TRACKING ANALYSIS RESULTS FOR NLMS AND APA

Sundar G. Sankaran[1] *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

We present the tracking properties of the Normalized LMS and affine projection class of algorithms for a randomly time-varying system under certain simplifying assumptions on the data. An expression is given for the steady-state mean-squared error. The dependence of the steady-state error and of the tracking properties on three user-selectable parameters, namely step size, number of vectors used for adaptation, and delay used to choose input vectors used for adaptation, is discussed. While the lag error depends on all of the above parameters, the fluctuation error depends only on stepsize. Increasing delay results in a linear increase in the lag error and hence the total steady-state mean-squared error. There is an optimum choice for stepsize and number of input vectors that minimizes the total mean-squared error. Simulation results are provided to corroborate the theoretical conclusions.

I. INTRODUCTION

Adaptive filtering is used in a wide range of applications including echo cancellation, noise cancellation, and equalization. In these applications, the environment in which the adaptive filter operates is often nonstationary. For satisfactory performance under nonstationary conditions, an adaptive filtering algorithm is required to follow the statistical variations of the environment. Tracking analysis provides insight into the ability of an adaptive filtering algorithm to track the changes in the surrounding environment. The tracking behavior of an algorithm is quite different from its convergence behavior. While convergence is a transient phenomenon, tracking is a steady-state phenomenon. An algorithm with good convergence properties does not necessarily track changes well.

Over the last decade a class of equivalent algorithms such as the Affine Projection Algorithm (APA), the Generalized Optimal Block Algorithm (GOBA), and Normalized LMS with Orthogonal Correction Factors (NLMS-OCF) has been developed to accelerate the convergence of NLMS [1, 2]. Mathematical analyses, as well as simulation results, have shown that the APA class of algorithms yields faster convergence than NLMS [1]. However, results on the tracking behavior of APA are not available in the existing literature. The objective of this paper is to analyze and to study the tracking behavior of APA and NLMS for a randomly time-varying channel. The tracking analysis is done based on certain simplifying assumptions such as the independence assumption and the discrete orientation assumption.

The paper is organized as follows. Section II provides the update equation of the APA class of algorithms. Section III begins with a list of assumptions used for the analysis. Some of the tracking properties of APA and NLMS are also discussed in Section III. Simulation results that support the derived theoretical results are presented in Section IV. A summary of results and concluding remarks are provided in Section V.

II. AFFINE PROJECTION ALGORITHM CLASS

Figure 1 shows an adaptive filter used in system identification mode. Here, 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 $\mathbf{x}_n$ is the input vector at the nth instant $\mathbf{x}_n = (x_n, x_{n-1}, \ldots, x_{n-N+1})^t$, 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.

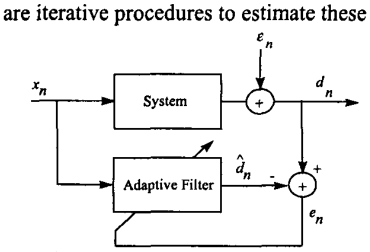

Figure 1. Adaptive Filtering for System Identification.

The APA class updates the weights on the basis of multiple input vectors. We use the weight update equation of the NLMS-OCF algorithm [2] 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 that follows. 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). The step size μ_k is a function of time n, even though it is not shown explicitly to avoid clutter.

[1] Currently with ArrayComm, Inc., San Jose, CA

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

where

$$\begin{aligned} e_n &= d_n - \mathbf{w}_n^H \mathbf{x}_n\ , \\ e_n^k &= d_{n-kD} - \mathbf{w}_n^{kH} \mathbf{x}_{n-kD},\ for\ k=1,2,\ldots,M,\ and \\ \mathbf{w}_n^k &= \mathbf{w}_n + \mu_0 \mathbf{x}_n + \mu_1 \mathbf{x}_n^1 + \ldots + \mu_{k-1}\mathbf{x}_n^{k-1}. \end{aligned} \tag{3}$$

The constant $\bar{\mu}$ 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 $D=1$. Furthermore, when $M=0$, NLMS-OCF reduces to conventional NLMS. In the next section we study the tracking behavior of (1) under certain simplifying assumptions.

