4. Tracking Behavior of the APA Class of Algorithms

In certain applications like echo cancellation, noise cancellation, and equalization, 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. For example, the popular Least Mean Squares (LMS) algorithm, which is known to exhibit very slow convergence, has better tracking behavior than the Recursive Least Squares (RLS) algorithm, which is known to yield fast convergence [49].

Results related to the tracking behavior of APA and NLMS are not available in the existing literature. In this chapter, we analyze the tracking behavior of APA for a randomly time-varying channel. The tracking analysis, as done in our convergence analysis, is based on certain simplifying assumptions, such as the independence assumption and the discrete orientation assumption.

4.1 Tracking Analysis of the Affine Projection Algorithm Class

In addition to Assumptions (A1)-(A3), listed in Section 3.1, we use the following assumption on the underlying system for the tracking analysis of the APA class of algorithms.

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

$$d_n = \mathbf{w}_n^{0^H} \mathbf{x}_n + \varepsilon_n \,. \tag{4.1a}$$

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

$$\mathbf{w}_n^0 = \mathbf{w}_{n-1}^0 + \boldsymbol{\omega}_{n-1} \tag{4.1b}$$

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

$$E\left(\boldsymbol{\omega}_n \boldsymbol{\omega}_n^H\right) = \omega^0 \mathbf{I} \tag{4.1c}$$

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

As in the convergence analysis, the NLMS-OCF adaptation equation is used for the tracking analysis and is first rewritten as in (3.4)-(3.6) using (A3). To analyze the tracking behavior of the APA adaptation equation, given in (3.4), the weight adaptation is rewritten in terms of the weight error vector $\tilde{\mathbf{w}}_n$, where $\tilde{\mathbf{w}}_n = \mathbf{w}_n^0 - \mathbf{w}_n$. The first step in this process is to rewrite (4.1b) as follows.

$$\mathbf{w}_n^0 = \mathbf{w}_{n-kD}^0 + \sum_{j=1}^{kD} \boldsymbol{\omega}_{n-j} \tag{4.2}$$

Using (4.1a) and (4.2), the estimation errors shown in (3.6) are rewritten in terms of the weight error vector as

$$e_n = \tilde{\mathbf{w}}_n^H \mathbf{x}_n + \varepsilon_n \text{, and} \tag{4.3a}$$

$$e_n^k = \tilde{\mathbf{w}}_n^H \mathbf{x}_{n-kD} - \sum_{j=1}^{kD} \boldsymbol{\omega}_{n-j}^H \mathbf{x}_{n-kD} + \varepsilon_{n-kD} \tag{4.3b}$$

Using (3.4), (3.5), and (4.3), the weight adaptation equation in error form is obtained as

$$\tilde{\mathbf{w}}_{n+1} = \left[\mathbf{I} - \sum_{j\in J_n} \bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}}\right] \tilde{\mathbf{w}}_n - \sum_{l\in J_n} \bar{\mu} \frac{\varepsilon_{n-lD}^* \mathbf{x}_{n-lD}}{\mathbf{x}_{n-lD}^H \mathbf{x}_{n-lD}} + \sum_{m\in J_n} \left[\bar{\mu} \frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H \mathbf{x}_{n-mD}} \left(\sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}\right)\right] + \boldsymbol{\omega}_n \tag{4.4}$$

where $J_n \subseteq \{0,1,2,...,M\}$ is a set of $M+1$ or fewer indices j for which the $\mathbf{x}_{n-jD}$ are orthogonal to each other, since $\mu_j = 0$ for $j \notin J_n$. Equation (4.4) is in a form suitable for tracking analysis. We begin the convergence analysis with the computation of the weight error vector covariance.

Using (4.4), the covariance of the weight error vector $\tilde{\mathbf{w}}_n$ is given by:

