Mean-Square Performance of Data-Reusing Adaptive Algorithms

Hyun-Chool Shin, *Member, IEEE*, Woo-Jin Song, *Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*

***Abstract*—This letter provides a unified mean-square performance analysis of the class of data reusing adaptive algorithms. The derivation relies on energy conservation arguments, and it does not restrict the regression data to being Gaussian. Simulation results show that there is a relatively good match between theory and practice.**

***Index Terms*—Affine projection algorithm (APA), data-reusing adaptive filters, energy conservation relation.**

I. Introduction

AFFINE projection (APA) and data-reusing LMS (DR-LMS) or normalized DR-LMS (NDR-LMS) algorithms have desirable convergence properties and computational costs. Several variants of APA and DR-LMS or NDR-LMS have been devised independently from different perspectives [1]–[4].

For APA algorithms, the mean-square error, tracking, and transient performances have been studied in [5], [6], and the references therein. However, the transient behaviors of DR-LMS and NDR-LMS are not as widely studied. The available results have progressed more qualitatively than quantitatively [9], [10]. In addition, although APA and DR-LMS or NDR-LMS algorithms have common features in that block errors and block regression data are used for updating the filter coefficients, each algorithm is usually studied separately in the literature under different assumptions. Such distinct treatments tend to obscure commonalities that exist between APA and DR-LMS algorithms.

In this letter, we provide a unified treatment of the convergence performance of DR-LMS and NDR-LMS algorithms, which can be used to clarify the common features among APA, DR-LMS, and NDR-LMS. To do so, we first introduce a uniform cost function from which APA, DR-LMS, and NDR-LMS can be motivated as instantaneous gradient approximations depending on the choice of a certain weighting matrix. Subsequently, the performance analysis is pursued using the energy conservation approach of [8, ch. 6 and 9] and in a manner similar to what was done earlier for APA in [6]. A feature of the analysis is that it does not restrict the regression data to being Gaussian.

Manuscript received April 27, 2005; revised July 6, 2005. This work was supported in part by the Brain Korea (BK) 21 Program funded by the Ministry of Education and in part by the HY-SDR Research Center at Hanyang University under the ITRC Program of MIC, Republic of Korea. The work of A. H. Sayed was supported in part by the National Science Foundation under Grant ECS-0401188 and Grant CCR-0208573. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dominic K. C. Ho.

H.-C. Shin is with the Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD 21205 USA (e-mail: shinhc@postech.ac.kr).

W.-J. Song is with the Division of Electronics and Computer Engineering, Pohang University of Science and Technology (POSTECH), 790-784 Kyungbuk, Korea (e-mail: wjsong@postech.ac.kr).

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

Digital Object Identifier 10.1109/LSP.2005.859500

Throughout the letter, the following notations are adopted.

$\|\mathbf{a}\|_\Sigma$	Weighted Euclidean norm of a vector, i.e., $\mathbf{a}^*\Sigma\mathbf{a}$.
$\mathrm{Tr}(\cdot)$	Trace of a matrix.
$\mathrm{diag}\{\cdots\}$	Diagonal matrix of its entries $\{\cdots\}$.
$A\otimes B$	Kronecker product of matrices A and B.
$\mathrm{vec}\{A\}$	Stacking the columns of a matrix A.
$\mathrm{vec}\{\mathbf{a}\}$	Writing $\mathrm{vec}\{\mathbf{a}\}$ for an $M^2\times 1$ column vector $\mathbf{a}$ results in an $M\times M$ matrix A whose entries are obtained from $\mathbf{a}$.
$\det(\cdot)$	Determinant of a matrix.
$\lambda_{\max}(\cdot)$	Largest eigenvalue of a matrix.
$\Re^+$	Set of positive real numbers.

II. Data-Reusing Algorithms

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) vector $\mathbf{u}_i=[u(i)\ \ u(i-1)\ \ \cdots\ \ u(i-M+1)]$ with a positive-definite covariance matrix $R_u=E[\mathbf{u}_i^*\mathbf{u}_i]$.