III. TRACKING PROPERTIES

The tracking 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

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

where $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_N)$, containing the eigenvalues of $\mathbf{R}$, and $\mathbf{V} = (\mathbf{v}_1\ \ \mathbf{v}_2\ \ \ldots\ \ \mathbf{v}_N)$ is unitary $(\mathbf{V}^H\mathbf{V} = \mathbf{I})$, containing the corresponding orthonormal eigenvectors.

(A2) The measurement noise ε_n is a zero mean white noise of variance ξ^0 that is independent of the input $\{\mathbf{x}_n\}$.

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

$$\mathbf{x}_n = s\, r\, \mathbf{v} \tag{5a}$$

$$\text{where } \begin{cases} P\{s=\pm 1\} = \dfrac{1}{2} \\ r \sim \|\mathbf{x}_n\| \\ P\{\mathbf{v}=\mathbf{v}_i\} = p_i = \dfrac{\lambda_i}{tr(R)}, \quad i=1,2,\ldots,N. \end{cases} \tag{5b}$$

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

(A4) The underlying system is modeled as an FIR filter of length N with time-varying weights $\mathbf{w}_n^0$. That is,

$$d_n = \mathbf{w}_n^{0^H} \mathbf{x}_n + \varepsilon_n\,. \tag{6a}$$

The time variation of the weights follows a random-walk model as shown below.

$$\mathbf{w}_n^0 = \mathbf{w}_{n-1}^0 + \boldsymbol{\omega}_{n-1} \tag{6b}$$

where $\{\boldsymbol{\omega}_n\}$ is a stationary sequence of independent zero-mean N dimensional vectors with covariance matrix

$$E\left(\boldsymbol{\omega}_n \boldsymbol{\omega}_n^H\right) = \omega^0 \mathbf{I} \tag{6c}$$

with $\mathbf{I}$ an identity matrix of appropriate dimension. Furthermore, $\{\boldsymbol{\omega}_n\}$ is independent of both $\{\mathbf{x}_n\}$ and ε_n.

The discrete orientation assumption (A3), first introduced by Slock [3] and later used by Sankaran and Beex [4] for the convergence analysis of APA, renders the tracking analysis tractable. It can be shown that the first- and second-order statistics of the signal under the discrete orientation assumption are consistent with that of the actual signal.

Under these assumptions, the steady-state error is given by [5]

$$\xi_\infty = \lim_{n\to\infty} \xi_n = \xi^0\left[1 + \frac{\bar{\mu}}{2-\bar{\mu}} E\left(\frac{1}{r^2}\right) tr(\mathbf{R})\right] + \sum_{i=1}^{N} \frac{\omega^0 \lambda_i (1+\alpha D\gamma_i)}{\alpha\beta_i} \tag{7}$$

The above steady-state error consists of two distinct (decoupled) parts, namely the fluctuation error ξ_∞^f, caused by the measurement noise, and the lag error ξ_∞^l, caused by variations in the environment, defined as follows.

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

$$\xi_\infty^l = \sum_{i=1}^{N} \frac{\omega^0 \lambda_i (1+\alpha D\gamma_i)}{\alpha\beta_i} \tag{8b}$$

The lag error characterizes the tracking property of the algorithm. A large lag error indicates that the algorithm is slow in tracking the variations in the environment and a small lag error indicates that the algorithm is able to track the changes.