$$
\begin{aligned}
\operatorname{cov}(\tilde{\mathbf{w}}_{n+1}) = & E\left(\left[\mathbf{I}-\sum_{j\in J_n}\bar{\mu}\frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H\mathbf{x}_{n-jD}}\right]\tilde{\mathbf{w}}_n\tilde{\mathbf{w}}_n^H\left[\mathbf{I}-\sum_{l\in J_n}\bar{\mu}\frac{\mathbf{x}_{n-lD}\mathbf{x}_{n-lD}^H}{\mathbf{x}_{n-lD}^H\mathbf{x}_{n-lD}}\right]\right) \\
& + E\left(\left[\sum_{j\in J_n}\bar{\mu}\frac{\varepsilon_{n-jD}^*\mathbf{x}_{n-jD}}{\mathbf{x}_{n-jD}^H\mathbf{x}_{n-jD}}\right]\left[\sum_{l\in J_n}\bar{\mu}\frac{\varepsilon_{n-lD}\mathbf{x}_{n-lD}^H}{\mathbf{x}_{n-lD}^H\mathbf{x}_{n-lD}}\right]\right) \\
& + E\left(\sum_{j\in J_n}\left[\bar{\mu}\frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H\mathbf{x}_{n-jD}}\left\{\sum_{l=1}^{jD}\boldsymbol{\omega}_{n-l}\right\}\right]\sum_{m\in J_n}\left[\bar{\mu}\left\{\sum_{p=1}^{mD}\boldsymbol{\omega}_{n-p}^H\right\}\frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H\mathbf{x}_{n-mD}}\right]\right) \\
& + E\left(\boldsymbol{\omega}_n\boldsymbol{\omega}_n^H\right) \\
& + E\left(\left[\mathbf{I}-\sum_{j\in J_n}\bar{\mu}\frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H\mathbf{x}_{n-jD}}\right]\tilde{\mathbf{w}}_n\sum_{m\in J_n}\left[\bar{\mu}\left\{\sum_{p=1}^{mD}\boldsymbol{\omega}_{n-p}^H\right\}\frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H\mathbf{x}_{n-mD}}\right]\right) \\
& + E\left(\sum_{j\in J_n}\left[\bar{\mu}\frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H\mathbf{x}_{n-jD}}\left\{\sum_{l=1}^{jD}\boldsymbol{\omega}_{n-l}\right\}\right]\tilde{\mathbf{w}}_n^H\left[\mathbf{I}-\sum_{l\in J_n}\bar{\mu}\frac{\mathbf{x}_{n-lD}\mathbf{x}_{n-lD}^H}{\mathbf{x}_{n-lD}^H\mathbf{x}_{n-lD}}\right]\right)
\end{aligned}
\tag{4.5}
$$

Note that the cross terms involving ε_n or $\boldsymbol{\omega}_n$ do not appear in the above expression, since these terms vanish. This can be proved by neglecting the dependency of $\tilde{\mathbf{w}}_n$ on past measurement noise, and by using that ε_n and $\boldsymbol{\omega}_n$ are of zero mean and that they are independent of each other and of $\mathbf{x}_n$.

Let us define the diagonal elements of the transformed covariance matrix $\mathbf{V}^H\operatorname{cov}(\tilde{\mathbf{w}}_n)\mathbf{V}$ as $\tilde{\lambda}_{n,i}$ for i = 1, 2, …, N. That is,

$$
\left[\mathbf{V}^H\operatorname{cov}(\tilde{\mathbf{w}}_n)\mathbf{V}\right]_{ii} = \mathbf{v}_i^H\operatorname{cov}(\tilde{\mathbf{w}}_n)\mathbf{v}_i = \tilde{\lambda}_{n,i}\,. \tag{4.6}
$$

Note that this does not mean that $\mathbf{V}^H\operatorname{cov}(\tilde{\mathbf{w}}_n)\mathbf{V}$ is a diagonal matrix.

With the above notation, the pre- and post-multiplication of (4.5) by $\mathbf{v}_i^H$ and $\mathbf{v}_i$ respectively and neglecting the dependency of $\tilde{\mathbf{w}}_n$ on the past input vectors that appear in the first term of (4.5) results in

$$
\begin{aligned}
\tilde{\lambda}_{n+1,i} &= E\left(\mathbf{v}_i^H \left[\mathbf{I} - \sum_{j\in J_n} \bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \right] \operatorname{cov}\left(\tilde{\mathbf{w}}_n \tilde{\mathbf{w}}_n^H\right) \left[\mathbf{I} - \sum_{l\in J_n} \bar{\mu} \frac{\mathbf{x}_{n-lD}\mathbf{x}_{n-lD}^H}{\mathbf{x}_{n-lD}^H \mathbf{x}_{n-lD}} \right] \mathbf{v}_i \right) \\
&+ E\left(\mathbf{v}_i^H \left[\sum_{j\in J_n} \bar{\mu} \frac{\varepsilon_{n-jD}^* \mathbf{x}_{n-jD}}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \right] \left[\sum_{l\in J_n} \bar{\mu} \frac{\varepsilon_{n-lD} \mathbf{x}_{n-lD}^H}{\mathbf{x}_{n-lD}^H \mathbf{x}_{n-lD}} \right] \mathbf{v}_i \right) \\
&+ E\left(\mathbf{v}_i^H \sum_{j\in J_n} \left[\bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \left\{ \sum_{l=1}^{jD} \boldsymbol{\omega}_{n-l} \right\} \right] \sum_{m\in J_n} \left[\bar{\mu} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H \mathbf{x}_{n-mD}} \right] \mathbf{v}_i \right) \\
&+ E\left(\mathbf{v}_i^H \boldsymbol{\omega}_n \boldsymbol{\omega}_n^H \mathbf{v}_i \right) \\
&+ E\left(\mathbf{v}_i^H \left[\mathbf{I} - \sum_{j\in J_n} \bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \right] \tilde{\mathbf{w}}_n \sum_{m\in J_n} \left[\bar{\mu} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H \mathbf{x}_{n-mD}} \right] \mathbf{v}_i \right) \\
&+ E\left(\mathbf{v}_i^H \sum_{j\in J_n} \left[\bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \left\{ \sum_{l=1}^{jD} \boldsymbol{\omega}_{n-l} \right\} \right] \tilde{\mathbf{w}}_n^H \left[\mathbf{I} - \sum_{l\in J_n} \bar{\mu} \frac{\mathbf{x}_{n-lD}\mathbf{x}_{n-lD}^H}{\mathbf{x}_{n-lD}^H \mathbf{x}_{n-lD}} \right] \mathbf{v}_i \right)
\end{aligned} \tag{4.7}
$$

