

# Mean-Square Performance of a Family of Affine Projection Algorithms

Hyun-Chool Shin and Ali H. Sayed, *Fellow, IEEE*

**Abstract**—Affine projection algorithms are useful adaptive filters whose main purpose is to speed the convergence of LMS-type filters. Most analytical results on affine projection algorithms assume special regression models or Gaussian regression data. The available analysis also treat different affine projection filters separately. This paper provides a unified treatment of the mean-square error, tracking, and transient performances of a family of affine projection algorithms. The treatment relies on energy conservation arguments and does not restrict the regressors to specific models or to a Gaussian distribution. Simulation results illustrate the analysis and the derived performance expressions.

**Index Terms**—Affine projection algorithm, energy-conservation, learning-curve, steady-state analysis, tracking analysis, transient analysis.

## I. INTRODUCTION

THE normalized least mean-squares (NLMS) algorithm is a widely used adaptive algorithm due to its computational simplicity and ease of implementation. However, colored input signals can deteriorate its convergence speed appreciably [1], [2]. To address this problem, Ozeki and Umeda [3] developed the basic form of an affine projection algorithm (APA) using affine subspace projections. APA is a useful family of adaptive filters whose main purpose is to speed the convergence of LMS-type filters, especially for correlated data, at a computational cost that is still comparable to that of LMS. This class of filters is particularly useful in echo cancellation applications, e.g., [4]. While NLMS updates the weights based only on the current input vector, APA updates the weights based on  $K$  previous input vectors. Since [3], many variants of APA have been devised independently from different perspectives such as the regularized APA (R-APA) [4], the partial rank algorithm (PRA) [5], the decorrelating algorithm (DA) [6], and NLMS with orthogonal correction factors (NLMS-OCF) [7]. We will refer to all these algorithms as belonging to the APA family (see also [8] and [9]).

Manuscript received October 23, 2002; revised April 11, 2003. This work was supported in part by the National Science Foundation under Grants ECS-9820765 and CCR-0208573. This work was performed while H. Shin was a visiting graduate student at the UCLA Adaptive Systems Laboratory. His work was supported in part by the Brain Korea (BK) 21 Program funded by the Ministry of Education and in part by HY-SDR Reserch Center at Hanyang University under the ITRC Program of MIC, Korea. The associate editor coordinating the review of this paper and approving it for publication was Dr. Behrouz Farhang-Boroujeny.

H.-C. Shin is with Division of Electronics and Computer Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Korea.

A. H. Sayed is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.edu).

Digital Object Identifier 10.1109/TSP.2003.820077

The transient behavior of affine projection algorithms is not as widely studied as that of NLMS. The available results have progressed more for some variations than others, and most analyses assume particular models for the regression data. For example, in [10], convergence analyses in the mean and in the mean-square senses are presented for the binormalized data-reusing LMS (BNDR-LMS) algorithm. Although the results show good agreement with simulations, the arguments are based on a particular model for the input signal and are applicable only to second-order APA. Likewise, the convergence results in [9] focus on NLMS-OCF and rely on a special model for the input signal vector. A convergence analysis of DA is given in [11], where the theoretical results of [6] are extended to the evaluation of learning curves assuming a Gaussian autoregressive input model. All these results provide useful design guidelines. However, each APA form is usually studied separately with specific techniques. Such distinct treatments tend to obscure commonalities that exist among algorithms.

In this paper, we provide a unified treatment of the transient performance of the APA family. In particular, we derive expressions for the mean-square error and tracking performances, as well as conditions on the step-size for mean-square stability. Our derivation relies on energy conservation arguments [12]–[18], and it does not restrict the regression data to being Gaussian or white. Extensive simulations at the end of the paper illustrate the derived results.

Throughout the paper, the following notations are adopted:

$\ \cdot\ $	Euclidean norm of a vector.
$\text{Tr}(\cdot)$	Trace of a matrix.
$\text{diag}\{\dots\}$	Diagonal matrix of its entries $\{\dots\}$ .
$(\cdot)^*$	Hermitian conjugation (complex conjugation for scalars).
$(\cdot)^T$	Transpose of a vector or a matrix.
$\det(\cdot)$	Determinant of a matrix.
$\lambda_{\max}(\cdot)$	Largest eigenvalue of a matrix.
$\mathbb{R}^+$	Set of positive real numbers.

In addition, small boldface letters are used to denote vectors, and capital letters are used to denote matrices, e.g.,  $\mathbf{c}$  and  $C$ . The symbol  $I$  denotes the identity matrix of appropriate dimensions. All vectors are *column vectors* except for the input data vector denoted by  $\mathbf{u}_i$ , which is taken to be a row vector for convenience of notation.

The paper is organized as follows. In the next section, the data model and reviews of the APA family are provided. In Section III, by examining the mean-square performance of the APA family, expressions for the steady-state mean-square error (MSE) are derived. Section IV studies the tracking ability of the APA family. In Section V, the transient performance is analyzed,

and then, the learning behavior is characterized. Section VI illustrates the theoretical results by giving several simulation results.

## II. DATA MODELS AND APA FAMILY

Consider reference data  $\{d(i)\}$  that arise from the linear model

$$d(i) = \mathbf{u}_i \mathbf{w}^\circ + v(i) \quad (1)$$

where  $\mathbf{w}^\circ$  is an unknown column vector that we wish to estimate,  $v(i)$  accounts for measurement noise, and  $\mathbf{u}_i$  denotes  $1 \times M$  row input (regressor) vectors with a positive-definite covariance matrix,  $R_u = E[\mathbf{u}_i^* \mathbf{u}_i]$ . In this paper, we focus on a general class of affine projection algorithms for estimating  $\mathbf{w}^\circ$  of the form

$$\mathbf{w}_i = \mathbf{w}_{i-1-\alpha(K-1)} + \mu U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i \quad (2)$$

where  $\mathbf{e}_i = \mathbf{d}_i - U_i \mathbf{w}_{i-1-\alpha(K-1)}$ ,  $\mathbf{w}_i$  is an estimate for  $\mathbf{w}^\circ$  at iteration  $i$ ,  $\mu$  is the step size, and

$$U_i = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{u}_{i-D} \\ \vdots \\ \mathbf{u}_{i-(K-1)D} \end{bmatrix}, \quad \mathbf{d}_i = \begin{bmatrix} d(i) \\ d(i-D) \\ \vdots \\ d(i-(K-1)D) \end{bmatrix}.$$

Different choices of the parameters  $\{K, \epsilon, \alpha, D\}$  result in different affine projection algorithms. Table I defines the parameters for some special cases. For example, the choices  $\epsilon = 0$ ,  $\alpha = 0$ , and  $D = 1$  result in the standard APA

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu U_i^* (U_i U_i^*)^{-1} \mathbf{e}_i.$$

For NLMS-OCF, it is further assumed that  $\mathbf{u}_{i-jD}$  is orthogonal to  $\mathbf{u}_i, \mathbf{u}_{i-D}, \dots, \mathbf{u}_{i-(j-1)D}$ . For PRA, it is understood that  $\mathbf{w}_m = \mathbf{w}_{i-K}$ , for  $m = i-1, \dots, i-K$ , i.e., the weight vector is updated once every  $K$  iterations.

Most algorithms assume  $K \leq M$ . Moreover, although we focus on (2), our approach can be extended to other APA algorithms such as DA, which is not covered by (2).

## III. MEAN SQUARE PERFORMANCE OF APA

Our first objective is to evaluate the steady-state mean-square error performance of the APA family (2), i.e., to compute

$$\text{MSE} = \lim_{i \rightarrow \infty} E|e(i)|^2$$

where

$$e(i) = d(i) - \mathbf{u}_i \mathbf{w}_{i-1}$$

is the output estimation error at time  $i$ . To do so, we will rely on energy-conservation arguments.

### A. Energy Conservation Relation

Let  $K' = \alpha(K-1)$ . Note that  $K' = 0$  for all algorithms listed in Table I, except PRA. Then, (2) becomes

$$\mathbf{w}_i = \mathbf{w}_{i-1-K'} + \mu U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i \quad (3)$$

TABLE I  
APA FAMILY WHERE  $\{\alpha, K, D\}$  ARE INTEGERS

Algorithm	$K$	$\epsilon$	$\alpha$	$D$
APA	$K \leq M$	$\epsilon = 0$	$\alpha = 0$	$D = 1$
BNDR-LMS	$K = 2$	$\epsilon = 0$	$\alpha = 0$	$D = 1$
R-APA	$K \leq M$	$\epsilon \neq 0$	$\alpha = 0$	$D = 1$
PRA	$K \leq M$	$\epsilon \neq 0$	$\alpha = 1$	$D = 1$
NLMS-OCF	$K \leq M$	$\epsilon = 0$	$\alpha = 0$	$D \geq 1$

which can be rewritten in terms of the weight-error vector  $\tilde{\mathbf{w}}_i = \mathbf{w}^\circ - \mathbf{w}_i$  as

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1-K'} - \mu U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i. \quad (4)$$

If we multiply both sides of (4) by  $U_i$  from the left, we find that

$$U_i \tilde{\mathbf{w}}_i = U_i \tilde{\mathbf{w}}_{i-1-K'} - \mu U_i U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i. \quad (5)$$

Introduce the *a posteriori* and *a priori* error vectors  $\{\mathbf{e}_{p,i}, \mathbf{e}_{a,i}\}$

$$\mathbf{e}_{p,i} = U_i \tilde{\mathbf{w}}_i \quad \text{and} \quad \mathbf{e}_{a,i} = U_i \tilde{\mathbf{w}}_{i-1-K'}.$$

Then, from (5), it holds that

$$\boxed{\mathbf{e}_{p,i} = \mathbf{e}_{a,i} - \mu U_i U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i}. \quad (6)$$

We can use (6) to solve for  $\mathbf{e}_i$ , assuming  $U_i U_i^*$  is invertible

$$(\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i = \frac{1}{\mu} (U_i U_i^*)^{-1} (\mathbf{e}_{a,i} - \mathbf{e}_{p,i})$$

and substitute into (4) to get

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1-K'} - U_i^* (U_i U_i^*)^{-1} (\mathbf{e}_{a,i} - \mathbf{e}_{p,i}) \quad (7)$$

which can be rearranged as

$$\tilde{\mathbf{w}}_i + U_i^* (U_i U_i^*)^{-1} \mathbf{e}_{a,i} = \tilde{\mathbf{w}}_{i-1-K'} + U_i^* (U_i U_i^*)^{-1} \mathbf{e}_{p,i}. \quad (8)$$

By evaluating the energies of both sides of this equation, we find that the following energy equality should hold:

$$\boxed{\|\tilde{\mathbf{w}}_i\|^2 + \mathbf{e}_{a,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{a,i} = \|\tilde{\mathbf{w}}_{i-1-K'}\|^2 + \mathbf{e}_{p,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{p,i}}. \quad (9)$$

The important fact to emphasize is that *no approximations* are used to establish the energy relation (9); it is an exact relation that shows how the energies of the weight-error vectors at two successive iterations are related to the weighted energies of the *a priori* and *a posteriori* estimation error vectors. Relation (9) is the extension to the APA case of the energy-conservation relation originally derived in [12] and [13] in the context of robustness analysis and subsequently used in [15]–[18] in the context of steady-state and transient performance analysis. See also [15]

