TRANSIENT BEHAVIOR OF AFFINE PROJECTION ALGORITHMS

Hyun-Chool Shin *

Department of Electrical Engineering,
Pohang University of Science and Technology
shinhc@postech.ac.kr

Ali H. Sayed †

Department of Electrical Engineering,
University of California at Los Angeles
sayed@ee.ucla.edu

ABSTRACT

Most analytical results on affine projection algorithms assume special regression models or Gaussian regression data. The available analyses also treat different affine projection filters separately. This paper provides a unified treatment of the transient performance of a family of affine projection algorithms. The treatment relies on energy conservation arguments and does not restrict the input data to being Gaussian or white. Simulation results illustrate the analysis and the derived performance expressions.

1. INTRODUCTION

The normalized least mean-squares (NLMS) algorithm is among the most widely used adaptive filters due to its computational simplicity and ease of implementation. However, colored input signals can deteriorate its convergence speed appreciably. To address this problem, Ozeki and Umeda [1] developed the basic form of an affine projection algorithm (APA) using affine subspace projections. While NLMS updates the weights based only on the current input vector, APA updates the weights based on K previous input vectors. Since [1], many variants of APA have been devised independently from different perspectives such as the regularized APA (R-APA), the partial rank algorithm (PRA) [2], and NLMS with orthogonal correction factors (NLMS-OCF) [3]. We shall refer to all these algorithms as belonging to the APA family.

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 [4] 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 [3] focus on NLMS-OCF and rely on a special model for the input signal vector. A convergence analysis given in [5] allows the evaluation of learning curves assuming a Gaussian autoregressive input model.

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 learning curves, as well as conditions on the step-size for mean-square stability. Our derivation relies on energy conservation arguments [6]–[10] and it does not restrict the regression data to being Gaussian or white. Simulations at the end of the paper illustrate the derived results.

2. DATA MODELS AND APA FAMILY

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

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

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, $R_u = E\left[\mathbf{u}_i^* \mathbf{u}_i\right]$. In this paper, we focus on a general class of affine projection algorithms for estimating $\mathbf{w}^\circ$ of the form:

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

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, μ is the step-size and

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

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

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

For NLMS-OCF, it is further assumed that $\mathbf{u}_{i-jD}$ is orthogonal to $\mathbf{u}_i, \mathbf{u}_{i-D}, \cdots, \mathbf{u}_{i-(j-1)D}$. For PRA, it is understood that $\mathbf{w}_m =$

Table 1. APA family where $\{\alpha, K, D\}$ are integers.

Algorithm	K	ϵ	α	D
APA	$K \le M$	$\epsilon = 0$	$\alpha = 0$	$D = 1$
BNDR-LMS	$K = 2$	$\epsilon = 0$	$\alpha = 0$	$D = 1$
R-APA	$K \le M$	$\epsilon \ne 0$	$\alpha = 0$	$D = 1$
PRA	$K \le M$	$\epsilon \ne 0$	$\alpha = 1$	$D = 1$
NLMS-OCF	$K \le M$	$\epsilon = 0$	$\alpha = 0$	$D \ge 1$

*This work was performed while H. Shin was a visiting graduate student at the UCLA Adaptive Systems Laboratory and supported 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.

†This work was supported in part by the National Science Foundation under grants ECS-9820765 and CCR-0208573.

$\mathbf{w}_{i-K}$, for $m = i-1, \cdots, i-K$, i.e., the weight vector is updated once every K iterations. In our discussions, K can be greater than M and the only restriction on K is $K > 0$, although most algorithms assume $K \leq M$.

3. TRANSIENT ANALYSIS OF APA

We now study the transient (i.e., learning curves, steady-state behavior and stability) performance of the APA family. To do this, we rely on energy conservation arguments.

3.1. Weighted Energy Relation

We shall assume, without loss of generality, that $\alpha = 0$. Then (2) becomes

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

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

$$\mathbf{e}_{p,i}^{\Sigma} = \mathbf{e}_{a,i}^{\Sigma} - \mu U_i \Sigma U_i^*(\epsilon I + U_i U_i^*)^{-1}\mathbf{e}_i \tag{4}$$

where $\mathbf{e}_{p,i}^{\Sigma} = U_i\Sigma\tilde{\mathbf{w}}_i$ and $\mathbf{e}_{a,i}^{\Sigma} = U_i\Sigma\tilde{\mathbf{w}}_{i-1}$. Solving for $\mathbf{e}_i$ and substituting into (3), we get

$$\boxed{\tilde{\mathbf{w}}_i + U_i^*(U_i\Sigma U_i^*)^{-1}\mathbf{e}_{a,i}^{\Sigma} = \tilde{\mathbf{w}}_{i-1} + U_i^*(U_i\Sigma U_i^*)^{-1}\mathbf{e}_{p,i}^{\Sigma}} \tag{5}$$