Now, we evaluate each of the terms in the above expression starting with the first term.

The first two terms of the above expression are the same as in the expression derived during convergence analysis. Hence, using (3.18)

$$
T_1 + T_2 = (1 - \alpha\beta_i)\tilde{\lambda}_{n,i} + \bar{\mu}^2 \xi^0 E\left(\frac{1}{r^2} \right) \beta_i \tag{4.8}
$$

Now, consider the third term of (4.7).

$$
T_3 = E\left(\mathbf{v}_i^H \sum_{j\in J_n} \left[\bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \left\{ \sum_{l=1}^{jD} \boldsymbol{\omega}_{n-l} \right\} \right] \sum_{m\in J_n} \left[\bar{\mu} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H \mathbf{x}_{n-mD}} \right] \mathbf{v}_i \right) \tag{4.9}
$$

Using (3.10) and (3.16), the product appearing in the expectation above vanishes unless $\mathbf{x}_{n-jD}$, for some $j \in J_n$, is parallel to $\mathbf{v}_i$. Furthermore, the result of the product depends on the value of j that satisfies this condition. Hence, to evaluate the required expected value, we need to consider all possible values of m for which $\mathbf{x}_{n-mD}$ is parallel to $\mathbf{v}_i$. This happens for $m = j$, if $j \in J_n$ $j \in J_n$ if and only if $\mathbf{x}_{n-jD}$ is not parallel to $\mathbf{x}_{n-mD}$ for $0 \le m < j$, i.e., $\mathbf{x}_{n-jD}$ is a new direction being used for adaptation during the current iteration. Equivalently, $j \in J_n$ and $\mathbf{x}_{n-jD}$ is parallel to $\mathbf{v}_i$ if and only if the minimum value of m for which $\mathbf{x}_{n-mD}$ is parallel to $\mathbf{v}_i$ is j.

Mathematically, the necessary and sufficient condition is $\min\{m \in \{0,1,\ldots,M\} : \mathbf{x}_{n-mD} \parallel \mathbf{v}_i\} = j$. Using this result along with (A3), (4.9) can be simplified as follows.

$$T_3 = \bar{\mu}^2 \mathbf{v}_i^H \sum_{j=1}^{M} \left[P\left(\min\{m \in \{0,1,\ldots,M\} : \mathbf{x}_{n-mD} \parallel \mathbf{v}_i\} = j\right) E\left\{ \left(\sum_{l=1}^{jD} \boldsymbol{\omega}_{n-l} \right) \left(\sum_{p=1}^{jD} \boldsymbol{\omega}_{n-p}^H \right) \right\} \right] \mathbf{v}_i \tag{4.10}$$

Now, we evaluate the probability that appears in the above expression.

$$\begin{aligned} P\left(\min\{m \in \{0,1,\ldots,M\} : \mathbf{x}_{n-mD} \parallel \mathbf{v}_i\} = j\right) &= P\left(\mathbf{x}_{n-jD} \parallel \mathbf{v}_i \;\&\; \mathbf{x}_{n-mD} \perp \mathbf{v}_i \;\forall m \in \{0,1,\ldots,j-1\}\right) \\ &= p_i (1-p_i)^j \end{aligned} \tag{4.11}$$

