}_n^k} & for\ k = 1,2,\ldots,M,\ if\ \|\mathbf{x}_n^k\| \neq 0 \\ 0 & otherwise \end{cases} \tag{2}$$

where

$$\begin{aligned} e_n &= d_n - \mathbf{w}_n^H \mathbf{x}_n, \\ e_n^k &= d_{n-kD} - \mathbf{w}_n^{kH} \mathbf{x}_{n-kD},\ for\ k = 1,2,\ldots,M,\ and \\ \mathbf{w}_n^k &= \mathbf{w}_n + \mu_0 \mathbf{x}_n + \mu_1 \mathbf{x}_n^1 + \ldots + \mu_{k-1} \mathbf{x}_n^{k-1}. \end{aligned} \tag{2}$$

The constant $\overline{\mu}$ is usually referred to as the step size.

The weight updates generated by APA and GOBA are equivalent to the special case of the weight updates generated by NLMS-OCF, shown in (1), with $D = 1$. PRA is the special case of APA where the APA weight adaptations are performed once

0-7803-5471-0/99/$10.00©1999 IEEE

every $M+1$ samples instead of every sample. In the next section we present the convergence behavior of (1) under certain simplifying assumptions.

III. Convergence Results

The convergence analysis is done based on the following assumptions on the signals and the underlying system:

(A1) The signal vectors $\{\mathbf{x}_n\}$ have zero mean, and are independent and identically distributed (i.i.d.) with covariance matrix

$$\mathbf{R} = E\left[\mathbf{x}_n \mathbf{x}_n^H\right] = \mathbf{V}\Lambda\mathbf{V}^H \tag{4}$$

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_N)$ and $\mathbf{V} = (\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_N)$.

Here, $\lambda_1, \lambda_2, ..., \lambda_N$ are the eigenvalues of $\mathbf{R}$ and $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N$ are the corresponding orthonormal eigenvectors $(\mathbf{V}^H\mathbf{V} = \mathbf{I})$. That is, $\mathbf{V}$ is a unitary matrix.

(A2) There exists a true adaptive filter weight $\mathbf{w}^0$ of dimension N such that the corresponding error signal

$$\begin{aligned} e_n &= d_n - w^{0H}\mathbf{x}_n \\ &\equiv \varepsilon_n \end{aligned} \tag{5}$$

inherits the properties of the measurement noise ε_n, which is a zero mean white noise of variance ξ^0 that is independent of $\{\mathbf{x}_n\}$.

(A3) The random vector $\mathbf{x}_n$ is the product of three independent random variables that are i.i.d. That is,

$$\mathbf{x}_n = s\, r\, \mathrm{V} \tag{6a}$$

$$\text{where} \begin{cases} P\{s = \pm 1\} = \dfrac{1}{2} \\ r \sim \|\mathbf{x}_n\| \\ P\{\mathrm{V} = \mathbf{v}_i\} = p_i = \dfrac{\lambda_i}{tr(R)}, \quad i = 1,2,...,N. \end{cases} \tag{6b}$$

where $r \sim \|\mathbf{x}_n\|$ means that r has the same distribution as the norm of the true input signal vectors.

Assumption (A3), first introduced by Slock [4], leads to a simple distribution for the vectors $\mathbf{x}_n$ consistent with the actual first- and second-order statistics of the input signal.

We define the weight error vector $\tilde{\mathbf{w}}_n$ as $\tilde{\mathbf{w}}_n = \mathbf{w}^0 - \mathbf{w}_n$. We say that the weights converge in the mean if the expectation of the weight-error vector $\tilde{\mathbf{w}}_n$ approaches zero as the number of iterations n approaches infinity. Convergence in the mean square means that the steady-state value of the covariance $\mathrm{cov}(\tilde{\mathbf{w}}_n)$ of the weight error vector is finite. If these two forms of convergence are satisfied, then the APA algorithm is said to be stable. To analyze the stability of the APA algorithm it suffices to know the time-update equation for the diagonal elements $\tilde{\lambda}_{n,i}$ of the transformed covariance matrix $\mathbf{V}^H \mathrm{cov}(\tilde{\mathbf{w}}_n)\mathbf{V}$. It can be shown that the desired update equation is given by [1]

$$\tilde{\lambda}_{n+1,i} = (1 - \alpha\beta_i)\tilde{\lambda}_{n,i} + \bar{\mu}^2 \xi^0 E\left(\frac{1}{r^2}\right)\beta_i \tag{7}$$

where $\alpha = \bar{\mu}(2 - \bar{\mu})$ and $\beta_i = 1 - (1 - p_i)^{M+1}$.