### B. Variance Relation for Steady-State Performance

The relevance of (9) to the mean-square analysis of affine projection algorithms can be seen as follows. Taking expectations of both sides of (9), we get

$$E[\|\tilde{\mathbf{w}}_i\|^2] + E[\mathbf{e}_{a,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{a,i}] = E[\|\tilde{\mathbf{w}}_{i-1-K'}\|^2] + E[\mathbf{e}_{p,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{p,i}]. \quad (10)$$

Taking the limit as  $i \rightarrow \infty$ , and using the steady-state condition  $E[\|\tilde{\mathbf{w}}_i\|^2] = E[\|\tilde{\mathbf{w}}_{i-K'}\|^2]$ , we obtain

$$E \left[ \mathbf{e}_{a,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{a,i} \right] = E \left[ \mathbf{e}_{p,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{p,i} \right]. \quad (11)$$

Substituting (6) into the right-hand side (RHS) of (11), we get

$$\begin{aligned} \text{RHS of (11)} &= E \left[ \mathbf{e}_{a,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{a,i} \right] \\ &\quad - \mu E \left[ \mathbf{e}_{a,i}^* B_i \mathbf{e}_i \right] - \mu E \left[ \mathbf{e}_i^* B_i \mathbf{e}_{a,i} \right] \\ &\quad + \mu^2 E \left[ \mathbf{e}_i^* A_i \mathbf{e}_i \right] \end{aligned} \quad (12)$$

where we are defining

$$A_i = (\epsilon I + U_i U_i^*)^{-1} U_i U_i^* (\epsilon I + U_i U_i^*)^{-1}$$

and

$$B_i = (\epsilon I + U_i U_i^*)^{-1}.$$

Using (12), equality (11) simplifies to

$$\boxed{\mu E \left[ \mathbf{e}_i^* A_i \mathbf{e}_i \right] = E \left[ \mathbf{e}_{a,i}^* B_i \mathbf{e}_i \right] + E \left[ \mathbf{e}_i^* B_i \mathbf{e}_{a,i} \right]} \quad (13)$$

as  $i \rightarrow \infty$ . This equation can now be used to evaluate the mean-square performance of affine projection algorithms.

### C. Mean-Square Performance

Introduce the noise vector

$$\mathbf{v}_i^T = [v(i) \ v(i-1) \ \dots \ v(i-(K-1)D)].$$

Then, (1) gives

$$\mathbf{e}_i = \mathbf{e}_{a,i} + \mathbf{v}_i$$

and under the often realistic assumption that

A.1) *the noise  $v(i)$  is i.i.d. and statistically independent of the regression matrix  $\{U_i\}$ .*

Neglecting the dependency of  $\tilde{\mathbf{w}}_{i-1}$  on past noises, we find that the variance relation (13) reduces to

$$\boxed{\mu E \left[ \mathbf{e}_{a,i}^* A_i \mathbf{e}_{a,i} \right] + \mu E \left[ \mathbf{v}_i^* A_i \mathbf{v}_i \right] = 2E \left[ \mathbf{e}_{a,i}^* B_i \mathbf{e}_{a,i} \right]} \quad (14)$$

as  $i \rightarrow \infty$ . This expression can be used to deduce an expression for the filter MSE or, equivalently, for the filter excess mean square error (EMSE), which is defined by

$$\text{EMSE} = \lim_{i \rightarrow \infty} E |e_a(i)|^2$$

where  $e_a(i) = \mathbf{u}_i^T \tilde{\mathbf{w}}_{i-1}$ . Now, from (1), we get

$$e(i) = e_a(i) + v(i)$$

and therefore, the MSE and EMSE define each other via

$$\text{MSE} = \text{EMSE} + \sigma_v^2.$$

In order to evaluate the EMSE, we need to deal with the expectations in (14). For this purpose, we shall rely on the following assumption.

A.2) *At steady-state,  $U_i$  is statistically independent of  $\mathbf{e}_{a,i}$ , and moreover,  $E[\mathbf{e}_{a,i} \mathbf{e}_{a,i}^*] = E|e_a(i)|^2 \cdot S$ , where  $S \approx I$  for small  $\mu$ , and  $S \approx (\mathbf{1} \cdot \mathbf{1}^T)$  for large  $\mu$ , where  $\mathbf{1}^T = [1 \ 0 \ \dots \ 0]$ .*

Note that since  $K' = 0$  for all algorithms listed in Table I, except PRA, then,  $e_a(i) = \mathbf{u}_i^T \tilde{\mathbf{w}}_{i-1}$ , and  $e_a(i)$  is the top entry of  $\mathbf{e}_{a,i}$ . For PRA,  $\mathbf{w}_{i-1} = \dots = \mathbf{w}_{i-K+1}$ , and therefore,  $e_a(i)$  is also equal to the top entry of  $\mathbf{e}_{a,i}$ . The condition on  $E[\mathbf{e}_{a,i} \mathbf{e}_{a,i}^*]$  is motivated in Appendix A. Using (14) and A.2), the first term on the left-hand side (LHS) of (14) becomes

$$\begin{aligned} \mu E \left[ \mathbf{e}_{a,i}^* A_i \mathbf{e}_{a,i} \right] &= \mu \text{Tr}(E[\mathbf{e}_{a,i} \mathbf{e}_{a,i}^* A_i]) \\ &= \mu E |e_a(i)|^2 \text{Tr}(S \cdot E[A_i]) \end{aligned} \quad (15)$$

as  $i \rightarrow \infty$ . Similar manipulations can be applied to the remaining terms in (14). Thus, we get

$$\mu E \left[ \mathbf{v}_i^* A_i \mathbf{v}_i \right] = \mu \sigma_v^2 \text{Tr}(E[A_i]) \quad (16)$$

and

$$2E \left[ \mathbf{e}_{a,i}^* B_i \mathbf{e}_{a,i} \right] = 2E |e_a(i)|^2 \text{Tr}(S \cdot E[B_i]) \quad (17)$$

as  $i \rightarrow \infty$ .

If we introduce the quantities (which are solely dependent on the statistics of the regression data):

$$\alpha_u \triangleq \text{Tr}(S \cdot E[A_i]) \quad \text{and} \quad \eta_u \triangleq \text{Tr}(S \cdot E[B_i]) \quad (18)$$

then (14) becomes

$$(2\eta_u - \mu\alpha_u) E |e_a(i)|^2 = \mu \sigma_v^2 \text{Tr}(E[A_i]) \quad (19)$$

as  $i \rightarrow \infty$ , and the EMSE of the filter is therefore given by

$$\boxed{\text{EMSE} = \frac{\mu \sigma_v^2 \text{Tr}(E[A_i])}{(2\eta_u - \mu\alpha_u)}} \quad (20)$$

and the steady-state MSE is

$$\boxed{\text{MSE} = \sigma_v^2 + \frac{\mu \sigma_v^2 \text{Tr}(E[A_i])}{(2\eta_u - \mu\alpha_u)}} \quad (21)$$

Two simplifications can be made when the regularization parameter  $\epsilon$  is small.

- If  $\epsilon$  is small enough so that its effect can be ignored, then  $A_i = B_i$ , and the definitions of  $\alpha_u$  and  $\eta_u$  will coincide. In this case, (20) reduces to

$$\text{EMSE} = \frac{\mu \sigma_v^2}{(2 - \mu)} \frac{\text{Tr}(E[A_i])}{\text{Tr}(S \cdot E[A_i])}. \quad (22)$$

If we use  $S \approx I$ , we obtain

$$\text{EMSE} = \frac{\mu \sigma_v^2}{(2 - \mu)}$$

and if we use  $S \approx (\mathbf{1} \cdot \mathbf{1}^T)$ , we get

$$\text{EMSE} = \frac{\mu\sigma_v^2}{(2-\mu)} \frac{\text{Tr}(E[A_i])}{E[A_i(1,1)]}.$$

- Another approximation assumes  $\epsilon$  is small and  $\mu$  is large and uses

$$\text{Tr}(E[A_i]) \approx E \left[ \frac{K}{\|\mathbf{u}_i\|^2} \right] \quad \text{and} \quad \frac{1}{E[A_i(1,1)]} \approx \text{Tr}(R_u)$$

to get

$$\text{EMSE} = \frac{\mu\sigma_v^2}{(2-\mu)} \text{Tr}(R_u) E \left[ \frac{K}{\|\mathbf{u}_i\|^2} \right]. \quad (23)$$

Note that this expression for the EMSE is proportional to  $K$ . In contrast, the expression given in [9] is

$$\text{EMSE} = \frac{\mu\sigma_v^2}{2-\mu} \text{Tr}(R_u) E \left[ \frac{1}{\|\mathbf{u}_i\|^2} \right]$$

which does not take into account the effect of  $K$ . Simulation results in Section VI (see Figs. 7–12) show that (22) and (23) provide good approximations for filter performance for relatively small step-size and order  $K$ .

#### IV. TRACKING PERFORMANCE OF APA

A similar analysis can be used to evaluate the performance of APA in nonstationary environments. Thus, assume that  $d(i) = \mathbf{u}_i \mathbf{w}_i^\circ + v(i)$ , where the unknown system  $\mathbf{w}_i^\circ$  is now time-variant. It is assumed that the variation in  $\mathbf{w}_i^\circ$  is according to the random-walk model (see, e.g., [1], [2], [15], and [19])

$$\mathbf{w}_i^\circ = \mathbf{w}_{i-1}^\circ + \mathbf{q}_i \quad (24)$$

where  $\mathbf{q}_i$  is an i.i.d. sequence with autocorrelation matrix  $Q = E[\mathbf{q}_i \mathbf{q}_i^*]$  and independent of the initial conditions  $\{\mathbf{w}_{-1}^\circ, \mathbf{w}_{-1}\}$  of the  $\{\mathbf{u}_j\}$  for all  $j$  and of the  $\{d(j)\}$  for all  $j < i$ . Let  $\tilde{\mathbf{w}}_i = \mathbf{w}_i^\circ - \mathbf{w}_i$ ,  $\mathbf{e}_{a,i} = U_i(\mathbf{w}_i^\circ - \mathbf{w}_{i-1-K})$ , and  $\mathbf{e}_{p,i} = U_i \tilde{\mathbf{w}}_i$ . Then

$$\mathbf{e}_{a,i} = U_i \left( \tilde{\mathbf{w}}_{i-1-K'} + \sum_{k=0}^{K'} \mathbf{q}_{i-k} \right)$$

and

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1-K'} + \sum_{k=0}^{K'} \mathbf{q}_{i-k} - \mu U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i. \quad (25)$$

If we multiply (25) by  $U_i$  from the left, we obtain that (6) still holds for the nonstationary case. Substituting (6) into (25), we get

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1-K'} + \sum_{k=0}^{K'} \mathbf{q}_{i-k} - U_i^* (U_i U_i^*)^{-1} (\mathbf{e}_{a,i} - \mathbf{e}_{p,i}). \quad (26)$$