Using (A4) and (4.11), we can rewrite (4.10) as

$$\begin{aligned} T_3 &= \bar{\mu}^2 \mathbf{v}_i^H \sum_{j=1}^{M} \left[p_i (1-p_i)^j E\left(\sum_{l=1}^{jD} \boldsymbol{\omega}_{n-l} \boldsymbol{\omega}_{n-l}^H \right) \right] \mathbf{v}_i \\ &= \bar{\mu}^2 \mathbf{v}_i^H \sum_{j=1}^{M} \left[p_i (1-p_i)^j \sum_{l=1}^{jD} E\left(\boldsymbol{\omega}_{n-l} \boldsymbol{\omega}_{n-l}^H \right) \right] \mathbf{v}_i \\ &= \bar{\mu}^2 \mathbf{v}_i^H \sum_{j=1}^{M} \left[p_i (1-p_i)^j \, jD\omega^0 \mathbf{I} \right] \mathbf{v}_i \\ &= \bar{\mu}^2 D\omega^0 p_i \left[\frac{(1-p_i)\beta_i}{p_i^2} - \frac{(M+1)(1-\beta_i)}{p_i} \right] \\ &= \bar{\mu}^2 D\omega^0 \gamma_i \end{aligned} \tag{4.12a}$$

where

$$\gamma_i = \frac{(1+Mp_i)\beta_i - (M+1)p_i}{p_i} \tag{4.12b}$$

and we used the following identity in going from line 3 to line 4 of (4.12a).

$$\sum_{n=1}^{N} n a^n = \frac{a\left(1-a^{N+1}\right)}{(1-a)^2} - \frac{(N+1)a^{N+1}}{(1-a)} \tag{4.13}$$

Using (A4), the fourth term of (4.7) is evaluated to be

$$T_4 = E\left(\mathbf{v}_i^H \boldsymbol{\omega}_n \boldsymbol{\omega}_n^H \mathbf{v}_i\right) = \mathbf{v}_i^H E\left(\boldsymbol{\omega}_n \boldsymbol{\omega}_n^H\right) \mathbf{v}_i = \omega^0 \tag{4.14}$$

Now, consider the fifth term of (4.7)

$$T_5 = E\left(\mathbf{v}_i^H \left[\mathbf{I} - \sum_{j\in J_n} \bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \right] \tilde{\mathbf{w}}_n \sum_{m\in J_n} \left[\bar{\mu} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H \mathbf{x}_{n-mD}} \right] \mathbf{v}_i \right)$$

$$= E\left(\mathbf{v}_i^H \tilde{\mathbf{w}}_n \sum_{m\in J_n} \left[\bar{\mu} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H \mathbf{x}_{n-mD}} \right] \mathbf{v}_i \right) \tag{4.15}$$

$$- E\left(\mathbf{v}_i^H \left[\sum_{j\in J_n} \bar{\mu} \frac{\mathbf{x}_{n-jD}\mathbf{x}_{n-jD}^H}{\mathbf{x}_{n-jD}^H \mathbf{x}_{n-jD}} \right] \tilde{\mathbf{w}}_n \sum_{m\in J_n} \left[\bar{\mu} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \frac{\mathbf{x}_{n-mD}\mathbf{x}_{n-mD}^H}{\mathbf{x}_{n-mD}^H \mathbf{x}_{n-mD}} \right] \mathbf{v}_i \right)$$

Neglecting the dependence of the current estimate for weights $\mathbf{w}_n$ on the past inputs and on the past $\boldsymbol{\omega}_n$, which - especially for small values of step size - does not introduce significant errors, we can write

$$E\left(\tilde{\mathbf{w}}_n \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \right) = E\left(\left\{ \mathbf{w}_n^0 - \mathbf{w}_n \right\} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \right)$$

$$= E\left(\mathbf{w}_n^0 \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \right) \tag{4.16}$$

$$= E\left(\left\{ \mathbf{w}_{n-mD}^0 + \sum_{j=1}^{mD} \boldsymbol{\omega}_{n-j} \right\} \left\{ \sum_{p=1}^{mD} \boldsymbol{\omega}_{n-p}^H \right\} \right)$$

$$= mD\omega^0 \mathbf{I}$$

Invoking (4.11) and (4.16), we can simplify (4.15) further as follows.

$$T_5 = \bar{\mu}(1-\bar{\mu}) \mathbf{v}_i^H \sum_{j=1}^{M} \left[p_i (1-p_i)^j \, jD\omega^0 \mathbf{I} \right] \mathbf{v}_i$$

$$= \bar{\mu}(1-\bar{\mu}) D\omega^0 p_i \left[\frac{(1-p_i)\beta_i}{p_i^2} - \frac{(M+1)(1-\beta_i)}{p_i} \right] \tag{4.17}$$

$$= \bar{\mu}(1-\bar{\mu}) D\omega^0 \gamma_i$$

Since T_6 is just the Hermitian transpose of the real-valued quantity T_5, we conclude that $T_5 = T_6$. Adding the six individual terms reduced above and invoking the definition $\alpha = \bar{\mu}(2-\bar{\mu})$, we get