Further analysis leads to the following results [1]:

Result 1. $0 < \bar{\mu} < 2$ is a necessary and sufficient condition for the APA class to be stable.

Result 2. The steady-state (final) mean-squared error is given by

$$\xi_\infty = \xi^0\left[1 + \frac{\bar{\mu}}{2 - \bar{\mu}} E\left(\frac{1}{r^2}\right) tr(\mathbf{R})\right] \tag{8}$$

Result 3. The misadjustment, defined as the ratio of excess mean-squared error to minimum mean-squared error, equals

$$\mathrm{M} = \frac{\xi_\infty - \xi^0}{\xi^0} = \frac{\bar{\mu}}{2 - \bar{\mu}} E\left(\frac{1}{r^2}\right) tr(\mathbf{R}) \tag{9}$$

Note the independence of (9) of M. In fact, it is the same as the misadjustment of the NLMS algorithm (NLMS is the special case of APA with $M = 0$) with the same $\bar{\mu}$. The independence of (9) of M is, perhaps, due to certain simplifications made during the analysis. Simulation results indicate a "weak" dependence of misadjustment on M.

Result 4. Assume that no *a priori* information on the system is available and also that the optimal weights satisfy the maximum entropy assumption (that is, $\mathbf{w}^0$ has equal components along all eigenvectors of $\mathbf{R}$). Hence, the typical initial estimate for the weights, $\mathbf{w}_0 = \mathbf{0}$, is used. For optimal weights we use

$$\mathbf{w}^0 = \sqrt{\frac{\sigma_d^2 - \xi^0}{tr(\mathbf{R})}}\mathbf{V}1_N, \tag{10}$$

where σ_d^2 is the variance of the output signal d_n and $1_N \equiv [1 \quad 1 \quad \cdots \quad 1]^T$, which satisfies the maximum entropy assumption. Under these conditions, the mean-squared error ξ_n is given by

$$\xi_n = \sigma_d^2 \sum_{i=1}^{N} (1 - \alpha\beta_i)^n p_i. \tag{11}$$

This suggests that the mean-squared error converges exponentially.

Result 5. The convergence rates of APA are the same for step sizes $\bar{\mu}$ and $2 - \bar{\mu}$ for $\bar{\mu} \in (0,2)$. However the misadjustment increases as $\bar{\mu}$ increases. In view of this, it is better to use a step size $\bar{\mu} \in (0,1]$. Furthermore, $\bar{\mu} = 1$ is the optimum step size value for fastest convergence.

Result 6. For large enough n, the slope of the learning curve (plot showing mean-squared error in dB versus iteration number) depends linearly on M. Hence, the convergence rate improves as M increases. However, if we define the time to (reach) steady state, T_{SS}, as a performance index of the algorithm, the rate at which the performance improves diminishes as M increases.

Result 7. When the input is white, the mean-squared error in dB can be written as

$$\xi_{n,dB} \approx 10\log_{10} \sigma_d^2 - 4.343\frac{n\alpha(M+1)}{N} \quad (12)$$

Thus the learning curve for a white input is linear and the mean squared error drops by about 20 dB in $5N/(M+1)$ iterations for $\bar{\mu}=1$. This also means that longer filters exhibit slower convergence. Hence, the convergence rate can be improved by starting with a smaller number of taps in the adaptive filter and then gradually increasing the number of taps until the desired order is reached.

Result 8. APA provides a way to increase the convergence rate without compromising too much on misadjustment and, hence, the steady state mean-squared error of APA.

Corollary. NLMS is the special case of APA with M = 0. If M = 0, then $\beta_i = p_i$ and difference equation (7), which describes the behavior of $\tilde{\lambda}_{n,i}$, becomes

$$\tilde{\lambda}_{n+1,i} = (1-\alpha p_i)\tilde{\lambda}_{n,i} + \bar{\mu}^2 \xi^0 E\left(\frac{1}{r^2}\right)p_i. \quad (13)$$

Similarly the NLMS mean-squared error convergence behavior is given by

$$\xi_n = \sigma_d^2 \sum_{i=1}^{N} (1-\alpha p_i)^n p_i. \quad (14)$$

These results match the earlier results derived for NLMS under the same assumptions [4]. From Result 7, the learning curve of NLMS drops by 20 dB in about $5N$ iterations for $\bar{\mu}=1$. This result conforms to Rupp's observation on the convergence speed of NLMS [5].