Evaluating the energies of both sides of (26) and taking expectations, we find that

$$\begin{aligned} & E[\|\tilde{\mathbf{w}}_i\|^2] + E[\mathbf{e}_{a,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{a,i}] \\ &= E \left[ \left\| \tilde{\mathbf{w}}_{i-1-K'} + \sum_{k=0}^{K'} \mathbf{q}_{i-k} \right\|^2 \right] + E[\mathbf{e}_{p,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{p,i}]. \quad (27) \end{aligned}$$

Using the random-walk model (24), we know that  $E[\tilde{\mathbf{w}}_{i-1-K'}^* \mathbf{q}_{i-k}] = 0$  for  $k = 0, \dots, K'$ , and therefore

$$E \left[ \left\| \tilde{\mathbf{w}}_{i-1-K'} + \sum_{k=0}^{K'} \mathbf{q}_{i-k} \right\|^2 \right] = E[\|\tilde{\mathbf{w}}_{i-1-K'}\|^2] + (1+K')\text{Tr}(Q). \quad (28)$$

Substituting into (27), we obtain

$$\begin{aligned} & E[\|\tilde{\mathbf{w}}_i\|^2] + E[\mathbf{e}_{a,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{a,i}] = E[\|\tilde{\mathbf{w}}_{i-1-K'}\|^2] \\ & \quad + (1+K')\text{Tr}(Q) + E[\mathbf{e}_{p,i}^* (U_i U_i^*)^{-1} \mathbf{e}_{p,i}]. \quad (29) \end{aligned}$$

Comparing with (10), we see that the only difference in the nonstationary case is the appearance of the additional term  $(1+K')\text{Tr}(Q)$ . Note that the other terms are identical. Therefore, similar manipulations to those in Section III lead to

$$(2\eta_u - \mu\alpha_u) E|e_a(i)|^2 = \mu\sigma_v^2 \text{Tr}(E[A_i]) + \frac{(1+K')\text{Tr}(Q)}{\mu} \quad (30)$$

as  $i \rightarrow \infty$ , and the EMSE is then given by

$$\boxed{\text{EMSE} = \frac{\mu\sigma_v^2 \text{Tr}(E[A_i]) + \mu^{-1}(1+K')\text{Tr}(Q)}{(2\eta_u - \mu\alpha_u)}} \quad (31)$$

The two simplifications of Section III can be used to get

$$\frac{1}{(2-\mu)} \left( \mu\sigma_v^2 + \frac{\mu^{-1}(1+K')\text{Tr}(Q)}{\text{Tr}(E[A_i])} \right) \quad (32)$$

or

$$\frac{\text{Tr}(R_u)}{(2-\mu)} \left( \mu\sigma_v^2 E \left[ \frac{K}{\|\mathbf{u}_i\|^2} \right] + \mu^{-1}(1+K')\text{Tr}(Q) \right). \quad (33)$$

From (32) and (33), we see that for a given  $K$ , there is an optimal  $\mu$  that minimizes the EMSE, and for a given  $\mu$ , there is an optimal  $K$  that minimizes the EMSE. Comparisons of the tracking performance among the APA family are given in Table II.

#### V. TRANSIENT ANALYSIS OF APA

We now study the transient (i.e., convergence and stability) performance of the APA family. This task is more challenging than mean-square performance. Nevertheless, the same energy conservation arguments of the previous section can still be used if we incorporate weighting into the energy relation and into the definition of the error quantities [14], [17], as we now explain. We will assume, without loss of generality, that  $\alpha = 0$ . Then, (2) becomes

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i.$$

In the following analysis, if we substitute  $\tilde{\mathbf{w}}_{i-1}$  by  $\tilde{\mathbf{w}}_{i-1-K'}$ , then the results for  $\alpha \neq 0$  would be obtained.

##### A. Weighted Energy Relation

Let  $\mathbf{e}_{p,i}^\Sigma = U_i \Sigma \tilde{\mathbf{w}}_i$  and  $\mathbf{e}_{a,i}^\Sigma = U_i \Sigma \tilde{\mathbf{w}}_{i-1}$ . If we multiply both sides of the above recursion by  $U_i \Sigma$  from the left, for any Hermitian positive-definite matrix  $\Sigma$ , we find that the *a priori* and *a posteriori* estimation errors  $\{\mathbf{e}_{p,i}^\Sigma, \mathbf{e}_{a,i}^\Sigma\}$  are related via

$$\mathbf{e}_{p,i}^\Sigma = \mathbf{e}_{a,i}^\Sigma - \mu U_i \Sigma U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i. \quad (34)$$

TABLE II  
EMSE OF APA FAMILY IN NONSTATIONARY ENVIRONMENTS WHERE  $A_i \approx (U_i U_i^*)^{-1}$

Algorithms	EMSE
NLMS	$\frac{1}{(2-\mu)} (\mu\sigma_v^2 + \mu^{-1}\text{Tr}(Q)\text{Tr}(R_u))$ <p style="text-align: center;">or</p> $\frac{\text{Tr}(R_u)}{(2-\mu)} \left( \mu\sigma_v^2 E \left[ \frac{1}{\ \mathbf{u}_i\ ^2} \right] + \mu^{-1}\text{Tr}(Q) \right)$
BNDR-LMS	$\frac{1}{(2-\mu)} (\mu\sigma_v^2 + \mu^{-1}\text{Tr}(Q)\text{Tr}(R_u))$ <p style="text-align: center;">or</p> $\frac{\text{Tr}(R_u)}{(2-\mu)} \left( 2\mu\sigma_v^2 E \left[ \frac{1}{\ \mathbf{u}_i\ ^2} \right] + \mu^{-1}\text{Tr}(Q) \right)$
APA, R-APA, NLMS-OCF (small $\epsilon$ )	$\frac{1}{(2-\mu)} \left( \mu\sigma_v^2 + \frac{\mu^{-1}\text{Tr}(Q)}{\text{Tr}(E[A_i])} \right)$ <p style="text-align: center;">or</p> $\frac{\text{Tr}(R_u)}{(2-\mu)} \left( \mu\sigma_v^2 E \left[ \frac{K}{\ \mathbf{u}_i\ ^2} \right] + \mu^{-1}\text{Tr}(Q) \right)$
PRA (small $\epsilon$ )	$\frac{1}{(2-\mu)} \left( \mu\sigma_v^2 + \frac{\mu^{-1}K\text{Tr}(Q)}{\text{Tr}(E[A_i])} \right)$ <p style="text-align: center;">or</p> $\frac{\text{Tr}(R_u)}{(2-\mu)} \left( \mu\sigma_v^2 E \left[ \frac{K}{\ \mathbf{u}_i\ ^2} \right] + \mu^{-1}K\text{Tr}(Q) \right)$

Similarly to the arguments in Section III, we can get

$$\boxed{\tilde{\mathbf{w}}_i + U_i^*(U_i \Sigma U_i^*)^{-1} \mathbf{e}_{a,i}^\Sigma = \tilde{\mathbf{w}}_{i-1} + U_i^*(U_i \Sigma U_i^*)^{-1} \mathbf{e}_{p,i}^\Sigma} \quad (35)$$

On each side of this identity, we have a combination of *a priori* and *a posteriori* errors. If we equate the weighted Euclidean norms of both sides of (35), we find that

$$\|\tilde{\mathbf{w}}_i\|_\Sigma^2 + \mathbf{e}_{a,i}^{*\Sigma} (U_i \Sigma U_i^*)^{-1} \mathbf{e}_{a,i}^\Sigma = \|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 + \mathbf{e}_{p,i}^{*\Sigma} (U_i \Sigma U_i^*)^{-1} \mathbf{e}_{p,i}^\Sigma. \quad (36)$$

The special choice  $\Sigma = I$  reduces to the energy relation (9). Moreover, since

$$\mathbf{e}_i = U_i \tilde{\mathbf{w}}_{i-1} + \mathbf{v}_i$$

we also get

$$\tilde{\mathbf{w}}_i = (I - \mu U_i^*(\epsilon I + U_i U_i^*)^{-1} U_i) \tilde{\mathbf{w}}_{i-1} - \mu U_i^*(\epsilon I + U_i U_i^*)^{-1} \mathbf{v}_i. \quad (37)$$

### B. Weighted Variance Relation

In transient analysis, we are interested in the time evolution of  $E\|\tilde{\mathbf{w}}_i\|_\Sigma^2$  for some desirable choices of  $\Sigma$ . For this reason, rather than eliminate the effect of the weight-error vector, the contributions of the other error quantities  $\{\mathbf{e}_{a,i}^\Sigma, \mathbf{e}_{p,i}^\Sigma, \mathbf{e}_i\}$  are instead expressed in terms of the weight-error vector itself. In so doing, the energy relation (36) will lead to a recursion that describes the evolution of  $E\|\tilde{\mathbf{w}}_i\|_\Sigma^2$ .

Replacing  $\mathbf{e}_{p,i}^\Sigma$  by its equivalent expression in (34), we get

$$\begin{aligned} \|\tilde{\mathbf{w}}_i\|_\Sigma^2 &= \|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 \\ &\quad + \mu^2 \mathbf{e}_i^*(\epsilon I + U_i U_i^*)^{-1} U_i \Sigma U_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i \\ &\quad - \mu \mathbf{e}_{a,i}^{*\Sigma} (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_i^* \\ &\quad - \mu \mathbf{e}_i^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_{a,i}^\Sigma. \end{aligned} \quad (38)$$

Using the relation  $\mathbf{e}_i = \mathbf{e}_{a,i} + \mathbf{v}_i$ , we can eliminate  $\mathbf{e}_i$ . Since most of the factors disappear under A.1) and expectation, we get

$$\begin{aligned} E[\|\tilde{\mathbf{w}}_i\|_\Sigma^2] &= E[\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2] \\ &\quad - \mu E[\mathbf{e}_{a,i}^{*\Sigma} (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_{a,i}^\Sigma] \\ &\quad - \mu E[\mathbf{e}_{a,i}^* (\epsilon I + U_i U_i^*)^{-1} \mathbf{e}_{a,i}^\Sigma] \\ &\quad + \mu^2 E[\mathbf{e}_{a,i}^* A_i^\Sigma \mathbf{e}_{a,i}^\Sigma] + \mu^2 E[\mathbf{v}_i^* A_i^\Sigma \mathbf{v}_i] \end{aligned} \quad (39)$$

where

$$A_i^\Sigma \triangleq (\epsilon I + U_i U_i^*)^{-1} U_i \Sigma U_i^* (\epsilon I + U_i U_i^*)^{-1}.$$

In addition,  $\{\mathbf{e}_{a,i}^\Sigma, \mathbf{e}_{a,i}^{*\Sigma}\}$  can be expressed in terms of  $\tilde{\mathbf{w}}_{i-1}$ . Thus, we have

$$E[\|\tilde{\mathbf{w}}_i\|_\Sigma^2] = E[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2] + \mu^2 E[\mathbf{v}_i^* A_i^\Sigma \mathbf{v}_i] \quad (40)$$

where

$$\begin{aligned} \Sigma' &\triangleq \Sigma - \mu \Sigma U_i^* (\epsilon I + U_i U_i^*)^{-1} U_i \\ &\quad - \mu U_i^* (\epsilon I + U_i U_i^*)^{-1} U_i \Sigma + \mu^2 (U_i^* A_i^\Sigma U_i). \end{aligned} \quad (41)$$

Recursion (40) provides a compact characterization of the time evolution of the weight-error variance. However, recursion

(40) is still hard to propagate due to the presence of the expectation

$$E [\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2] = E [\tilde{\mathbf{w}}_{i-1}^* \Sigma' \tilde{\mathbf{w}}_{i-1}].$$

This expectation is difficult to evaluate due to the dependence of  $\Sigma'$  on  $U_i$  and of  $\tilde{\mathbf{w}}_{i-1}$  on prior regressors. One way to overcome this difficulty is to introduce an *independence assumption* on the regressor sequence  $U_i$ , namely, to assume the following.