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 (5) we find that

$$\begin{aligned} &\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 + \mathbf{e}_{a,i}^{*\Sigma}\left(U_i\Sigma U_i^*\right)^{-1}\mathbf{e}_{a,i}^{\Sigma} = \\ &\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma}^2 + \mathbf{e}_{p,i}^{*\Sigma}\left(U_i\Sigma U_i^*\right)^{-1}\mathbf{e}_{p,i}^{\Sigma} \end{aligned} \tag{6}$$

where $\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 = \tilde{\mathbf{w}}_i^*\Sigma\tilde{\mathbf{w}}_i$. The important fact to emphasize is that *no approximations* have been used to establish the energy relation (6); it is an exact relation that shows 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. Relation (6) is an extension to the APA case of the energy conservation relation originally derived in [6] in the context of robustness analysis and subsequently used in [7]–[10] in the context of steady-state and transient performance analysis.

3.2. 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 Σ (e.g., $\Sigma = I$ or $\Sigma = R_u$). 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, expressing $\{\mathbf{e}_{a,i}, \mathbf{e}_{a,i}^{\Sigma}, \mathbf{e}_{p,i}^{\Sigma}\}$ in terms of $\tilde{\mathbf{w}}_{i-1}$ and taking expectations of both sides, relation (6) becomes

$$E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2\right] + \mu^2 E\left[\mathbf{v}_i^* A_i^{\Sigma}\mathbf{v}_i\right] \tag{7}$$

where

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

and

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

The expectation $E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2\right]$ in (7) is difficult to evaluate due to the dependence of Σ' on U_i and of $\tilde{\mathbf{w}}_{i-1}$ on prior regressors. One common way to overcome this difficulty is to introduce *an independence assumption* on the regressor sequence U_i, namely, to assume that

A.2 *The matrix sequence $\{U_i\}$ is independent and identically distributed.*

This assumption guarantees that $\tilde{\mathbf{w}}_{i-1}$ is independent of both Σ' and U_i. Clearly, A.2 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.). Observe from the expansion for Σ' that it is sufficient for our purposes to require

A.2' $\tilde{\mathbf{w}}_{i-1}$ *is independent of* $U_i^*(\epsilon I + U_i U_i^*)^{-1}U_i$

which is a weaker assumption and more likely to hold. In this way, recursion (7) reduces to

$$E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2\right] + \mu^2 E\left[\mathbf{v}_i^* A_i^{\Sigma}\mathbf{v}_i\right] \tag{8}$$

where now

$$\begin{aligned} \Sigma' = \;&\Sigma - \mu\Sigma E\left[U_i^*(\epsilon I + U_i U_i^*)^{-1}U_i\right] \\ &-\mu E\left[U_i^*(\epsilon I + U_i U_i^*)^{-1}U_i\right]\Sigma + \mu^2 E\left[U_i^* A_i^{\Sigma} U_i\right] \end{aligned}$$

with expectations appearing in Σ'. Also taking expectations of both sides of (3) and using assumption A.1, we obtain the following result for the evolution of the mean of the weight-error vector:

$$E\left[\tilde{\mathbf{w}}_i\right] = E\left[I - \mu U_i^*(\epsilon I + U_i U_i^*)^{-1}U_i\right]E\left[\tilde{\mathbf{w}}_{i-1}\right] \tag{9}$$

Relations (8) and (9) 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.

Using the following property of the Kronecker product of matrices,

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

and introducing the vector notations $\sigma' = \text{vec}\{\Sigma'\}$ and $\sigma = \text{vec}\{\Sigma\}$, we find that[1]

$$\sigma' = F\sigma \tag{10}$$

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

$$F = I - \mu(E[P_i^T] \otimes I + I \otimes E[P_i]) + \mu^2 E[P_i^T \otimes P_i] \tag{11}$$

with

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

We can rewrite the recursion for $E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right]$ in (7) by using the vectors $\{\sigma', \sigma\}$ instead of the matrices $\{\Sigma', \Sigma\}$ as follows

$$E\left[\|\tilde{\mathbf{w}}_i\|_{\text{vec}\{\sigma\}}^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\text{vec}\{\sigma'\}}^2\right] + \mu^2\sigma_v^2(\gamma^T\sigma) \tag{12}$$

[1]The vec($\cdot$) operation stacks the columns of a matrix into a vector.