A Special Comment for PRA. The partial rank algorithm attempts to reduce the complexity of APA by adapting the weights once every $M+1$ samples instead of every sample. Hence the above results can easily be "customized" for PRA.

IV. Verification Using Simulation

In this section, we demonstrate the validity of the analytical results presented in Section III. Simulation and theoretical results corresponding to three different types of signals, viz. white, reasonably colored, and highly colored, are shown. The reasonably and highly colored signals are generated as a Gaussian first-order autoregressive process with a pole at 0.25 and 0.95, respectively. The system to be identified has a 32-point long impulse response computed according to (10) for each case and hence the impulse response satisfies the maximum entropy assumption. The delay line of the adaptive filter is initialized with true data values (soft initialization) in all simulations and $\mathbf{w}_0 = \mathbf{0}$ is used as the initial estimate for the weights. The measurement noise is assumed to be absent, $\xi^0 = 0$, unless noted otherwise. The simulation results shown are obtained by ensemble averaging over 100 independent trials of the experiment.

Figure 1 shows the results obtained using a white input signal. The weight updates are performed with 11 input vectors (i.e., $M=10$). The steady-state MSE is limited in simulation to around –325 dB because of the quantization errors introduced in the calculations. We see that the theoretical result, as given by (11), is very close to the simulated result when D = 32, and that there is an appreciable deviation between the theoretical and simulated results when D = 1. This is because of the independence assumption that we used in the analysis. The input vectors used for a particular weight update are truly independent when D = 32, while this is not true when D = 1.

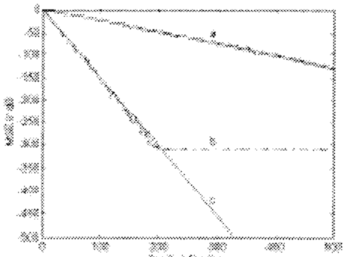

Figure 1. Learning Curves of APA for White Input Using $\bar{\mu}$=1.0 (a) Simulated with D=1, (b) Simulated with D=32, and (c) Theoretical.

The results obtained using the reasonably colored signal as input are shown in Figure 2. The simulation result is closer to the theoretical result when D = 32 than when D = 1, since the input vectors used for weight updates are more nearly independent when D = 32 than when D = 1.

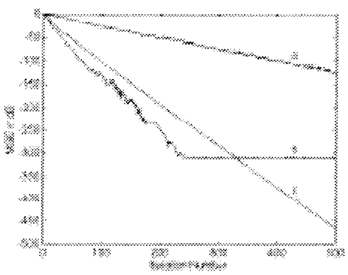

Figure 2. Learning Curves of APA for Reasonably Colored Input Using $\bar{\mu}$=1.0 (a) Simulated with D=1, (b) Simulated with D=32, and (c) Theoretical.

Results, for the highly colored signal as input, similar to the results shown in Figures 1 and 2, are shown in Figure 3. We see that there is a larger deviation between the theoretical and simulation results in this case than in the white noise and reasonably colored case. One would expect this behavior, since the highly correlated input violates the independence assumption more strongly than the other two inputs.

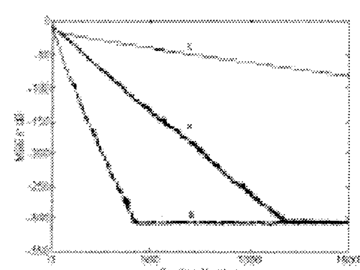

Figure 3. Learning Curves of APA for Highly Colored Input Using $\bar{\mu}$=1.0 (a) Simulated with D=1, (b) Simulated with D=32, and (c) Theoretical.

Next, we simulate the effect of varying the number of vectors (M+1) used for adaptation. The simulation results with white noise input, for different values of M, are shown in Figure 4. While for M = 0 the steady state is reached in about 1800 iterations, the steady state is reached for M = 2 and 8 in about 750 and 250 iterations respectively. Thus the improvement in time-to-steady-state T_{SS} achieved by increasing M from 2 to 8 is less than the improvement achieved by increasing M from 0 to 2. This confirms Result 6 – the T_{SS} improvement rate diminishes as M increases.

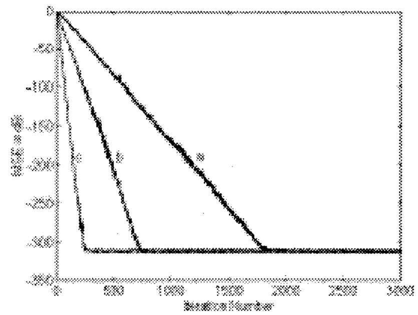

Figure 4. Simulated Learning Curves of APA for White Input – Various M (a) M=0 (NLMS), (b) M=2, and (c) M=8.