A.3) *The matrix sequence  $\{U_i\}$  is independent and identically distributed.*

This assumption guarantees that  $\tilde{\mathbf{w}}_{i-1}$  is independent of both  $\Sigma'$  and  $U_i$ . Clearly, A.3) is a strong assumption (it is actually stronger than the usual independence assumption, which only requires the  $\{\mathbf{u}_i\}$  to be i.i.d [1], [2]). Observe, however, from (41) for  $\Sigma'$  that it is sufficient for our purposes to require the following:

A.3')  *$\tilde{\mathbf{w}}_{i-1}$  is independent of  $U_i^*(\epsilon I + U_i U_i^*)^{-1} U_i$ .*

This is generally a weaker assumption. In this way, recursion (40) reduces to

$$E [\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2] = E [\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2] + \mu^2 E [\mathbf{v}_i^* A_i^{\Sigma} \mathbf{v}_i] \quad (42)$$

where now

$$\Sigma' = \Sigma - \mu \Sigma E [U_i^*(\epsilon I + U_i U_i^*)^{-1} U_i] - \mu E [U_i^*(\epsilon I + U_i U_i^*)^{-1} U_i] \Sigma + \mu^2 E [U_i^* A_i^{\Sigma} U_i] \quad (43)$$

with expectations appearing in (43). In addition, taking expectations of both sides of (37) and using assumption A.1), we obtain the following result for the evolution of the mean of the weight-error vector:

$$E [\tilde{\mathbf{w}}_i] = E [I - \mu U_i^*(\epsilon I + U_i U_i^*)^{-1} U_i] E [\tilde{\mathbf{w}}_{i-1}] \quad (44)$$

Relations (42) and (44) can be used to derive conditions for mean-square stability, as well as expressions for the steady-state MSE and mean-square deviation (MSD) of the APA family. To see this, we introduce some notation. The  $\text{vec}\{\cdot\}$  notation, e.g.,  $\sigma = \text{vec}(\Sigma)$ , allows us to replace an  $M \times M$  arbitrary matrix  $\Sigma$  by an  $M^2 \times 1$  column vector  $\sigma$  whose entries are formed by stacking the successive columns of the matrix on top of each other. On the other hand, writing  $\text{vec}\{\sigma\}$  for an  $M^2 \times 1$  column vector  $\sigma$  results in an  $M \times M$  matrix  $\Sigma$  whose entries are obtained from  $\sigma$ . Therefore, we also write  $\Sigma = \text{vec}\{\sigma\}$ . The  $\text{vec}\{\cdot\}$  notation is convenient when working with Kronecker products. The Kronecker product of two matrices  $A$  and  $B$ , say of dimensions  $m_a \times n_a$  and  $m_b \times n_b$ , respectively, is denoted by  $A \otimes B$  [20]. For any matrices  $\{P, \Sigma, Q\}$  of compatible dimensions, it holds that

$$\text{vec}\{P\Sigma Q\} = (Q^T \otimes P) \text{vec}(\Sigma). \quad (45)$$

Applying (45) to (40), we find that it leads to the vector relation

$$\sigma' = F\sigma \quad (46)$$

where the coefficient matrix  $F$  is  $M^2 \times M^2$  and defined by

$$F \triangleq I - \mu(E[P_i^T] \otimes I + I \otimes E[P_i]) + \mu^2 E[P_i^T \otimes P_i] \quad (47)$$

with

$$P_i = U_i^*(\epsilon I + U_i U_i^*)^{-1} U_i.$$

We can rewrite the recursion for  $E [\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2]$  in (40) by using the vectors  $\{\sigma', \sigma\}$  instead of the matrices  $\{\Sigma', \Sigma\}$ , say, as

$$E [\|\tilde{\mathbf{w}}_i\|_{\text{vec}\{\sigma\}}^2] = E [\|\tilde{\mathbf{w}}_{i-1}\|_{\text{vec}\{\sigma'\}}^2] + \mu^2 \sigma_v^2 (\gamma^T \sigma) \quad (48)$$

where, for the last term, we used the fact that

$$\text{Tr} (E [(\epsilon I + U_i U_i^*)^{-1} U_i \Sigma U_i^* (\epsilon I + U_i U_i^*)^{-1}]) = \gamma^T \sigma$$

where  $\gamma \equiv \text{vec}\{E [U_i^*(\epsilon I + U_i U_i^*)^{-1} U_i]\}$ . For compactness of notation, we drop the  $\text{vec}\{\cdot\}$  notation from the subscripts and keep the vectors so that the above is simply rewritten as

$$E [\|\tilde{\mathbf{w}}_i\|_{\sigma}^2] = E [\|\tilde{\mathbf{w}}_{i-1}\|_{F\sigma}^2] + \mu^2 \sigma_v^2 (\gamma^T \sigma) \quad (49)$$

In addition, we obtain the following result for the evolution of the mean of the weight-error vector:

$$E [\tilde{\mathbf{w}}_i] = (I - \mu E[P_i]) E [\tilde{\mathbf{w}}_{i-1}] \quad (50)$$

Recursion (49) shows that in order to evaluate  $E [\|\tilde{\mathbf{w}}_i\|_{\sigma}^2]$ , we need to know  $E [\|\tilde{\mathbf{w}}_{i-1}\|_{F\sigma}^2]$ , with a weighting matrix whose entries are determined by  $F\sigma$ . Now, the quantity  $E [\|\tilde{\mathbf{w}}_i\|_{F\sigma}^2]$  can be inferred from (49) by writing the recursion for  $F\sigma$ , i.e.,

$$E [\|\tilde{\mathbf{w}}_i\|_{F\sigma}^2] = E [\|\tilde{\mathbf{w}}_{i-1}\|_{F^2\sigma}^2] + \mu^2 \sigma_v^2 (\gamma^T F\sigma).$$

We again find that in order to evaluate  $E [\|\tilde{\mathbf{w}}_i\|_{F\sigma}^2]$ , we need to know  $E [\|\tilde{\mathbf{w}}_{i-1}\|_{F^2\sigma}^2]$ . The natural question is whether this procedure terminates. Fortunately, as in [14] and [17], this procedure does terminate. This is because once we write (48) by substituting  $\sigma$  by  $F^{M^2-1}\sigma$ , we get

$$E [\|\tilde{\mathbf{w}}_i\|_{F^{M^2-1}\sigma}^2] = E [\|\tilde{\mathbf{w}}_{i-1}\|_{F^{M^2}\sigma}^2] + \mu^2 \sigma_v^2 (\gamma^T F^{M^2-1}\sigma)$$

where the weighting matrix on the RHS is  $F^{M^2}\sigma$ . This term can be deduced from the prior weighting factors. Indeed, let  $p(x)$  denote the characteristic polynomial of  $F$ ,

$$p(x) \equiv \det(xI - F).$$

It is a polynomial of order  $M^2$  in  $x$

$$p(x) = x^{M^2} + p_{M^2-1} x^{M^2-1} + \dots + p_1 x + p_0$$

with coefficients  $\{p_k\}$ . Now, the Cayley–Hamilton theorem guarantees that  $p(F) = 0$  so that

$$E [\|\tilde{\mathbf{w}}_i\|_{F^{M^2-1}\sigma}^2] = - \sum_{m=0}^{M^2-1} p_m E [\|\tilde{\mathbf{w}}_i\|_{F^m\sigma}^2]. \quad (51)$$

*Theorem 1 [Transient Performance]:* Under assumptions A.1) and A.3'), the transient performance of the APA family (2) for  $\alpha = 0$  is described by the state recursion

$$\mathcal{W}_i = \mathcal{F} \mathcal{W}_{i-1} + \mu^2 \sigma_v^2 \mathcal{Y} \quad (52)$$

where

$$\mathcal{F} = \begin{bmatrix} 0 & 1 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -p_0 & -p_1 & -p_2 & \dots & -p_{M^2-1} \end{bmatrix}$$

$$\mathcal{W}_i = \begin{bmatrix} E[\|\tilde{\mathbf{w}}_i\|_{\sigma}^2] \\ E[\|\tilde{\mathbf{w}}_i\|_{F\sigma}^2] \\ \vdots \\ E[\|\tilde{\mathbf{w}}_i\|_{F^{M^2-2}\sigma}^2] \\ E[\|\tilde{\mathbf{w}}_i\|_{F^{M^2-1}\sigma}^2] \end{bmatrix}, \quad \mathcal{Y} = \begin{bmatrix} \gamma^T \sigma \\ \gamma^T F \sigma \\ \vdots \\ \gamma^T F^{M^2-2} \sigma \\ \gamma^T F^{M^2-1} \sigma \end{bmatrix}$$

$\gamma = \text{vec}\{E[U_i^*(\epsilon I + U_i U_i^*)^{-2} U_i]\}$ ,  $\sigma = \text{vec}\{\Sigma\}$ ,  $F = I - \mu(E[P_i^T] \otimes I + I \otimes E[P_i]) + \mu^2 E[P_i^T \otimes P_i]$ , and  $p_i$  are coefficients of the characteristic polynomial of  $F$ .  $\square$

Observe that the eigenvalues of  $\mathcal{F}$  coincide with those of  $F$ .

### C. Learning Curves

The learning curve of an adaptive filter describes the time evolution of the variance  $E|e_a(i)|^2$ . Now, if the  $\mathbf{u}_i$  are assumed to be i.i.d., then

$$E|e_a(i)|^2 = E|\mathbf{u}_i \tilde{\mathbf{w}}_{i-1}|^2 = E\|\tilde{\mathbf{w}}_{i-1}\|_{R_u}^2$$

and the learning curve can be evaluated by computing  $E\|\tilde{\mathbf{w}}_{i-1}\|_{R_u}^2$  for each  $i$ . This task can be accomplished recursively from (48) by iterating it and setting  $\sigma = \mathbf{r} = \text{vec}(R_u)$ . This yields

$$E[\|\tilde{\mathbf{w}}_i\|_{\mathbf{r}}^2] = E[\|\tilde{\mathbf{w}}_{i-1}\|_{F^i \mathbf{r}}^2] + \mu^2 \sigma_v^2 (\gamma^T (I + \dots + F^{i-1}) \mathbf{r}). \quad (53)$$

That is

$$E[\|\tilde{\mathbf{w}}_{i-1}\|_{\mathbf{r}}^2] = E[\|\tilde{\mathbf{w}}_{i-1}\|_{\mathbf{f}_{i-1}}^2] + \mu^2 \sigma_v^2 g(i-1) \quad (54)$$

where the vector  $\mathbf{f}_i$  and the scalar  $g(i)$  satisfy the recursions

$$\begin{aligned} \mathbf{f}_{i-1} &= F \mathbf{f}_{i-2}, \quad \mathbf{f}_0 = \mathbf{r} \\ g(i-1) &= g(i-2) + \gamma^T \mathbf{f}_{i-1}, \quad g(-1) = 0. \end{aligned}$$

### D. Mean-Square Stability

From (50), the convergence in the mean of the APA family is guaranteed for any  $\mu$  satisfying

$$\mu < \frac{2}{\lambda_{\max}(E[P_i])}. \quad (55)$$

Moreover, recursion (49) is stable if, and only if, the matrix  $F$  is stable. Thus, let  $C = E[P_i^T] \otimes I + I \otimes E[P_i]$  and  $D = E[P_i^T \otimes P_i]$  so that  $F = I - \mu C + \mu^2 D$ . The following holds.