1. Dependence on Step size.

We begin by computing the derivative of the two components of the steady-state error, shown in (8), with respect to step size.

$$\frac{d}{d\bar{\mu}}\xi_\infty^f = \frac{2}{(2-\bar{\mu})^2}\xi^0 E\left(\frac{1}{r^2}\right) tr(\mathbf{R}) > 0 \quad \forall \bar{\mu} \in (0,2) \tag{9a}$$

$$\frac{d}{d\bar{\mu}}\xi_\infty^l = -\frac{2(1-\bar{\mu})}{\alpha^2}\omega^0 \sum_{i=1}^{N}\left(\frac{\lambda_i}{\beta_i}\right) \le 0 \quad \forall \bar{\mu} \in (0,1] \tag{9b}$$

$$\frac{d}{d\bar{\mu}}\xi_\infty^l = -\frac{2(1-\bar{\mu})}{\alpha^2}\omega^0 \sum_{i=1}^{N}\left(\frac{\lambda_i}{\beta_i}\right) > 0 \quad \forall \bar{\mu} \in (1,2) \tag{9c}$$

From (9a), we can conclude that the steady-state fluctuation error due to measurement noise increases as the step size $\bar{\mu}$ is increased. We deduce from (9b) and (9c) that the steady-state lag error due to variations in the environment decreases as $\bar{\mu}$ is increased if $\bar{\mu} \in (0,1]$ and it increases as $\bar{\mu}$ is increased if $\bar{\mu} \in (1,2)$. However, the fluctuation error keeps increasing as the step size is increased in the range $(0,2)$. Hence, we do not gain by using any step size in the range $(1,2)$ and it would be judicious to restrict step size to $(0,1]$. For this restricted range of step size, a larger value of step size results in better tracking, as well as in higher fluctuations due to measurement noise.

Since the total steady-state mean-squared error is the sum of two components--one increasing and the other decreasing, as $\bar{\mu}$ is increased--there is an optimum choice of $\bar{\mu}$, say $\bar{\mu}_{opt}$, for which the total error is minimized. This is illustrated in Figure 2.

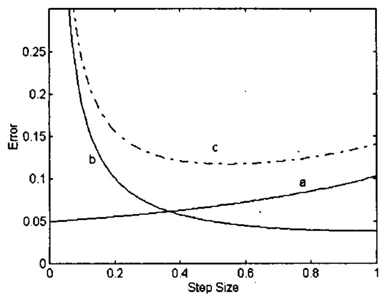

Figure 2. Mean-Squared Error for Different Step Sizes: (a) Fluctuation Error, (b) Lag Error, and (c) Total Error.

Now, we proceed to compute the optimum value of $\bar{\mu}$, which results in the minimum steady-state error. The optimum value of $\bar{\mu}$ can be evaluated by setting the derivative of the steady-state error ξ_∞ with respect to step size $\bar{\mu}$ to zero. The optimum value thus obtained is given below.

$$\bar{\mu}_{opt} = \frac{\sqrt{c_2^2 + 4c_1c_2} - c_2}{2c_1} \tag{10}$$

where $c_1 = \xi^0 E\left(\frac{1}{r^2}\right)tr(\mathbf{R})$ and $c_2 = \omega^0 \sum_{i=1}^{N}\left(\frac{\lambda_i}{\beta_i}\right)$.

2. Dependence on Input Vector Delay.

The fluctuation error shown in (8a) is independent of the delay D. Hence, APA has the same fluctuation error (by our analysis, the same fluctuation error as NLMS) irrespective of the choice of D (assuming that the rest of the parameters is fixed). However, the lag error increases as D is increased (since the derivative of (8b) with respect to D is strictly positive). Consequently, in the non-stationary case, the total mean-squared error increases as D is increased.

It is known that larger D results in faster convergence under most stationary circumstances [6]. On the other hand, the tracking property of the algorithm diminishes as D increases. Thus, for any choice of D, there is a trade-off between convergence rate (transient behavior) and tracking error (steady-state behavior).

3. Dependence on Number of OCFs.

The fluctuation error of APA is independent of the number of orthogonal correction factors as well. In fact, by our analysis, the fluctuation error of APA (with any D and M) is the same as the fluctuation error of NLMS. This property of APA has been exploited to accelerate the convergence of adaptive filters without compromising on the steady-state error under stationary conditions [4]. In this subsection, we show that the same holds true for nonstationary conditions.