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

$$= (1-\alpha\beta_i)\tilde{\lambda}_{n,i} + \bar{\mu}^2 \xi^0 E\left(\frac{1}{r^2} \right) \beta_i + \omega^0 (1+\alpha D\gamma_i) \tag{4.18}$$

As we showed while analyzing convergence in Section 3.1, if $\bar{\mu} \in (0,2)$, then the recursive equation for $\tilde{\lambda}_{n,i}$ shown in (4.18) is stable. Assuming that the stability condition is met, from (4.18), the steady-state value of $\tilde{\lambda}_{n,i}$ is given by

$$\lim_{n\to\infty} \tilde{\lambda}_{n,i} = \frac{\bar{\mu}}{2-\bar{\mu}} \xi^0 E\left(\frac{1}{r^2}\right) + \frac{\omega^0 (1+\alpha D\gamma_i)}{\alpha\beta_i} \tag{4.19}$$

We use the above result to calculate the steady-state mean-squared output estimation error as follows. From (3.21), the mean-squared error is given by

$$\xi_n = \xi^0 + \sum_{i=1}^{N} \lambda_i \tilde{\lambda}_{n,i} \tag{4.20}$$

Using (4.19) and (4.20), the steady-state error is given by

$$\xi_\infty = \lim_{n\to\infty} \xi_n = \xi^0 \left[1 + \frac{\bar{\mu}}{2-\bar{\mu}} E\left(\frac{1}{r^2}\right) tr(\mathbf{R})\right] + \sum_{i=1}^{N} \frac{\omega^0 \lambda_i (1+\alpha D\gamma_i)}{\alpha\beta_i} \tag{4.21}$$

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

$$\xi_\infty^f = \xi^0 \left[1 + \frac{\bar{\mu}}{2-\bar{\mu}} E\left(\frac{1}{r^2}\right) tr(\mathbf{R})\right] \tag{4.22a}$$

$$\xi_\infty^l = \sum_{i=1}^{N} \frac{\omega^0 \lambda_i (1+\alpha D\gamma_i)}{\alpha\beta_i} \tag{4.22b}$$

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. In the next section, we list the tracking properties of the APA class of algorithms.

4.2 Tracking Properties of the APA Class of Algorithms

In this section, we discuss the dependence of the tracking rate and the total mean-squared error on each of the user-selectable parameters of APA, namely the step size $\bar{\mu}$, the input vector delay D, and the number of orthogonal correction factors M, assuming that the rest are kept constant.

1. Dependence on Step size:

We begin by computing the derivative of the two components of the steady-state error, shown in (4.22), with respect to step size.

$$\frac{d}{d\bar{\mu}}\xi_{\infty}^{f} = \frac{2}{(2-\bar{\mu})^2}\xi^0 E\left(\frac{1}{r^2}\right)tr(\mathbf{R}) > 0 \quad \forall\bar{\mu}\in(0,2) \tag{4.23a}$$

$$\frac{d}{d\bar{\mu}}\xi_{\infty}^{l} = -\frac{2(1-\bar{\mu})}{\alpha^2}\omega^0\sum_{i=1}^{N}\left(\frac{\lambda_i}{\beta_i}\right) \leq 0 \quad \forall\bar{\mu}\in(0,1] \tag{4.23b}$$

$$\frac{d}{d\bar{\mu}}\xi_{\infty}^{l} = -\frac{2(1-\bar{\mu})}{\alpha^2}\omega^0\sum_{i=1}^{N}\left(\frac{\lambda_i}{\beta_i}\right) > 0 \quad \forall\bar{\mu}\in(1,2) \tag{4.23c}$$

From (4.23a), 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 (4.23b) and (4.23c) 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 $\bar{\mu}\in(1,2)$.

It is evident from (4.22b) that the lag error depends on $\alpha=\bar{\mu}(2-\bar{\mu})$, and hence on the step size. More importantly, the lag error depends on step size only through α, since none of the variables other than α, appearing on the right hand side of (4.22b) depend on $\bar{\mu}$. Necause of the symmetry of α about $\bar{\mu}=1$, for any value of $\bar{\mu}$, say $\bar{\mu}^*$, there is the value $\bar{\mu}=2-\bar{\mu}^*$ that results in the same lag error as $\bar{\mu}=\bar{\mu}^*$. 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 4.1. It is also possible to minimize a convex combination of the fluctuation and lag errors, if desired.