*Theorem 2 [Stability]:* The convergence in the mean-square sense of the APA family is guaranteed for any  $\mu$  in the range

$$0 < \mu < \min \left\{ \frac{1}{\lambda_{\max}(C^{-1}D)}, \frac{1}{\max(\lambda(H) \in \mathfrak{R}^+)} \right\}$$

where  $C = E[P_i^T] \otimes I + I \otimes E[P_i]$ ,  $D = E[P_i^T \otimes P_i]$ , and  $H = \begin{bmatrix} \frac{1}{2}C & -\frac{1}{2}D \\ I & 0 \end{bmatrix}$ .  $\square$

The above condition on  $\mu$  is in terms of the largest positive eigenvalue of  $H$  when it exists. The theorem is proved Ap-

TABLE III  
STABILITY BOUNDS COMPUTED BY THEOREM II (GAUSSIAN INPUT)

	$\frac{2}{\lambda_{\max}(E[P_i])}$	$\frac{1}{\lambda_{\max}(C^{-1}D)}$	$\frac{1}{\max(\lambda(H) \in \mathfrak{R}^+)}$	$\mu_{\max}$
$K=1$	4.3419	2.0003	2.9624	2.0003
$K=2$	4.0328	2.0006	3.1241	2.0006
$K=4$	3.3043	2.0026	3.7518	2.0026
$K=8$	2.6053	2.0014	3.5945	2.0014

TABLE IV  
STABILITY BOUNDS COMPUTED BY THEOREM II (UNIFORM INPUT)

	$\frac{2}{\lambda_{\max}(E[P_i])}$	$\frac{1}{\lambda_{\max}(C^{-1}D)}$	$\frac{1}{\max(\lambda(H) \in \mathfrak{R}^+)}$	$\mu_{\max}$
$K=1$	13.6903	2.0001	7.0952	2.0001
$K=2$	9.8496	2.0002	5.3601	2.0002
$K=4$	6.0767	2.0003	3.7029	2.0003
$K=8$	3.3642	2.0006	4.2959	2.0006

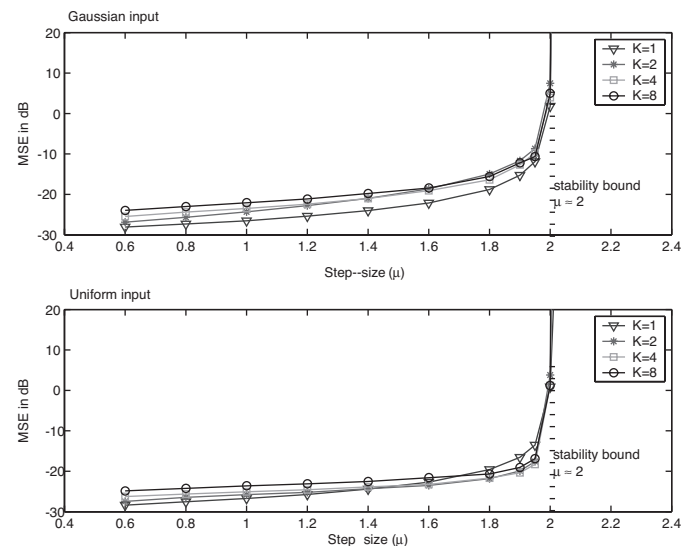


Fig. 1. Simulated MSE of APA as a function of the step size.

pendix B. By combining (55) and Theorem 2, a bound on the step-size for both mean and mean-square stability is obtained.

Theorem 2 provides an explicit and unified stability bound for a general class of input signals and various affine projection algorithms.

### E. Steady-State Behavior

In the above, we used the variance relation (49) to characterize the transient behavior of the APA family in terms of a state recursion. We can use the same variance relation to shed further light on the mean-square performance of the APA family. In particular, we shall re-examine the EMSE, as well as study the mean-square deviation (MSD), which is defined as

$$\text{MSD} = \lim_{i \rightarrow \infty} E\|\tilde{\mathbf{w}}_i\|^2.$$

Assuming the step-size  $\mu$  is chosen to guarantee filter stability, recursion (49) becomes in steady-state

$$E[\|\tilde{\mathbf{w}}_{\infty}\|_{\sigma}^2] = E[\|\tilde{\mathbf{w}}_{\infty}\|_{F\sigma}^2] + \mu^2 \sigma_v^2 (\gamma^T \sigma) \quad (56)$$

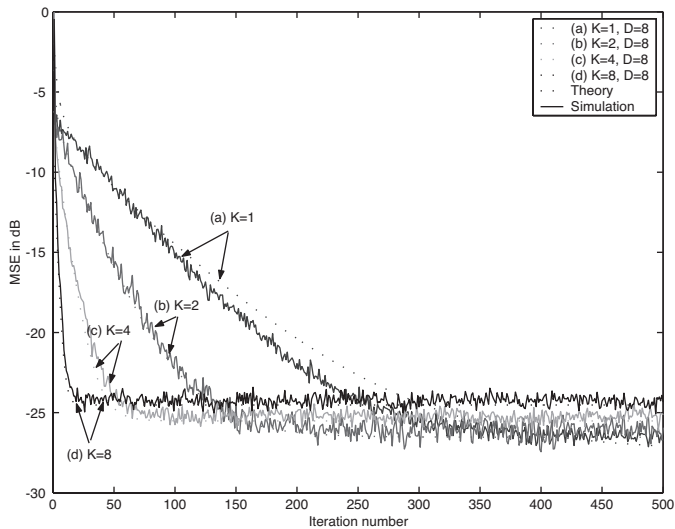


Fig. 2. Learning curves of the APA family for colored Gaussian input using  $\mu = 1.0$  and  $D = 8$ . (a)  $K = 1$ . (b)  $K = 2$ . (c)  $K = 4$ . (d)  $K = 8$  [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

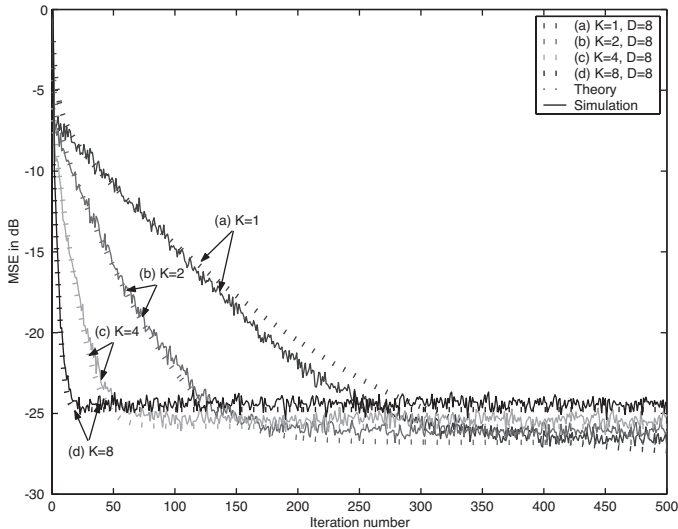


Fig. 3. Learning curves of the APA family for colored uniform input using  $\mu = 1.0$  and  $D = 8$ . (a)  $K = 1$ . (b)  $K = 2$ . (c)  $K = 4$ . (d)  $K = 8$  [Input: uniform AR(1), pole at 0.5. System: FIR (16)].

which is equivalent to

$$E \left[ \|\tilde{\mathbf{w}}_\infty\|_{(I-F)\sigma}^2 \right] = \mu^2 \sigma_v^2 (\gamma^T \sigma). \quad (57)$$

We choose  $\sigma$  to reduce the weight into the identity matrix. Thus, it needs to be selected as the solution to the linear system of equations  $(I - F)\sigma = \text{vec}\{I\}$ , i.e.,  $\sigma = (I - F)^{-1} \text{vec}\{I\}$ . In this case, the weighting quantity that appears in (57) reduces to the vector of unit entries. Then, the left-hand side of (57) becomes the filter MSD, and (57) leads to

$$\text{MSD} = \mu^2 \sigma_v^2 \gamma^T (I - F)^{-1} \text{vec}\{I\} \quad (58)$$

In a similar way, let us evaluate the EMSE of the APA family. Note that since

$$E|e_a(i)|^2 = E \left[ \|\tilde{\mathbf{w}}_{i-1}\|_{R_u}^2 \right]$$

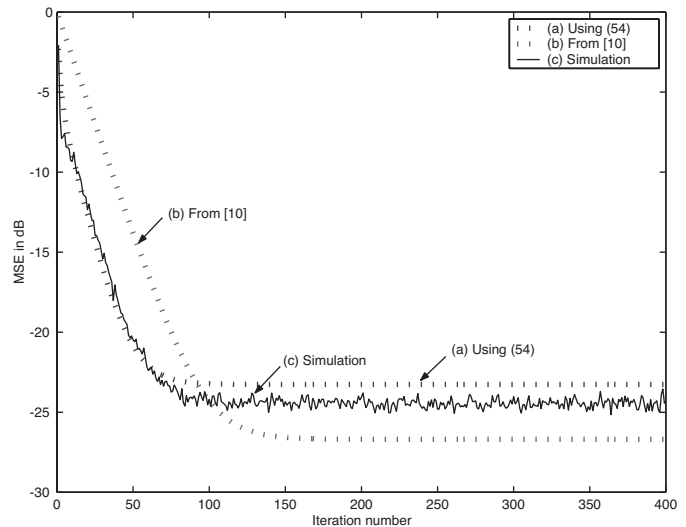


Fig. 4. Comparison of learning curves for colored Gaussian input using  $K = 2$ ,  $\mu = 1.0$ , and  $D = 8$ . (a) Using (54). (b) Using the results of [10]. (c) Simulation [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