The steady-state lag error, shown in (8b), is a function of M through β_i and γ_i. Unfortunately, there is no simple closed form expression for the optimal choice for M, say M_{opt}, for which the lag-error is minimized. The optimal value M_{opt}, which must be an integer, depends on parameters including D and $\bar{\mu}$. By setting the number of orthogonal correction factors to M_{opt}, the lag error (and hence the total steady-state error) can be minimized. It is known that the convergence rate improves as M is increased [4]. Hence, APA provides a way to improve convergence rate, while simultaneously reducing the steady-state error (equivalently, improving the tracking performance).

4. Tracking Properties of NLMS Algorithm.

Since NLMS is the special case of NLMS-OCF for which $M = 0$, the steady-state fluctuation and lag errors of NLMS can be obtained by setting $M = 0$ in (8). When $M = 0$, we have $\gamma_i = 0$ and $\beta_i = p_i$. Hence,

$$\xi^f_{\infty,NLMS} = \xi^0\left[1 + \frac{\bar{\mu}}{2-\bar{\mu}} E\left(\frac{1}{r^2}\right)tr(\mathbf{R})\right] \tag{11a}$$

$$\xi^l_{\infty,NLMS} = \frac{N\omega^0 tr(\mathbf{R})}{\alpha} \tag{11b}$$

The steady-state mean-squared error and tracking performance of NLMS depend on the step size in the same way as APA. NLMS also has an optimal choice of step size that minimizes the steady-state mean-squared error. The optimal step size can be obtained by substituting $M = 0$ in (10).

IV. VERIFICATION USING SIMULATION

In this section, we provide simulation results to corroborate the properties derived analytically in earlier sections. The system to be identified has a 32-point long impulse response that varies in time according to the random-walk model shown in (6b). The variance ω^0 of the weight increment $\omega_{n,i}$ is assumed to be 10^{-4}. The delay line of the adaptive filter is initialized with true data values (soft initialization) in all simulations. The measurement noise is assumed to be absent, $\xi^0 = 0$, since we are mainly interested in the tracking behavior. In all simulations, the steady-state mean-squared error is computed by time averaging of the instantaneous squared error over 8000 iterations.

Figure 3 compares the steady-state lag error, computed on the basis of our analytical result, and the result from simulation for APA with two OCFs ($M = 2$) and a 32-point delay ($D = 32$). We observe that the two results agree quite well.

Next we show the effect of varying the number of orthogonal correction factors, while keeping the delay and step size fixed at 1 and 0.1, respectively. Figure 4 shows the steady-state lag error as computed from the theoretical result and as determined from simulation, with $D = 8$. We observe that the theoretical results and simulation results agree well. The theoretically predicted optimum value for M in this case is 6, while the optimum value obtained from simulations is 7.

Figure 5 illustrates the dependence of the steady-state lag error on the input vector delay. We observe that the steady-state lag error, in theory, increases linearly as the delay is increased. It is evident from Figure 5 that the simulation results also indicate a linear dependence of the steady-state lag error on the input delay used. While the slope of the theoretical steady-state lag error vs. delay curve is 0.0031, the corresponding simulated result has a slope of 0.0024.

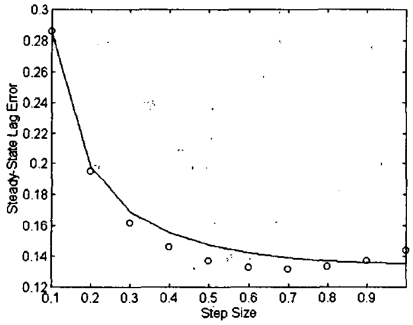

Figure 3. Theoretical (solid) and Simulated (o) Steady-State Lag Error vs. Step Size, with Input "Close" to Assumptions. (White Input with $N=32$, $M=2$, $D=32$, and $\omega^0=10^{-4}$)