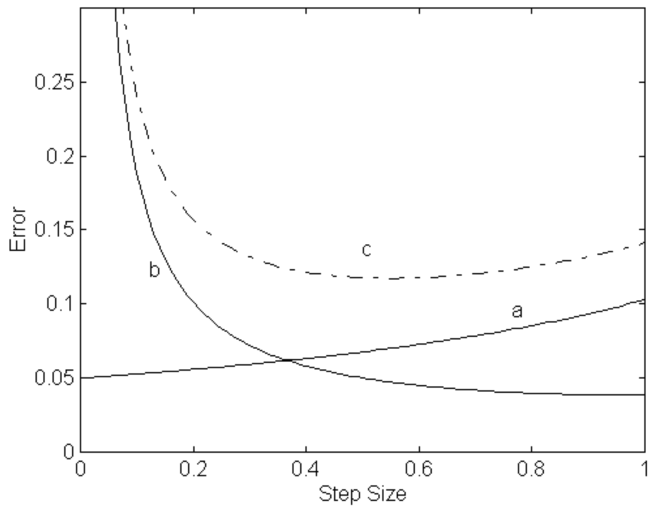

Figure 4.1 Mean-Squared Error for Different Step Sizes: (a) Fluctuation Error, (b) Lag Error, and (c) Total Error.

Now, we proceed to compute the optimum value of $\bar{\mu}$, which results in 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 shown below.

$$\bar{\mu}_{opt} = \frac{\sqrt{c_2^2 + 4c_1c_2} - c_2}{2c_1} \tag{4.24a}$$

where

$$c_1 = \xi^0 E\left(\frac{1}{r^2}\right) tr(\mathbf{R}) \tag{4.24b}$$

$$c_2 = \omega^0 \sum_{i=1}^{N} \left(\frac{\lambda_i}{\beta_i}\right) \tag{4.24c}$$

2. Dependence on Delay.

The fluctuation error shown in (4.22a) 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 (4.22b) with respect to D is strictly positive). Consequently, in the non-stationary case, the total mean-squared error increases as D is increased.

In Chapter 5, it will be shown that larger D results in faster convergence under most stationary circumstances. 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 M

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. In Chapter 3, we showed that this property of APA can be exploited to accelerate the convergence of adaptive filters without compromising on the steady-state error under stationary conditions. In this subsection, we show that the same holds for nonstationary conditions as well.

The steady-state lag error, shown in (4.22b), 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 has to 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. In Chapter 3, we showed that the convergence rate improves as M is increased. Hence, APA provides a way to improve convergence rate, while simultaneously reducing the steady-state error (equivalently, improving the tracking performance).

4.3 Tracking Properties of the NLMS Algorithm

Since NLMS is a special case of NLMS-OCF with $M = 0$, the steady-state fluctuation and lag errors of NLMS can be obtained by setting $M = 0$ in (4.22). When $M = 0$, we have $\gamma_i = 0$ and $\beta_i = p_i$. Hence,

$$\xi^{f}_{\infty,NLMS} = \xi^{0}\left[1 + \frac{\overline{\mu}}{2-\overline{\mu}} E\left(\frac{1}{r^2}\right) tr(\mathbf{R})\right] \tag{4.25a}$$

$$\xi^{l}_{\infty,NLMS} = \frac{N\omega^{0} tr(\mathbf{R})}{\alpha} \tag{4.25b}$$

The steady-state mean-squared error and tracking performance of NLMS depend on the step size in the same way as for 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 (4.24). When $M = 0$, using (3.3b)

$$\sum_{i=1}^{N}\left(\frac{\lambda_i}{\beta_i}\right) = N\ tr(\mathbf{R}) \tag{4.26}$$

Hence,

$$\overline{\mu}_{opt,NLMS} = \frac{\sqrt{c_2^2 + 4c_1c_2} - c_2}{2c_1} \tag{4.27a}$$

where

$$c_1 = \xi^{0} E\left(\frac{1}{r^2}\right) \tag{4.27b}$$

$$c_2 = N\omega^{0} \tag{4.27c}$$

4.4 Simulation Results

In this section, we provide simulation results to corroborate the properties derived analytically in earlier sections and to also discuss limitations introduced by the assumptions used in the analysis. The system to be identified has a 32-point long impulse response that varies in time according to the random-walk model shown in (4.1b). 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 4.2 compares the steady-state lag error computed based on 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 reasonably well. A similar result corresponding to unit delay ($D = 1$) is shown in Figure 4.3. We see that there is a larger deviation between the theoretical prediction and the simulation result when D=1 than when $D = 32$. This is explained by 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.

Even though there is a deviation between the theoretical and the simulation results when the assumptions are not satisfied (as seen in Figure 4.3), this deviation is not very large. Furthermore, the general trend of the result predicted by theory (sharp fall followed by a slow reduction) matches the trend of the simulation result. Hence, the analytical result is still useful.