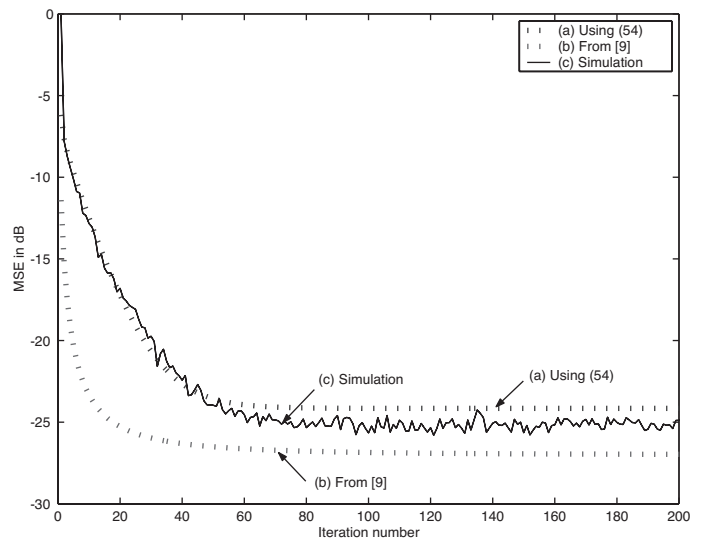


Fig. 5. Comparison of learning curves for colored Gaussian input using  $K = 4$ ,  $\mu = 1.0$ , and  $D = 8$ . (a) Using (54). (b) Using the results of [9]. (c) Simulation [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

we need to evaluate  $E \left[ \|\tilde{\mathbf{w}}_\infty\|_{\mathbf{r}}^2 \right]$ , where the weighting factor is  $\mathbf{r} = \text{vec}\{R_u\}$ . Assume we select  $\sigma$  as the solution to the linear system of equations  $(I - F)\sigma = \mathbf{r}$ . In this case, the weighting quantity that appears in (57) reduces to  $R_u$ . Then, the LHS of (57) becomes the filter EMSE, and (57) leads to the desired result

$$\text{EMSE} = \mu^2 \sigma_v^2 \gamma^T (I - F)^{-1} \text{vec}\{R_u\} \quad (59)$$

## VI. SIMULATION RESULTS

We illustrate the theoretical results presented in this paper by carrying out computer simulations in a channel estimation scenario. The unknown channel has 16 taps and is randomly generated. Two different types of signals, viz., Gaussian and



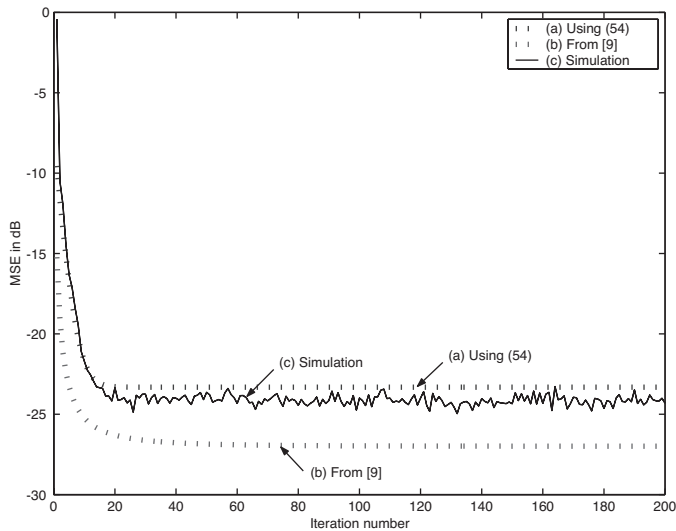


Fig. 6. Comparison of learning curves for colored Gaussian input using  $K = 8$ ,  $\mu = 1.0$ , and  $D = 8$ . (a) Using (54). (b) Using the results of [9]. (c) Simulation [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

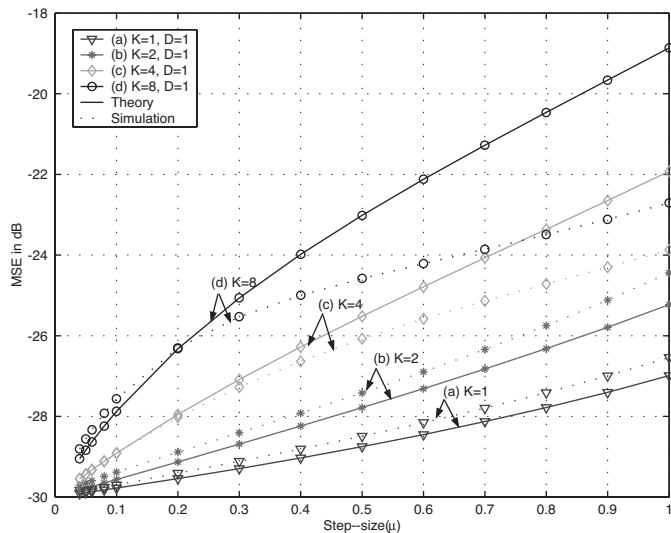


Fig. 7. Steady-state MSE curves of the APA family for colored Gaussian input using  $D = 1$  in stationary environments. (a)  $K = 1$ . (b)  $K = 2$ . (c)  $K = 4$ . (d)  $K = 8$  [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

uniformly distributed signals, are used for the input signal  $u(i)$ , viz.,

$$u(i) = \tau u(i-1) + \rho(i)$$

which is a first-order autoregressive (AR) process with a pole at  $\tau$ . For the Gaussian case,  $\rho(i)$  is a white, zero-mean, Gaussian random sequence having unit variance, and  $\tau$  is set to 0.9. As a result, a highly colored Gaussian signal is generated. For the uniform case,  $\rho(i)$  is a uniform random sequence between  $-1.0$  and  $1.0$ , and  $\tau$  is set to 0.5. In Tables III and IV we evaluate the bounds in (55) and Theorem 2. These tables indicate that the stability bound on  $\mu$  is approximately  $0 < \mu < 2$  for both Gaussian input (which is consistent with [9] and uniform input signals). This fact is further verified by simulation in Fig. 1,

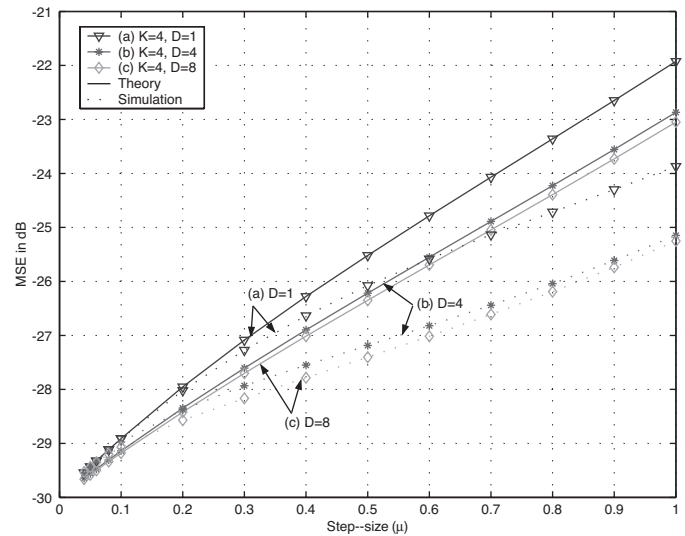


Fig. 8. Steady-state MSE curves of the APA family for colored Gaussian input using  $K = 4$  in stationary environments. (a)  $D = 1$ . (b)  $D = 4$ . (c)  $D = 8$  [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

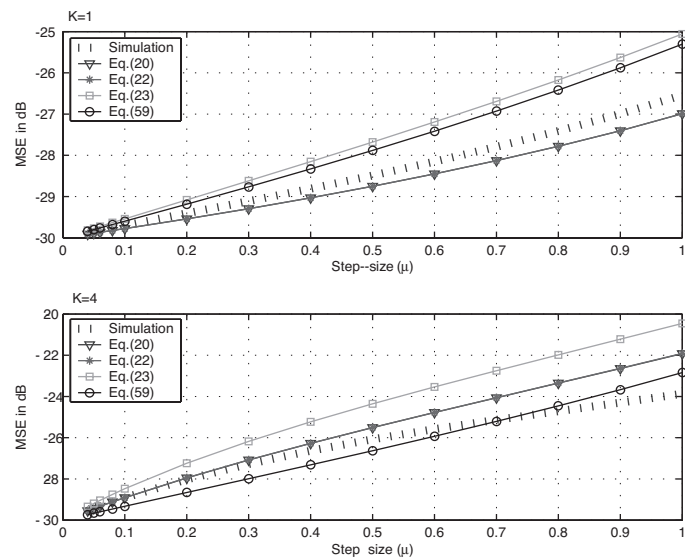


Fig. 9. Comparison of MSE expressions when  $K = 1$  or  $K = 4$  and  $D = 1$  [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

where MSE curves are plotted as a function of the step size. The expectations involved in evaluating  $C$  and  $D$  are estimated via ensemble averaging.

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

$$\text{SNR} = 10 \log \left( \frac{E[y^2(i)]}{E[v^2(i)]} \right)$$

where  $y(i) = \mathbf{u}_i \mathbf{w}^o$ . The measurement noise  $v(i)$  is added to  $y(i)$  such that  $\text{SNR} = 30$  dB. The adaptive filter and the unknown channel are assumed to have the same number of taps. All adaptive filter coefficients are initialized to zero. In addition, the regularization parameter  $\epsilon$  is set to 0.001. We set  $\alpha = 0$ . The simulation results shown are obtained by ensemble averaging over 200 independent trials.

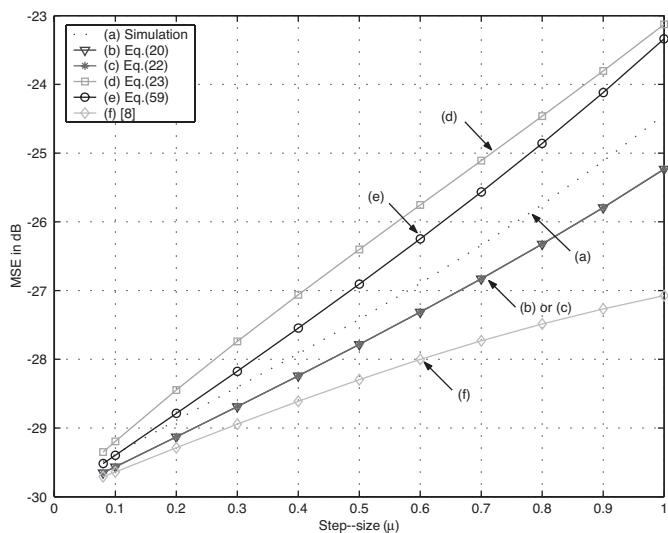


Fig. 10. Comparison of MSE when  $K = 2$  and  $D = 1$  [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

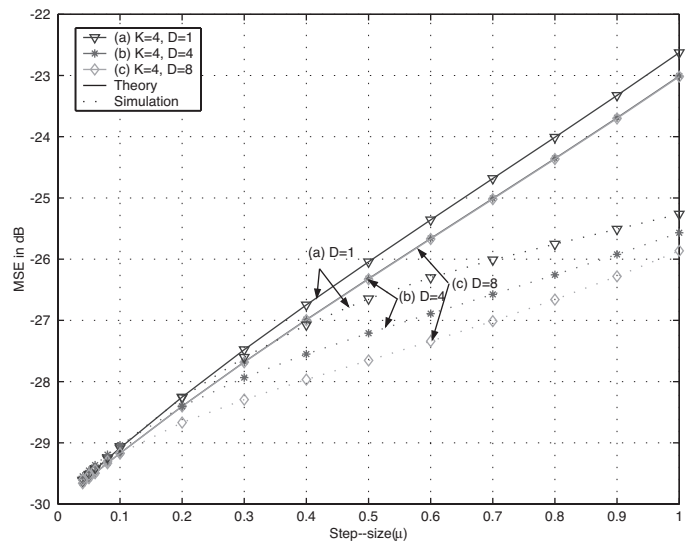


Fig. 12. Steady-state MSE curves of the APA family for colored uniform input using  $K = 4$  in stationary environments. (a)  $D = 1$ . (b)  $D = 4$ . (c)  $D = 8$  [Input: uniform AR(1), pole at 0.5. System: FIR (16)].