We use the result shown in Figure 4 to corroborate Result 7 as well. While the theoretical predictions for the slope of the learning curves (for white input) for M = 0, 2, and 8, using (12), are 0.14, 0.41, and 1.2 dB/iteration respectively, the corresponding slopes estimated from the simulation results are about 0.17, 0.42, and 1.3 dB/iteration respectively.

Result 8 suggested that APA provides a way to improve the convergence rate without compromising on misadjustment. The following experiment corroborates this observation. Figure 5a shows the learning curve of NLMS with a step size $\bar{\mu}$ of 0.25. We see that the algorithm takes about 8000 iterations to converge. The misadjustment $\mathcal{M}$ is 0.2062 for this case. An improvement in convergence can be achieved either by using a larger value of step size $\bar{\mu}$ or by using the affine projection algorithm (that is, by using more input vectors for the weight update). Figures 5b and 5c show the learning curves obtained by using NLMS with $\bar{\mu}$ = 1 and by using APA with M = 2 (and $\bar{\mu}$ = 0.25), respectively. In both these cases we see faster convergence than for NLMS with $\bar{\mu}$ = 0.25. It is evident that their individual convergence rates are nearly comparable, while the resulting misadjustments are quite different. NLMS with $\bar{\mu}$ = 1 has a misadjustment $\mathcal{M}$ of 1.1164 while APA with M = 2 has a misadjustment $\mathcal{M}$ of 0.2904. In other words, the steady-state error of APA with M = 2 is at least 2 dB less than the steady state error of NLMS with $\bar{\mu}$ = 1, while their convergence rates are comparable. This suggests that it would be better to use APA to get improved convergence than to use NLMS with large step size.

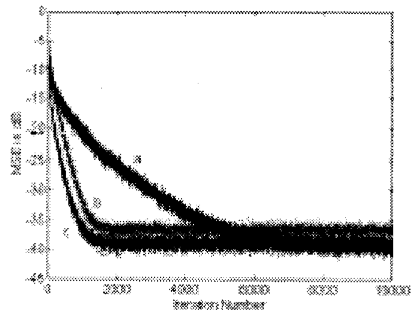

Figure 5. Simulated Learning Curves of APA - Misadjustment/Convergence Rate Trade-Off (a) M=0 (NLMS) and $\bar{\mu}$ =0.25, (b) M=0 (NLMS) and $\bar{\mu}$ =1.0, (c) M=2 and $\bar{\mu}$ =0.25.

V. Conclusion

The APA class of algorithms provides an improvement in convergence rate over NLMS, especially for colored input signals. A theoretical expression for the convergence behavior of the mean-squared error is given. As the signal color and/or step-sizes tend towards satisfying the independence assumption the simulated results tend to the theoretical results, while there is a mismatch otherwise. The convergence rate is exponential and it improves with an increase in the number of input signal vectors used for adaptation. However, the rate of improvement in terms of time-to-steady-state diminishes as the number of input vectors used for adaptation increases. The APA provides a way to increase the convergence rate without compromising too much on misadjustment. For white input the mean squared error drops by 20 dB in about $5N/(M+1)$ iterations, where N is the number of taps in the adaptive filter and M is the number of vectors used for adaptation. Simulation results corroborate our findings.

References

[1] S. G. Sankaran and A. A. (Louis) Beex, "Convergence Behavior of Affine Projection Algorithms," Submitted to *IEEE Transactions on Signal Processing*, March 1998.

[2] D. R. Morgan and S. G. Kratzer, "On a Class of Computationally Efficient, Rapidly Converging, Generalized NLMS Algorithms," *IEEE Signal Processing Letters*, vol. 3, no. 8, pp. 245 – 247, August 1996.

[3] S. G. Sankaran and A. A. (Louis) Beex, "Normalized LMS Algorithm with Orthogonal Correction Factors," *Proceedings of the Thirty First Asilomar Conference On Signals, Systems, and Computers*, Pacific Grove, CA, pp. 1670-1673, 2-5 November 1997.

[4] D. T. M. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2811 – 2825, September 1993.

[5] M. Rupp, "A Family of Adaptive Filter Algorithms with Decorrelating Properties," *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 771 – 775, September 1999.