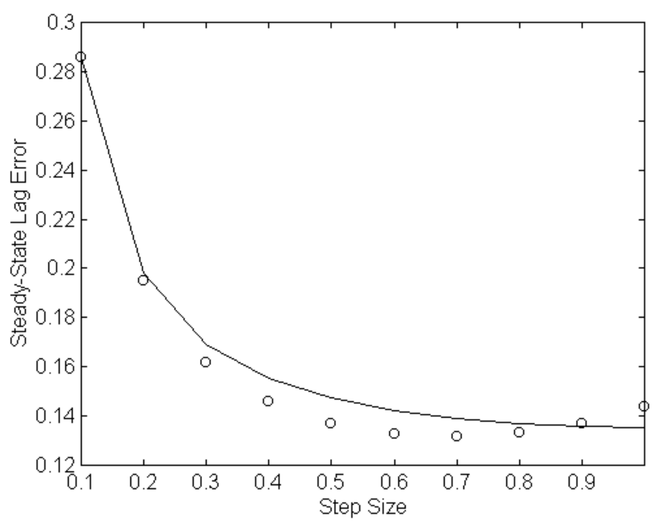

Figure 4.2 Theoretical (solid) and Simulated (o) Steady-State Lag Error for Different Step Sizes with Input "Close" to Assumptions.
(White Input with $N = 32$, $M = 2$, $D = 32$, and $\omega^0 = 10^{-4}$)

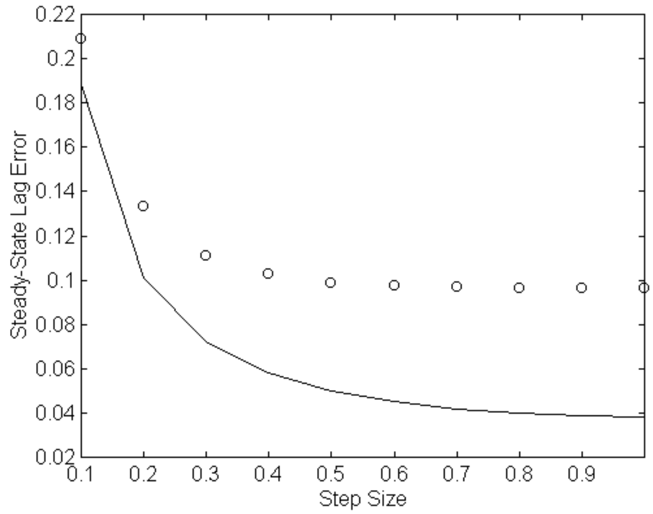

Figure 4.3 Theoretical (solid) and Simulated (o) Steady-State Lag Error for Different Step Sizes with Input Not Matching Assumptions.
(White Input with $N = 32$, $M = 2$, $D = 1$, and $\omega^0 = 10^{-4}$)

Now, 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 4.4 shows the steady-state lag error as computed from the theoretical result and as determined from simulation. Firstly, we observe that there is good agreement between theory and simulation in the results shown in Figure 4.4. Secondly, it is evident from Figure 4.4 that there is an optimum value of M for which the lag error is minimized. While the theoretically predicted optimum value is 19, the optimum value found on the basis of the simulation results is 20. (Note that the simulation result is based on the lag-error estimates from a finite number of trials and hence has some error due to the variance of the estimator.) This also suggests that there is a reasonable match between theory and simulation.

Figure 4.5 shows results similar to the result shown in Figure 4.4, but with $D = 8$. Here also 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.

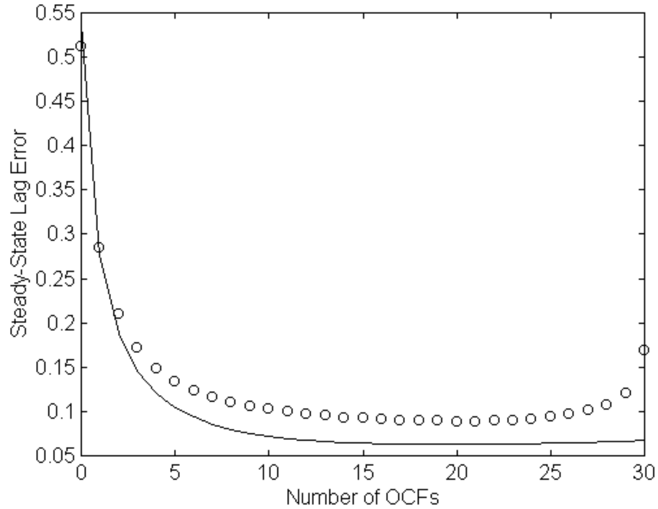

Figure 4.4 Theoretical (solid) and Simulated (o) Steady-State Lag Errors for Different Number of OCFs with Short Delay.
(White Input with $N = 32$, $D = 1$, $\bar{\mu} = 0.1$, and $\omega^0 = 10^{-4}$)