A. DR-LMS and NDR-LMS Adaptive Algorithms

The DR-LMS and NDR-LMS algorithms for estimating $\mathbf{w}^\circ$ take the forms

$$\mathbf{w}_i=\mathbf{w}_{i-1}+\mu\mathbf{U}_i^*\mathbf{e}_i \tag{2}$$
$$\mathbf{w}_i=\mathbf{w}_{i-1}+\mu\mathbf{U}_i^*\mathbf{D}_i\mathbf{e}_i \tag{3}$$

respectively, where $\mathbf{w}_i$ is an estimate for $\mathbf{w}^\circ$ at iteration i,μ is the step size, $\mathbf{e}_i=\mathbf{d}_i-\mathbf{U}_i\mathbf{w}_{i-1},\mathbf{D}_i=\mathrm{diag}\{1/\|\mathbf{u}_i\|^2,\ldots,1/\|\mathbf{u}_{i-K+1}\|^2\}$, and usually, $K\le M$

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

B. *Unified Cost Function*

These algorithms can be motivated as instantaneous gradient approximations as follows. Consider a cost function of the form

$$J(i) = E[\mathbf{e}_i^* \Pi \mathbf{e}_i]/2 \tag{4}$$

for some general positive-definite weighting matrix Π. The gradient vector of $J(i)$ with respect to $\mathbf{w}_{i-1}$ is

$$\frac{\partial J(i)}{\partial \mathbf{w}_{i-1}} = -E\left[\mathbf{e}_i^* \Pi \mathbf{U}_i\right]. \tag{5}$$

Replacing the expected value in (5) by its instantaneous value (i.e., removing the expectation sign), we can motivate the stochastic gradient algorithm

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{U}_i^* \Pi \mathbf{e}_i \tag{6}$$

where μ is the step size that controls the convergence of the adaptation process. Different choices of the parameter Π result in different algorithms. For example, if we choose Π as the identity matrix I, then the DR-LMS algorithm (2) is obtained. Also, by choosing $\Pi = \mathbf{D}_i$, we get the NDR-LMS algorithm (3). Likewise, the choice $\Pi = (\mathbf{U}_i \mathbf{U}_i^*)^{-1}$ results in APA

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

where it is assumed that $\mathbf{U}_i \mathbf{U}_i^*$ is invertible. The equation in (6) covers data reusing over K pairs, not covering the case where the same data matrix U_i is repeatedly used [7].

III. Mean-Square Performance Analysis

We now examine the mean-square performance of general adaptive algorithms of the form (6). To do so, we rely on the energy conservation approach of [8, ch. 6 and 9], which was used in [6] to study the performance of APA algorithms. Thus, note that (6) can be rewritten in terms of the weight-error vector $\tilde{\mathbf{w}}_i = \mathbf{w}^\circ - \mathbf{w}_i$ as

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu \mathbf{U}_i^* \Pi \mathbf{e}_i. \tag{7}$$

Now introduce the *a priori* and *a posteriori* weighted estimation errors

$$\mathbf{e}_{p,i}^{\Sigma} = \mathbf{U}_i \Sigma \tilde{\mathbf{w}}_i \quad \mathbf{e}_{a,i}^{\Sigma} = \mathbf{U}_i \Sigma \tilde{\mathbf{w}}_{i-1} \tag{8}$$

for any Hermitian positive-definite matrix Σ to be chosen later; different choices for Σ allow us to answer different questions pertaining to the mean-square performance of the filter [8], [11]. If we multiply both sides of (7) by $\mathbf{U}_i \Sigma$ from the left, we find that

$$\mathbf{e}_{p,i}^{\Sigma} = \mathbf{e}_{a,i}^{\Sigma} - \mu \mathbf{U}_i \Sigma \mathbf{U}_i^* \Pi \mathbf{e}_i. \tag{9}$$

Solving for $\mathbf{e}_i$ and substituting into (7), we get

$$\tilde{\mathbf{w}}_i + \mathbf{U}_i^* (\mathbf{U}_i \Sigma \mathbf{U}_i^*)^{-1} \mathbf{e}_{a,i}^{\Sigma} = \tilde{\mathbf{w}}_{i-1} + \mathbf{U}_i^* (\mathbf{U}_i \Sigma \mathbf{U}_i^*)^{-1} \mathbf{e}_{p,i}^{\Sigma}. \tag{10}$$