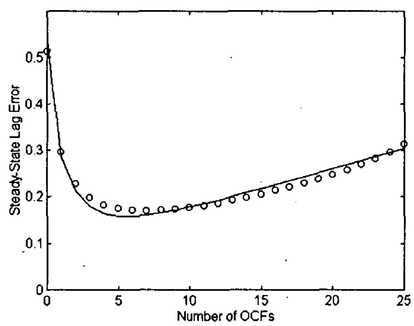

Figure 4. Theoretical (solid) and Simulated (o) Steady-State Lag Error vs. Number of OCFs for Longer Delay. (White Input with $N=32$, $D=8$, $\bar{\mu}=0.1$, and $\omega^0=10^{-4}$)

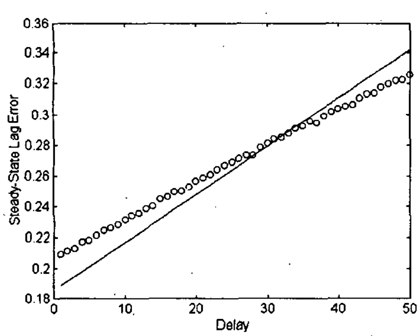

Figure 5. Theoretical (solid) and Simulated (o) Steady-State Lag Error vs. Input Vector Delay. (White Input with $N=32$, $M=2$, $\bar{\mu}=0.1$, and $\omega^0=10^{-4}$)

Figure 6 shows the effect of varying the step size on the tracking performance of NLMS. We observe a very good match between the theoretical simulation results. We also see that as the step size is increased, the tracking property of NLMS improves. Observe that the theoretical and simulated lag errors agree better for NLMS than for NLMS-OCF (shown in Figure 3). This is due to certain approximations, such as neglecting a few terms, made while deriving the expression for lag error. Some of these neglected terms truly vanish when $M=0$.

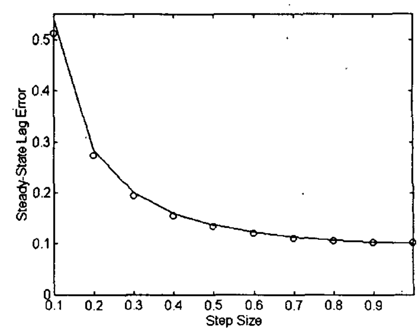

Figure 6. Theoretical (solid) and Simulated (o) NLMS Steady-State Lag Error vs. Step Size. (White Input with $N=32$ and $\omega^0=10^{-4}$)

V. CONCLUSION

An expression for the steady-state mean-squared error of the APA class of algorithms for a randomly time-varying system is presented. The steady-state error consists of two parts, namely a lag error caused by variations in the environment and a fluctuation error caused by measurement noise. The dependence of the steady-state error and of the tracking properties on the user-selectable parameters of APA is discussed. While the lag error depends on all of the parameters (namely $\bar{\mu}$, D, and M), the fluctuation error depends only on the step size $\bar{\mu}$. Increasing D always results in a linear increase in the lag error and hence a corresponding linearly increasing component of the total steady state mean-squared error. There is an optimum choice for $\bar{\mu}$ and M that minimizes the total mean-squared error. Simulation results support our theoretical predictions.

VI. REFERENCES

[1] 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.

[2] 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, November 1997.

[3] 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.

[4] S. G. Sankaran and A. A. (Louis) Beex, "Convergence Behavior of Affine Projection Algorithms," *IEEE Transactions on Signal Processing*, vol. 48, no. 4, pp. 1086-1096, April 2000.

[5] S. G. Sankaran and A. A. (Louis) Beex, "Tracking Properties of Affine Projection Algorithms," Submitted to *IEEE Transactions on Signal Processing*, July 2001.

[6] S. G. Sankaran and A. A. (Louis) Beex, "Fast Generalized Affine Projection Algorithm," *International Journal of Adaptive Control and Signal Processing*, vol. 14, no. 6, pp. 623-641, September 2000.