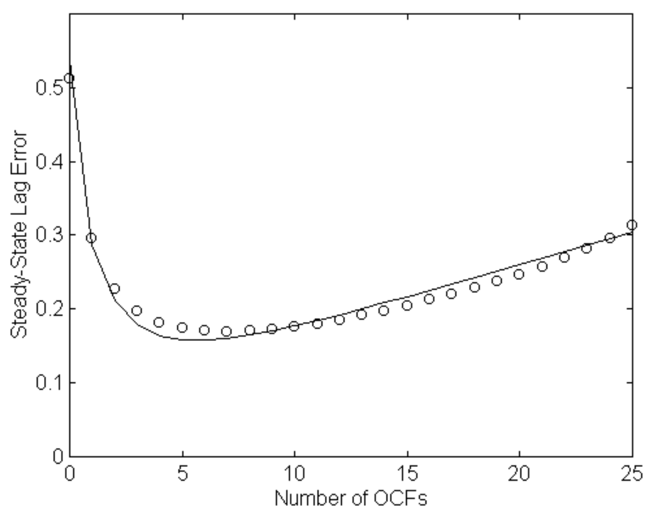

Figure 4.5 Theoretical (solid) and Simulated (o) Steady-State Lag Error for Different Number of OCFs with Long Delay.
(White Input with $N = 32$, $D = 8$, $\bar{\mu} = 0.1$, and $\omega^0 = 10^{-4}$)

Figure 4.6 illustrates the dependence of the steady-state lag error on the delay used for OCF generation. We observe that the steady-state lag error, in theory, increases linearly as the delay is increased. It is evident from Figure 7 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. This difference is due to the approximations made in the tracking analysis.

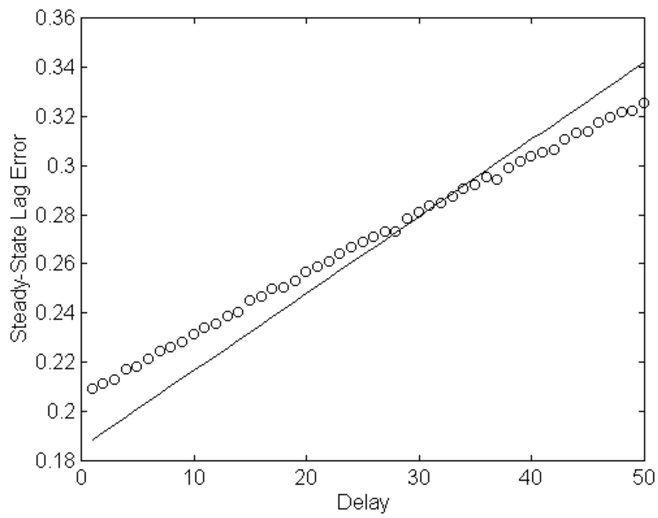

Figure 4.6 Theoretical (solid) and Simulated (o) Steady-State Lag Error for Different Delays.
(White Input with $N = 32$, $M = 2$, $\bar{\mu} = 0.1$, and $\omega^0 = 10^{-4}$)

Figure 4.7 shows the effect of varying the step size on the tracking performance of NLMS. We observe a very good match between the theoretical results and the simulation results shown in Figure 4.7 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 Figures 4.2 and 4.3). 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 $M = 0$. This eliminates the error due to approximation in the case of NLMS.

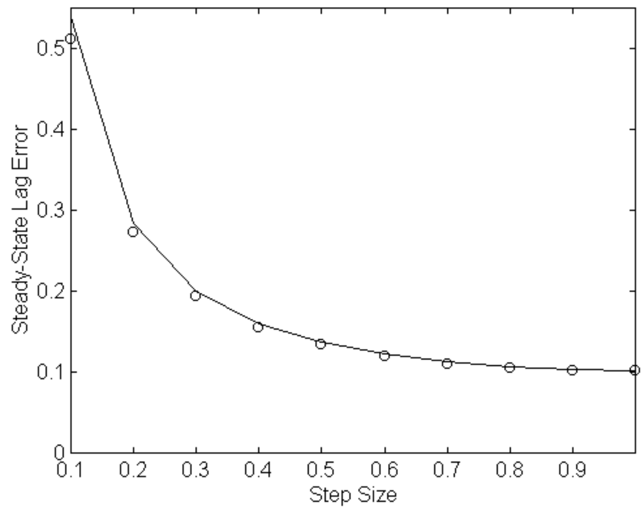

Figure 4.7 Theoretical (solid) and Simulated (o) NLMS Steady-State Lag Error for Different Step Sizes.
(White Input with $N = 32$ and $\omega^0 = 10^{-4}$)

4.5 Conclusion

The tracking behavior of the APA class of algorithms for a randomly time-varying system is analyzed under certain simplifying assumptions on the data. An expression for the steady-state mean-squared error is derived. 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 conclusions.