If we equate the weighted Euclidean norms of both sides of (10), we find that

$$\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 + \mathbf{e}_{a,i}^{*\Sigma} \left(\mathbf{U}_i \Sigma \mathbf{U}_i^*\right)^{-1} \mathbf{e}_{a,i}^{\Sigma} = \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma}^2 + \mathbf{e}_{p,i}^{*\Sigma} \left(\mathbf{U}_i \Sigma \mathbf{U}_i^*\right)^{-1} \mathbf{e}_{p,i}^{\Sigma} \tag{11}$$

where $\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 = \tilde{\mathbf{w}}_i^* \Sigma \tilde{\mathbf{w}}_i$, and using (8), the crossed terms are eliminated since they are identical on both sides. Note that no approximations have been used to establish the energy relation (12).

Although we have started with the generalized adaptive form (6), the resulting weighted energy relation (12) is fortunately of the same form as [6, eq. (36)]. This is because in the relation (9) between $\mathbf{e}_{p,i}^{\Sigma}$ and $\mathbf{e}_{a,i}^{\Sigma}$, the matrix Π is included. This is the authors' intention to make the remaining analysis similar to [6]. Thus, here we will briefly describe the remaining manipulation, focusing on the main results. Note that the symbols for notations look similar, but the actual definitions are different from [6].

A. *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 $\Sigma = R_u$). The following is the often realistic assumption.

A.1: The noise $v(i)$ is independent and identically distributed (i.i.d.) and statistically independent of the regression matrix $\{\mathbf{U}_i\}$.

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 (12) gives

$$E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2\right] + \mu^2 E\left[\mathbf{v}_i^* A_i^{\Sigma} \mathbf{v}_i\right] \tag{12}$$

where $\mathbf{v}_i = [v(i)\ v(i-1)\ \cdots v(i-K+1)]^T, \Sigma' \triangleq \Sigma - \mu \Sigma \mathbf{U}_i^* \Pi \mathbf{U}_i - \mu \mathbf{U}_i^* \Pi \mathbf{U}_i \Sigma + \mu^2 (\mathbf{U}_i^* A_i^{\Sigma} \mathbf{U}_i)$, and $A_i^{\Sigma} \triangleq \Pi \mathbf{U}_i \Sigma \mathbf{U}_i^* \Pi$. The expectation $E[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2]$ in (12) is difficult to evaluate due to the dependence of Σ' on $\mathbf{U}_i$ and of $\tilde{\mathbf{w}}_{i-1}$ on prior regressors. One common way to overcome this difficulty, especially for small step sizes, is to assume the following.

A.2: $\tilde{\mathbf{w}}_{i-1}$ is independent of $\mathbf{U}_i^* \Pi \mathbf{U}_i$.

In this way, Σ' in recursion (12) reduces to $\Sigma' = \Sigma - \mu \Sigma E[\mathbf{U}_i^* \Pi \mathbf{U}_i] - \mu E[\mathbf{U}_i^* \Pi \mathbf{U}_i] \Sigma + \mu^2 E[\mathbf{U}_i^* A_i^{\Sigma} \mathbf{U}_i]$ with expectations appearing in Σ'.

Using the following property of the Kronecker product of matrices, $\mathrm{vec}\{P \Sigma Q\} = (Q^T \otimes P) \mathrm{vec}(\Sigma)$ and introducing the vector notations $\sigma' = \mathrm{vec}\{\Sigma'\}$ and $\sigma = \mathrm{vec}\{\Sigma\}$, we find that

$$\sigma' = F \sigma \tag{13}$$

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

$$F = I - \mu \left(E\left[P_i^T\right] \otimes I + I \otimes E[P_i]\right) + \mu^2 E\left[P_i^T \otimes P_i\right] \tag{14}$$

with $P_i = \mathbf{U}_i^* \Pi \mathbf{U}_i$. We can rewrite the recursion for $E[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2]$ in (12) by using the vectors $\{\sigma', \sigma\}$ instead of the matrices $\{\Sigma', \Sigma\}$ as follows:

$$E\left[\|\tilde{\mathbf{w}}_i\|_{\sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{\sigma'}^2\right] + \mu^2 \sigma_v^2 (\gamma^T \sigma) \tag{15}$$