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

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

where $\gamma = \text{vec}\{E\left[U_i^*(\epsilon I + U_i U_i^*)^{-2} U_i\right]\}$. For compactness of notation, we drop the vec$\{\cdot\}$ notation from the subscripts and keep the vectors, so that the above is simply rewritten as

$$\boxed{E\left[\|\tilde{\mathbf{w}}_i\|_\sigma^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{F\sigma}^2\right] + \mu^2 \sigma_v^2 (\gamma^T \sigma)} \tag{13}$$

Also we obtain the following result for the evolution of the mean of the weight-error vector:

$$\boxed{E\left[\tilde{\mathbf{w}}_i\right] = (I - \mu E[P_i]) E\left[\tilde{\mathbf{w}}_{i-1}\right]} \tag{14}$$

Recursion (13) shows that in order to evaluate $E\left[\|\tilde{\mathbf{w}}_i\|_\sigma^2\right]$ we need to know $E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{F\sigma}^2\right]$, with a weighting matrix whose entries are determined by $F\sigma$. Now the quantity $E\left[\|\tilde{\mathbf{w}}_i\|_{F\sigma}^2\right]$ can be inferred from (13) by writing the recursion for $F\sigma$, i.e.,

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

We again find that in order to evaluate $E\left[\|\tilde{\mathbf{w}}_i\|_{F\sigma}^2\right]$ we need to know $E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{F^2\sigma}^2\right]$. Fortunately, as in [9], this procedure terminates. This is because any matrix F satisfies $p(F) = 0$ where

$$p(x) = \det(xI - F)$$

denotes its characteristic polynomial, say

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

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

$$\boxed{\mathcal{W}_i = \mathcal{F}\mathcal{W}_{i-1} + \mu^2 \sigma_v^2 \mathcal{Y}}$$

where

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

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

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

3.3. 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

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

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 relation (13) by iterating it and setting $\mathbf{r} = \text{vec}(R_u)$. This yields

$$E\left[\|\tilde{\mathbf{w}}_i\|_{\mathbf{r}}^2\right] = E\left[\|\tilde{\mathbf{w}}_{-1}\|_{F^i\mathbf{r}}^2\right] + \mu^2 \sigma_v^2 \left(\gamma^T (I + \cdots + F^{i-1})\mathbf{r}\right) \tag{15}$$

That is,

$$E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\mathbf{r}}^2\right] = E\left[\|\tilde{\mathbf{w}}_{-1}\|_{\mathbf{f}_{i-1}}^2\right] + \mu^2 \sigma_v^2 g(i-1) \tag{16}$$

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

$$\mathbf{f}_{i-1} = F\mathbf{f}_{i-2}$$

$$g(i-1) = g(i-2) + \gamma^T \mathbf{f}_{i-1}$$

with initial condition $\mathbf{f}_0 = \mathbf{r}$ and $g(-1) = 0$.

3.4. Mean-Square Stability

From (14) the convergence in the mean of the APA family is guaranteed for any μ satisfying

$$\mu < \frac{2}{\lambda\text{max}(E[P_i])} \tag{17}$$

Moreover, recursion (13) 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 (see [9]).

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

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

where $H = \begin{bmatrix} \frac{1}{2}C & -\frac{1}{2}D \\ I & 0 \end{bmatrix}$. The above condition on μ is in terms of the largest positive eigenvalue of H when it exists. By combining (17) and Theorem 2, a bound on the step-size for both mean and mean-square stability is obtained.

3.5. Steady-State Behavior

Assuming the step-size μ is chosen to guarantee filter stability, recursion (13) becomes in steady-state

$$E\left[\|\tilde{\mathbf{w}}_\infty\|_\sigma^2\right] = E\left[\|\tilde{\mathbf{w}}_\infty\|_{F\sigma}^2\right] + \mu^2 \sigma_v^2 (\gamma^T \sigma) \tag{18}$$

which is equivalent to

$$E\left[\|\tilde{\mathbf{w}}_\infty\|_{(I-F)\sigma}^2\right] = \mu^2 \sigma_v^2 (\gamma^T \sigma) \tag{19}$$

Assume that we select σ as the solution to the linear system of equations $(I - F)\sigma = \text{vec}\{I\}$. In this case, the weighting quantity that appears in (19) reduces to the vector of unit entries. Then the left-hand side of (19) becomes the filter MSD and (19) leads to

$$\boxed{\text{MSD} = \mu^2 \sigma_v^2 \gamma^T (I - F)^{-1} \text{vec}\{I\}} \tag{20}$$

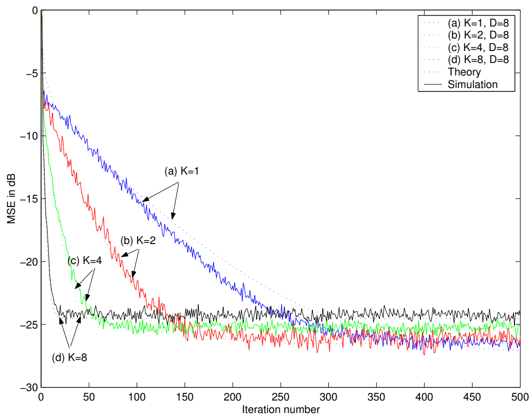

Fig. 1. Learning curves of the APA family for colored Gaussian input using $\mu = 1.0$ and $D = 8$ (a) $K = 1$ (b) $K = 2$ (c) $K = 4$ (d) $K = 8$