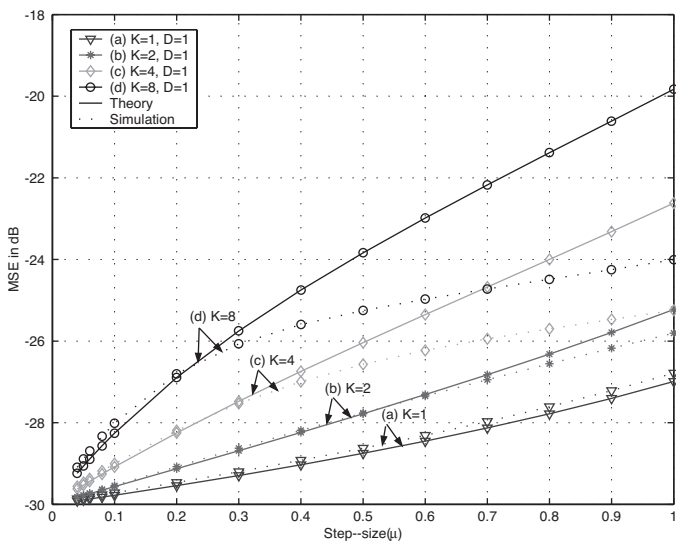


Fig. 11. Steady-state MSE curves of the APA family for colored uniform input using  $D = 1$  in stationary environments. (a)  $K = 1$ . (b)  $K = 2$ . (c)  $K = 4$ . (d)  $K = 8$  [Input: uniform AR(1), pole at 0.5. System: FIR (16)].

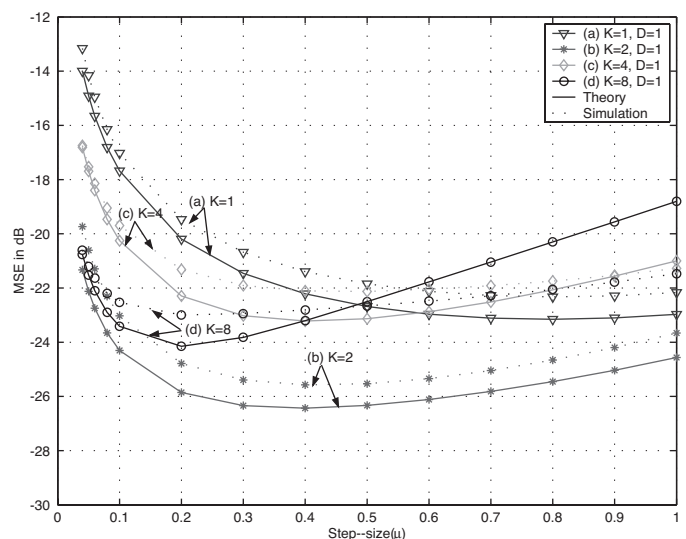


Fig. 13. Steady-state MSE curves of the APA family for colored Gaussian input using  $D = 1$  in nonstationary environments. (a)  $K = 1$ . (b)  $K = 2$ . (c)  $K = 4$ . (d)  $K = 8$  [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

### A. Transient Performance

Figs. 2–6 show the learning curves of the APA family. The step size is set to  $\mu = 1$ , and the delay parameter is set to 8. Fig. 2 shows how close the simulation results are to the theoretical results (54), where  $F$  and  $\gamma$  were evaluated via ensemble averaging. The theoretical results are very close to the simulated results, although there is some discrepancy when  $K = 1$ . In Fig. 3, the colored uniform input signal is used for the simulation. For generating the input signal,  $\tau$  is set to 0.5, unlike the Gaussian case. In Figs. 4–6, the learning curves in Fig. 2 are compared with the theoretical results in [9] and [10].

### B. Steady-State Performance

Fig. 7 shows the steady-state MSE curves of the APA family for colored Gaussian input as a function of the step size. The step size  $\mu$  varies from 0.04 to 1.0. This range guarantees stability as

mentioned before. The theoretical results are calculated using (22), and the simulation results are obtained by averaging more than 1000 instantaneous square errors in the steady-state and then averaging 200 independent trials. The simulation results present good agreement with the theoretical results for small step size but deviates from the theoretical one for a larger step size and larger  $K$ . The theoretical MSE in [9] is almost the same as the curve corresponding to  $K = 1$  in Fig. 7; the MSE expression in [9] is independent of  $K$  and is therefore not able to predict the variations in MSE as a function of  $K$ . Fig. 8 shows the steady state MSE for different delay parameters  $D$ . As  $D$  increases, the MSE decreases. To compare the EMSE expressions in Sections III and V, theoretical MSE curves using (20), (22), (23), and (59) are plotted in Fig. 9. The EMSE curves using (20) and (22) show good agreement with the simulation results.

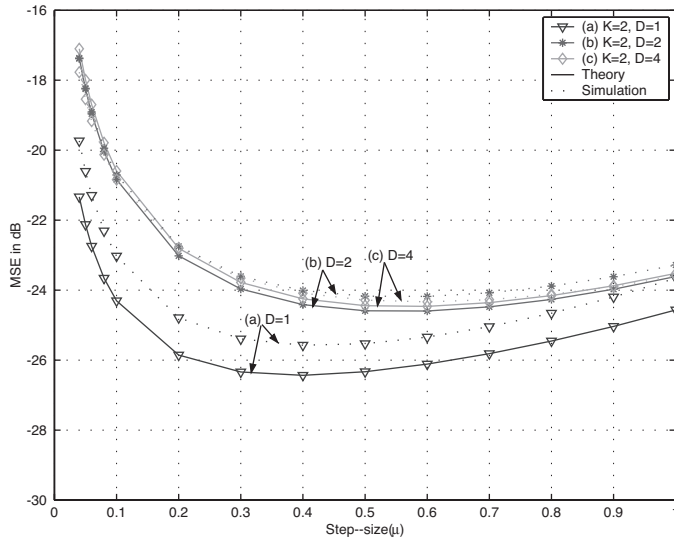


Fig. 14. Steady-state MSE curves of the APA family for colored Gaussian input  $K = 2$  in nonstationary environments. (a)  $D = 1$ . (b)  $D = 2$ . (c)  $D = 4$  [Input: Gaussian AR(1), pole at 0.9. System: FIR (16)].

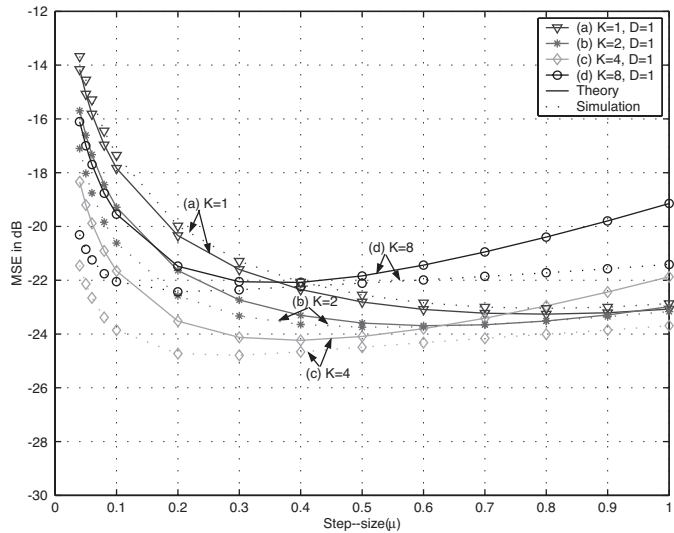


Fig. 15. Steady-state MSE curves of the APA family for colored uniform input using  $D = 1$  in nonstationary environments. (a)  $K = 1$ . (b)  $K = 2$ . (c)  $K = 4$ . (d)  $K = 8$  [Input: uniform AR(1), pole at 0.9. System: FIR (16)].

Fig. 10 shows comparison of MSE with [10]. Figs. 11 and 12 present the results for a colored uniform input signal.

### C. Tracking Performance

Figs. 13–16 show the steady-state MSE tracking performance of the APA family in a nonstationary environment. The steady-state tracking MSE in (31) is not a monotonically increasing function of  $\mu$ . Therefore, there exists an optimal value of step size  $\mu^\circ$  that minimizes the MSE in the nonstationary case. To see this, the range of the step-size is set from 0.04 to 1.0. We are using an i.i.d. sequence  $\mathbf{q}_i$  with autocorrelation matrix  $\mathbf{Q} = \sigma_q^2 \mathbf{I}$ , where  $\sigma_q^2 = 0.5\sigma_v^2$ . Fig. 13 shows the theoretical and simulated results for colored Gaussian input for the different value of  $K$ . For a given  $K$ , there exists an optimal  $\mu$  that minimizes the MSE, and for a given  $\mu$ , there exists an optimal  $K$ , which minimizes the MSE. Fig. 14 shows the tracking performance for

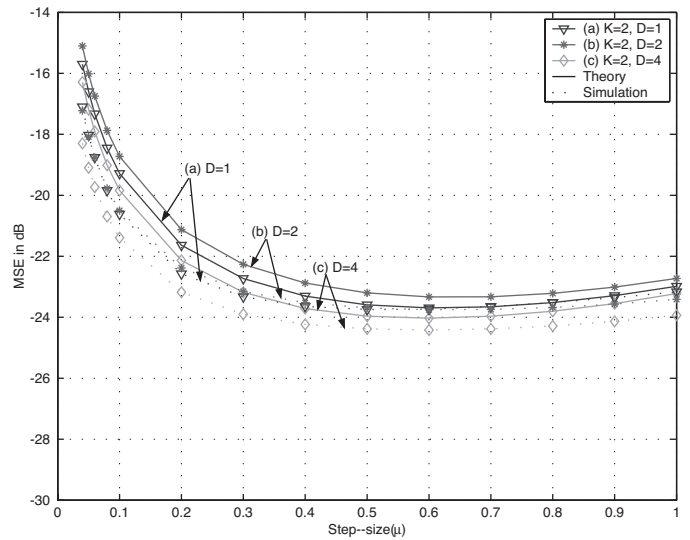


Fig. 16. Steady-state MSE curves of the APA family for colored uniform input using  $K = 2$  in nonstationary environments. (a)  $D = 1$ . (b)  $D = 2$ . (c)  $D = 4$  [Input: uniform AR(1), pole at 0.9. System: FIR (16)].

different values of  $D$ . The simulation results show the dependence of the tracking performance on  $D$ . For a given  $\mu$ , there exists an optimal  $D$  that minimizes the MSE. Figs. 15 and 16 show the theoretical and simulated results for colored uniform input signal.