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

$$\mathrm{Tr}\left(E[\Pi \mathbf{U}_i \Sigma \mathbf{U}_i^* \Pi]\right) = \gamma^T \sigma$$

where $\gamma = \text{vec}\{E[\mathbf{U}_i^* \Pi^2 \mathbf{U}_i]\}$. 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

$$E\left[\|\tilde{\mathbf{w}}_i\|_\sigma^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|_{F\sigma}^2\right] + \mu^2\sigma_v^2(\gamma^T\sigma). \quad (16)$$

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

$$E[\tilde{\mathbf{w}}_i] = (I - \mu E[P_i])E[\tilde{\mathbf{w}}_{i-1}]. \quad (17)$$

B. Mean-Square Stability

From (17), the convergence in the mean of the adaptive filter is guaranteed for any μ satisfying

$$\mu < 2/\lambda_{\max}(E[P_i]). \quad (18)$$

Moreover, as in [6], recursion (16) 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 [8, ch. 9]).

Theorem 1 (Stability): Under conditions A.1 and A.2, the convergence in the mean-square sense of the class of adaptive filters (6) is guaranteed for any μ in the range

$$0 < \mu < \min\{1/\lambda_{\max}(C^{-1}D), 1/\max(\lambda(H) \in \Re^+)\}$$

where

$$H = \begin{bmatrix} (1/2)C & -(1/2)D \\ I & 0 \end{bmatrix}.$$

C. Steady-State Behavior

As in [6] and [11], assuming the step size μ is chosen to guarantee filter stability, recursion (16) becomes in steady state

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

which is equivalent to

$$E\left[\|\tilde{\mathbf{w}}_\infty\|_{(I-F)\sigma}^2\right] = \mu^2\sigma_v^2(\gamma^T\sigma). \quad (20)$$

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 (20) reduces to the vector of unit entries. Then the left-hand side of (20) becomes the filter mean-square deviation (MSD), and (20) leads to

$$\text{MSD} = \mu^2\sigma_v^2\gamma^T(I - F)^{-1}\text{vec}\{I\}. \quad (21)$$

In a similar way, since $E\,|\,e_a(i)|^2 = E\|\tilde{\mathbf{w}}_{i-1}\|_{R_u}^2$, we can determine the excess mean-square error (EMSE) by evaluating $E\|\tilde{\mathbf{w}}_\infty\|_{\mathbf{r}}^2$, where the weighting factor is $\mathbf{r} = \text{vec}\{R_u\}$, i.e.,

$$\text{EMSE} = \mu^2\sigma_v^2\gamma^T(I - F)^{-1}\text{vec}(R_u). \quad (22)$$

IV. Simulation Results

We illustrate the theoretical results presented in this letter by carrying out computer simulations in a channel estimation scenario. The unknown channel has 16 taps and is randomly generated. The adaptive filter and the unknown channel are assumed to have the same number of taps. The input signal $u(i)$ is obtained by filtering a white, zero-mean, Gaussian random sequence through the system

$$G(z) = \frac{1 + 0.5z^{-1} + 0.81z^{-2}}{1 - 0.59z^{-1} + 0.4z^{-2}}.$$

TABLE I
MEAN-SQUARE STABILITY BOUNDS OF DR-LMS ($\Pi = I$)

	$\frac{1}{\lambda_{\max}(C^{-1}D)}$	$\frac{1}{\max(\lambda(H)\in\Re^+)}$	$\mu_{\max}$
$K = 1$	0.098	0.3788	0.098
$K = 2$	0.063	0.2028	0.063
$K = 4$	0.041	0.1087	0.041
$K = 8$	0.031	0.079	0.031

TABLE II
MEAN-SQUARE STABILITY BOUNDS OF NDR-LMS
($\Pi = \text{diag}\{1/\|\mathbf{u}_i\|^2, \ldots, 1/\|\mathbf{u}_{i-K+1}\|^2\}$)

	$\frac{1}{\lambda_{\max}(C^{-1}D)}$	$\frac{1}{\max(\lambda(H)\in\Re^+)}$	$\mu_{\max}$
$K = 1$	2.0002	6.3131	2.0002
$K = 2$	1.2903	3.3727	1.2903
$K = 4$	0.9123	1.8396	0.9123
$K = 8$	0.6638	1.1625	0.6638