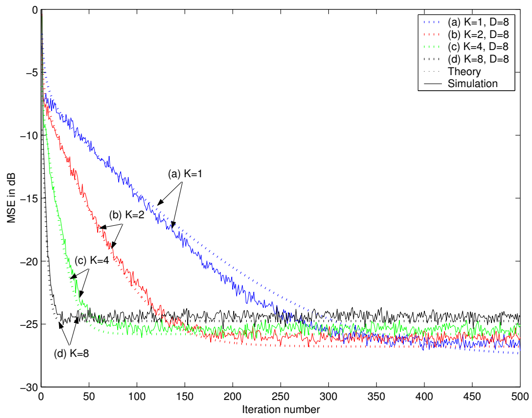

Fig. 2. Learning curves of the APA family for colored uniform input using $\mu = 1.0$ and $D = 8$ (a) $K = 1$ (b) $K = 2$ (c) $K = 4$ (d) $K = 8$

In a similar way, since

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

we can determine the EMSE by evaluating $E\left[\|\tilde{\mathbf{w}}_\infty\|^2_{\mathbf{r}}\right]$, where the weighting factor is $\mathbf{r} = \text{vec}\{R_u\}$. Assume we select σ as the solution to the linear system of equations $(I - F)\sigma = \mathbf{r}$. In this case, the weighting quantity that appears in (19) reduces to R_u. Then the LHS of (19) becomes the filter EMSE and (19) leads to the desired result

$$\boxed{\text{EMSE} = \mu^2\sigma_v^2\gamma^T(I - F)^{-1}\text{vec}(R_u)} \tag{21}$$

4. 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. Two different types of signals, viz., Gaussian and 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 τ. For the Gaussian case, $\rho(i)$ is a white, zero-mean, Gaussian random sequence having unit variance and τ 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 τ is set to 0.9. The signal-to-noise ratio (SNR) is calculated by $\text{SNR} = 10\log(E[y^2(i)]/E[v^2(i)])$ where $y(i) = \mathbf{u}_i\mathbf{w}^\circ$. The measurement noise $v(i)$ is added to $y(i)$ such that SNR $=$ 30dB. The adaptive filter and the unknown channel are assumed to have the same number of taps. All adaptive filter coefficients are initialized to zero. Also, the regularization parameter ϵ is set to 0.001. We set $\alpha = 0, \mu = 1.0$ and $D = 8$. The simulation results shown are obtained by ensemble averaging over 200 independent trials. Fig.1–2 shows the learning curves for both the theoretical and the simulation results.

5. CONCLUSIONS

Using energy conservation arguments, the paper analyzed the steady-state and transient performances of the APA family and derived stability conditions without restricting the regression data to being Gaussian or white.

Acknowledgment

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

6. REFERENCES

[1] 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, pp. 19–27, 1984.

[2] S. G. Kratzer and D. R. Morgan, "The partial-rank algorithm for adaptive beamforming," *Proc. SPIE Int. Soc. Optic. Eng.*, vol. 564, pp. 9–14, 1985.

[3] S. G. Sankaran and A. A. (Louis) Beex, "Convergence behavior of affine projection algorithms," *IEEE Trans. Signal Processing*, vol. 48, no. 4, pp. 1086–1096, April 2000.

[4] J. Apolinário, Jr., M. L. R. Campos, and P. S. R. Diniz, "Convergence analysis of the binormailzed data-reusing LMS algorithm," *IEEE Trans. Signal Processing*, vol. 48, no. 11, pp. 3235–3242, Nov. 2000.

[5] N. J. Bershad and D. Linebarger and S. McLaughlin, "A stochastic analysis of the affine projection algorithm for Gaussian autoregressive inputs," *Proc. ICASSP*, , pp. 3837–3840, Saly Lake City, 2001.

[6] 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, San Diego, CA, Jul. 1995.

[7] N. R. Yousef and A. H. Sayed, "A unified aproach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. Signal Processing*, vol. 49, pp. 314–324, Feb. 2001.

[8] N. R. Yousef and A. H. Sayed, "Ability of adaptive filters to track carrier offsets and random channel nonstationarities," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1533–1544, July 2002.

[9] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Processing*, vol. 51, no. 3, Mar. 2003.

[10] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Processing*, vo. 51, no. 3, Mar. 2003.