## VII. CONCLUSIONS

In this paper, we carried out a rather detailed mean-square performance evaluation of the family of affine projection algorithms under the assumptions A1), A2), and A3'). Using energy-conservation arguments, we were able to derive expressions for the steady-state mean-square error and mean-square deviation without restricting the distribution of the input data to being Gaussian or white and without assuming any particular model for the input signals. Both stationary and nonstationary environments were considered. We also characterized the transient behavior of the filters by means of a first-order state-space model, whose stability was shown to determine the mean-square stability of the adaptive filter. Several simulation results were included to illustrate the application of the theory. In particular, it was seen that there is relatively good match between theory and practice.

### APPENDIX A EVALUATION OF $E[\mathbf{e}_{a,i}\mathbf{e}_{a,i}^*]$

Recall that the *a priori* and *a posteriori* error vectors are defined by

$$\mathbf{e}_{a,i} = \begin{bmatrix} \mathbf{u}_i \tilde{\mathbf{w}}_{i-1} \\ \mathbf{u}_{i-1} \tilde{\mathbf{w}}_{i-1} \\ \vdots \\ \mathbf{u}_{i-K+1} \tilde{\mathbf{w}}_{i-1} \end{bmatrix}, \quad \mathbf{e}_{p,i} = \begin{bmatrix} \mathbf{u}_i \tilde{\mathbf{w}}_i \\ \mathbf{u}_{i-1} \tilde{\mathbf{w}}_i \\ \vdots \\ \mathbf{u}_{i-K+1} \tilde{\mathbf{w}}_i \end{bmatrix}$$

where we are assuming  $\alpha = 0$  and  $D = 1$  without loss of generality. From (6), we know that

$$\mathbf{e}_{p,i} = \mathbf{e}_{a,i} - \mu \mathbf{e}_i = (1 - \mu)\mathbf{e}_{a,i} - \mu \mathbf{v}_i$$

when  $\epsilon$  is small. Then, the following relations hold:

$$\begin{aligned} E|\mathbf{u}_i \tilde{\mathbf{w}}_i|^2 &= (1 - \mu)^2 E|\mathbf{u}_i \tilde{\mathbf{w}}_{i-1}|^2 + \mu^2 \sigma^2 \\ E|\mathbf{u}_{i-1} \tilde{\mathbf{w}}_i|^2 &= (1 - \mu)^2 E|\mathbf{u}_{i-1} \tilde{\mathbf{w}}_{i-1}|^2 + \mu^2 \sigma^2 \\ &\vdots \\ E|\mathbf{u}_{i-K+1} \tilde{\mathbf{w}}_i|^2 &= (1 - \mu)^2 E|\mathbf{u}_{i-K+1} \tilde{\mathbf{w}}_{i-1}|^2 + \mu^2 \sigma^2. \end{aligned}$$

From these relations, we also get

$$\begin{aligned} E|\mathbf{u}_{i-1} \tilde{\mathbf{w}}_{i-1}|^2 &= (1 - \mu)^2 E|\mathbf{u}_{i-1} \tilde{\mathbf{w}}_{i-2}|^2 + \mu^2 \sigma^2 \\ E|\mathbf{u}_{i-2} \tilde{\mathbf{w}}_{i-1}|^2 &= (1 - \mu)^2 E|\mathbf{u}_{i-2} \tilde{\mathbf{w}}_{i-2}|^2 + \mu^2 \sigma^2 \\ &\vdots \\ E|\mathbf{u}_{i-K+1} \tilde{\mathbf{w}}_{i-1}|^2 &= (1 - \mu)^2 E|\mathbf{u}_{i-K+2} \tilde{\mathbf{w}}_{i-2}|^2 + \mu^2 \sigma^2 \end{aligned}$$

but since in steady-state

$$E|e_a(i)|^2 = E|e_a(i-1)|^2 = \dots = E|e_a(i-K+1)|^2$$

and neglecting off-diagonal terms in  $E[\mathbf{e}_{a,i} \mathbf{e}_{a,i}^*]$ , we find that

$$E[\mathbf{e}_{a,i} \mathbf{e}_{a,i}^*] = E|e_a(i)|^2 S_1 + \mu^2 \sigma_v^2 S_2 \quad (60)$$

where the diagonal matrices ( $S_1, S_2$ ) are given by

$$S_1 = \begin{bmatrix} 1 & & & \\ & (1 - \mu)^2 & & \\ & & \ddots & \\ & & & (1 - \mu)^{2(K-1)} \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 + \dots + (1 - \mu)^{2K} \end{bmatrix}.$$

Note that when  $\mu$  is small,  $S_1 \approx I$  and  $\mu^2 \sigma_v^2 S_2 \approx \mathbf{0}$ . In addition, when  $\mu$  is close to 1 and when SNR is high,  $S_1 \approx \mathbf{11}^T$  and  $\mu^2 \sigma_v^2 S_2 \approx \mathbf{0}$  so that (60) agrees with our assumption A.2. Expression (60) suggests that other choices for  $S_1$  and  $S_2$  are possible for assumption A.2). However, simulations show that the simpler conditions in A.2) lead to good results.

#### APPENDIX B PROOF OF THEOREM 2

From properties of Kronecker products, we know that the eigenvalues of

$$C = (E[P_i^T] \otimes I) + (I \otimes E[P_i])$$

are all the combinations  $\{\lambda_i + \lambda_j\}$  for all  $1 \leq i, j \leq M$ , where  $\lambda_i$  are the eigenvalues of  $E[P_i]$ . Since  $\lambda_i > 0$ ,  $C$  is positive definite. Moreover,  $D = E[P_i^T \otimes P_i]$  is non-negative definite.

Now, we want to determine conditions on  $\mu$  in order to guarantee  $|\lambda(F(\mu))| < 1$ , where

$$F(\mu) = I - \mu C + \mu^2 D.$$

Following the same argument used in [17, App. A], we can establish the condition

$$\mu < \min \left\{ \frac{1}{\lambda_{\max}(C^{-1}D)}, \frac{1}{\max(\lambda(H) \in \mathfrak{R}^+)} \right\}$$

#### ACKNOWLEDGMENT

The authors would like to thank Prof. W.-J. Song for his support of the first author's visit to the UCLA Adaptive Systems Laboratory.

#### REFERENCES

- [1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [2] S. Haykin, *Adaptive Filter Theory*, 3rd ed, NJ: Prentice-Hall, 1996.
- [3] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19–27, 1984.
- [4] S. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication*. Boston, MA: Kluwer, 2000.
- [5] S. G. Kratzer and D. R. Morgan, "The partial-rank algorithm for adaptive beamforming," in *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 564, 1985, pp. 9–14.
- [6] M. Rupp, "A family of adaptive filter algorithms with decorrelating properties," *IEEE Trans. Signal Processing*, vol. 46, pp. 771–775, Mar. 1998.
- [7] S. G. Sankaran and A. A. (Louis) Beex, "Normalized LMS algorithm with orthogonal correction factors," in *Proc. 31st Annu. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1997, pp. 1670–1673.
- [8] D. R. Morgan and S. G. Kratzer, "On a class of computationally efficient, rapidly converging, generalized NLMS algorithms," *IEEE Signal Processing Lett.*, vol. 3, pp. 245–247, Aug. 1996.
- [9] S. G. Sankaran and A. A. (Louis) Beex, "Convergence behavior of affine projection algorithms," *IEEE Trans. Signal Processing*, vol. 48, pp. 1086–1096, Apr. 2000.
- [10] J. Apolinário, Jr., M. L. R. Campos, and P. S. R. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," *IEEE Trans. Signal Processing*, vol. 48, pp. 3235–3242, Nov. 2000.
- [11] N. J. Bershad, D. Linebarger, and S. McLaughlin, "A stochastic analysis of the affine projection algorithm for gaussian autoregressive inputs," in *Proc. ICASSP*, Salt Lake City, UT, 2001, pp. 3837–3840.
- [12] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive algorithms via the small gain theorem," *Proc. SPIE*, vol. 2563, pp. 458–469, July 1995.
- [13] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," in *IEEE Trans. Signal Processing*, June 1996, vol. 44, pp. 1428–1439.
- [14] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
- [15] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyzes of adaptive filters," *IEEE Trans. Signal Processing*, vol. 49, pp. 314–324, Feb. 2001.
- [16] —, "Ability of adaptive filters to track carrier offsets and random channel nonstationarities," *IEEE Trans. Signal Processing*, vol. 50, pp. 1533–1544, July 2002.
- [17] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Processing*, vol. 51, pp. 639–652, Mar. 2003.
- [18] —, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Processing*, vol. 51, pp. 653–663, Mar. 2003.
- [19] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, vol. 42, pp. 2937–2944, Nov. 1994.
- [20] G. Alexander, *Kronecker Products and Matrix Calculus With Applications*. New York: Halsted, 1981.



**Hyun-Chool Shin** was born in Seoul, Korea, in 1974. He received the B.Sc. and M.Sc. degrees in electronic and electrical engineering from Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 1997 and 1999, respectively. Since 1997, he has been a Research Assistant with the Department of Electronic and Electrical Engineering, POSTECH, where he is currently pursuing the Ph.D. degree.

His research interests include adaptive filter theory and methods applied to channel equalization

and identification.



**Ali H. Sayed** (F'01) received the Ph.D. degree in electrical engineering in 1992 from Stanford University, Stanford, CA.

He is currently Professor and Vice Chair of electrical engineering at the University of California, Los Angeles. He is also the Principal Investigator of the UCLA Adaptive Systems Laboratory ([www.ee.ucla.edu/asl](http://www.ee.ucla.edu/asl)). He has over 190 journal and conference publications, is the author of the textbook *Fundamentals of Adaptive Filtering* (New York: Wiley, 2003), is coauthor of the research monograph

*Indefinite Quadratic Estimation and Control* (Philadelphia, PA: SIAM, 1999) and of the graduate-level textbook *Linear Estimation* (Englewood Cliffs, NJ: Prentice-Hall, 2000). He is also co-editor of the volume *Fast Reliable Algorithms for Matrices with Structure* (Philadelphia, PA: SIAM, 1999). He is a member of the editorial boards of the *SIAM Journal on Matrix Analysis and Its Applications* and the *International Journal of Adaptive Control and Signal Processing* and has served as coeditor of special issues of the journal *Linear Algebra and Its Applications*. He has contributed several articles to engineering and mathematical encyclopedias and handbooks and has served on the program committees of several international meetings. He has also consulted with industry in the areas of adaptive filtering, adaptive equalization, and echo cancellation. His research interests span several areas, including adaptive and statistical signal processing, filtering and estimation theories, signal processing for communications, interplays between signal processing and control methodologies, system theory, and fast algorithms for large-scale problems.

Dr. Sayed is recipient of the 1996 IEEE Donald G. Fink Award, a 2002 Best Paper Award from the IEEE Signal Processing Society in the area of Signal Processing Theory and Methods, and co-author of two Best Student Paper awards at international meetings. He is also a member of the technical committees on Signal Processing Theory and Methods (SPTM) and on Signal Processing for Communications (SPCOM), both of the IEEE Signal Processing Society. He is a member of the editorial board of the IEEE SIGNAL PROCESSING MAGAZINE. He has also served twice as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, of which he is now serving as Editor-in-Chief.