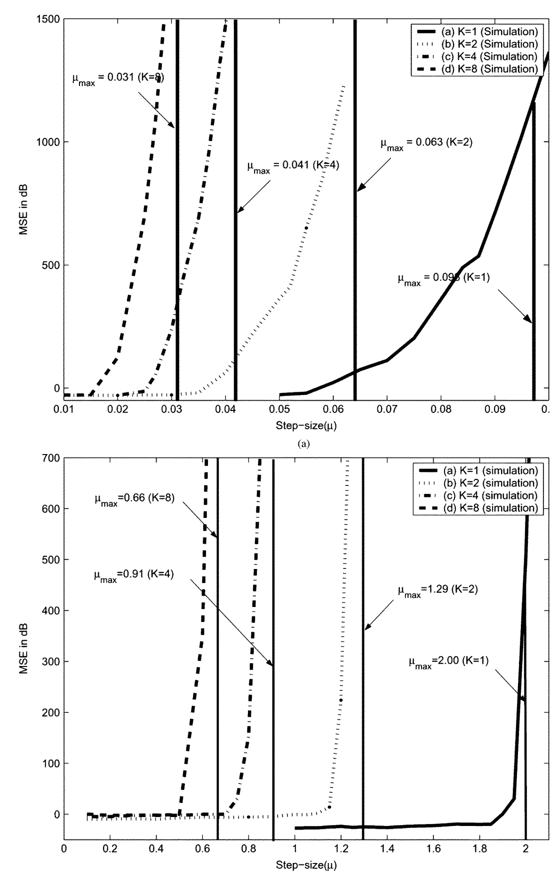

Fig. 1. Simulated MSE of DR-LMS and NDR-LMS as a function of the step size. (a) DR-LMS. (b) NDR-LMS.

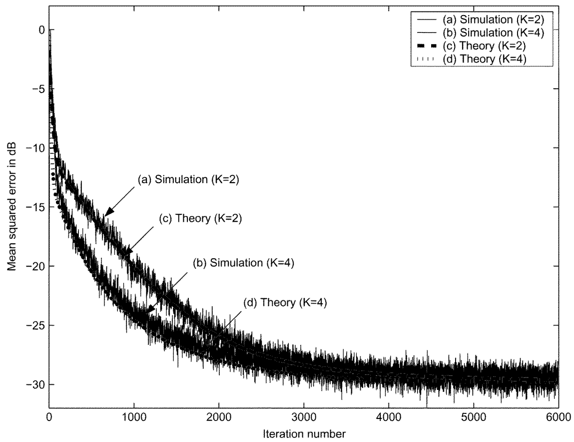

Fig. 2. Learning curves of DR-LMS for colored Gaussian input using $\mu = 0.005$ for $K = 2$ and $K = 4$ [Input: Gaussian ARMA(2, 2); System: FIR (16)].

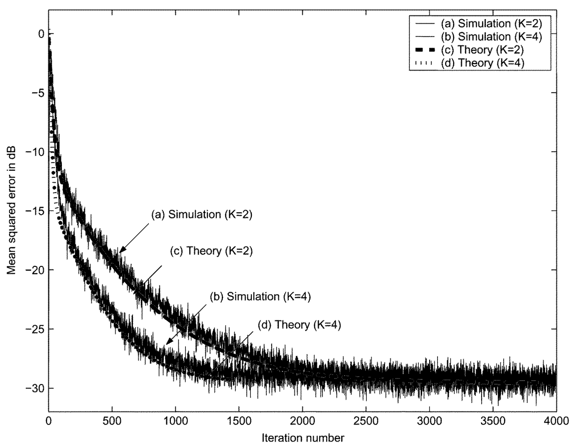

Fig. 3. Learning curves of NDR-LMS for colored Gaussian input using $\mu = 0.01$ for $K = 2$ and $K = 4$ [Input: Gaussian ARMA(2, 2); System: FIR (16)].

As a result, a highly correlated Gaussian signal of which the eigenvalue spread is around 105 is generated. The measurement noise $v(i)$ is added to $y(i)$ such that the signal-to-noise ratio (SNR) is 30 dB. The simulation results shown are obtained by ensemble averaging over 100 independent trials. In Tables I and II, we evaluate the stability bounds from Theorem 1 for the DR-LMS and NDR-LMS algorithms. This fact is numerically verified in Fig. 1, where simulated MSE curves are plotted as a function of the step size. Fig. 2 shows the learning curves of the DR-LMS algorithm for colored Gaussian input. The step size is set to $\mu = 0.005$. Fig. 2 shows how close the simulation results are to the theoretical results, where F and γ were evaluated via ensemble averaging. In Fig. 3, the learning curves for the NDR-LMS algorithm are shown for $\mu = 0.01$. Fig. 4 shows the steady-state MSE curve of the NDR-LMS algorithm for colored Gaussian input as a function of the step size. The step size μ varies from 0.01 to 0.3. 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 steady state and then averaging 200 independent trials. The simulation results are in good agreement with the theoretical results for small step sizes, but they deviate from the theoretical value for larger step sizes and

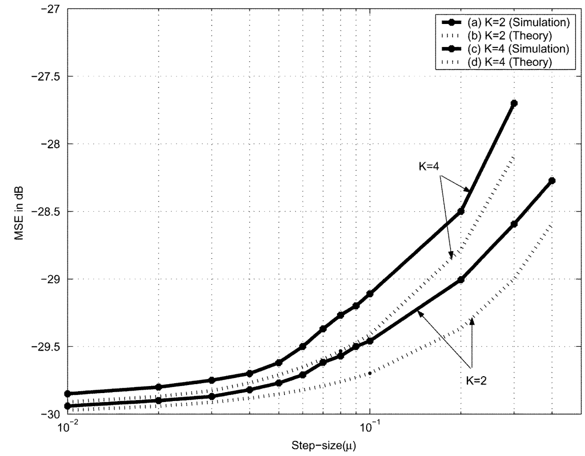

Fig. 4. Steady-state MSE curves of NDR-LMS for colored Gaussian input in stationary environments for $K = 2$ and $K = 4$ [Input: Gaussian ARMA(2, 2); System: FIR (16)].

larger K. This is because of the independent assumption A.2. The steady-state MSE curve of the DR-LMS algorithm has a similar behavior to Fig. 4.

V. Conclusion

In this letter, we performed a mean-square performance analysis of the class of data-reusing adaptive algorithms, such as DR-LMS and NDR-LMS, without restricting the distribution of the input data to being Gaussian. The arguments were based on the energy conservation approach of [8], and they bring forth commonalities between this class of algorithms and affine projection methods.

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. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunications*. Norwell, MA: Kluwer, 2000.

[3] B. A. Schnaufer and W. K. Jenkins, "New data-reusing LMS algorithms for improved covergence," in *Proc. Asilomar Conf.*, Pacific Groves, CA, May 1993, pp. 1584–1588.

[4] R. A. Soni, W. K. Jenkins, and K. A. Gallivan, "Acceleration of normalized adaptive filtering data-reusing methods using the Tchebyshev and conjugate gradient methods," in *Proc. Int. Symp. Circuits Systems*, vol. 5, 1998, pp. 309–312.

[5] S. G. Sankaran and A. A. L. Beex, "Convergence behavior of affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1086–1096, Apr. 2000.

[6] H.-C. Shin and A. H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 90–102, Jan. 2004.

[7] R. A. Soni, K. A. Gallivan, and W. K. Jenkins, "Low-complexity data-reusing techniques for adaptive filtering," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 394–405, Feb. 2004.

[8] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[9] R. A. Soni, K. A. Gallivan, and W. K. Jenkins, "Convergence properties of affine projection and normalized data reusing methods," in *Proc. Asilomar Conf.*, vol. 2, Pacific Groves, CA, Nov. 1998, pp. 1166–1170.

[10] M. L. R. Campos, P. S. R. Diniz, and J. Apolinário Jr., "On normalized data-reusing and affine projection algorithms," in *Proc. Int. Conf. Electronics, Circuits, Systems*, vol. 2, Sep. 1999, pp. 843–846.

[11] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639–652, Mar. 2003.