Mean-Square Performance of Adaptive Filters Using Averaging Theory

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

Abstract—

This paper uses averaging analysis to study the mean-square performance of adaptive filters, not only in terms of stability conditions but also in terms of expressions for the mean-square error and the mean-square deviation of the filters, as well as in terms of the transient performance of the corresponding partially averaged systems. The treatment relies on energy conservation arguments. Simulation results illustrate the analysis and the derived performance expressions.

I. Introduction

There are several ingenious approaches that study the performance of adaptive filters without relying on the independence assumptions such as averaging analysis and the ODE method (e.g., [1]–[4]). Studies based on these methods assume small step-sizes and they are primarily concerned with stability statements as opposed to performance statements. For example, in the ODE method, one would replace a difference equation characterizing the update equation of an adaptive filter by a differential equation. Subsequently, one would infer conditions for the stability of the adaptive filter from the stability of the corresponding ODE. In such studies, one seldom progresses beyond stability studies to derive expressions for the mean-square performance of an adaptive filter (e.g., in terms of its excess mean-square error (EMSE) or its mean-square deviation (MSD)) or to discuss the filter transient response and its learning curve behavior. The contribution of this work is to use averaging analysis to characterize the mean-square performance of adaptive filters in terms of their EMSE and MSD and in terms of a state-space characterization for their transient behavior. The results provide further support to the conclusion that results obtained by an independence analysis tend to agree with adaptive filter performance when the step-size is sufficiently small [5], [6].

In the sequel, small boldface letters are used to denote vectors and capital letters are used to denote matrices, e.g., $\mathbf{c}$ and C. All vectors are *column vectors* except for regression vector denoted by $\mathbf{u}_i$, which is taken to be a row vector for convenience of notation.

II. Data Model

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, $v(i)$ is measurement noise, and $\mathbf{u}_i$ denotes $1 \times M$ row input (regressor) vectors with a positive-definite covariance matrix, $R_u = E\left[\mathbf{u}_i^* \mathbf{u}_i\right]$. Consider further adaptive filter weight-error vector updates of the form

$$\boxed{\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} + \mu \mathbf{f}(i, \tilde{\mathbf{w}}_{i-1})} \tag{2}$$

for some stochastic vector function $\mathbf{f}(\cdot,\cdot)$ and where $\tilde{\mathbf{w}}_i = \mathbf{w}^\circ - \mathbf{w}_i$. For example, for LMS we have

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu \mathbf{u}_i^*(\mathbf{u}_i \tilde{\mathbf{w}}_{i-1} + v(i)) \tag{3}$$

for which $\mathbf{f}(i, \tilde{\mathbf{w}}_{i-1}) = -\mathbf{u}_i^*(\mathbf{u}_i \tilde{\mathbf{w}}_{i-1} + v(i))$. Introduce the averaged function $\mathbf{f}_{av}(i, \tilde{\mathbf{w}}_{i-1}) = E\mathbf{f}(i, \tilde{\mathbf{w}}_{i-1})$, where $\tilde{\mathbf{w}}_{i-1}$ is considered *constant* for the computation of the expected value, and define the *averaged system*

$$\boxed{\tilde{\mathbf{w}}_i^{av} = \tilde{\mathbf{w}}_{i-1}^{av} + \mu \mathbf{f}_{av}(i, \tilde{\mathbf{w}}_{i-1}^{av}), \quad \tilde{\mathbf{w}}_{-1}^{av} = \tilde{\mathbf{w}}_{-1}} \tag{4}$$

where the stochastic function $\mathbf{f}(\cdot,\cdot)$ in (2) is replaced by its averaged value and, accordingly, the corresponding weight-error vectors are denoted by $\tilde{\mathbf{w}}_i^{av}$. Define also the *partially averaged system*:

$$\boxed{\tilde{\mathbf{w}}_i^{pav} = [I + \mu \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)] \tilde{\mathbf{w}}_{i-1}^{pav} + \mu[\mathbf{f}(i,0) - \mathbf{f}_{av}(i,0)]} \tag{5}$$

where $\nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)$ denotes the gradient vector of $\mathbf{f}_{av}(i, \tilde{\mathbf{w}})$ with respect to $\tilde{\mathbf{w}}$ evaluated at the origin. Again, for LMS, we have $\mathbf{f}_{av,\text{LMS}}(i, \tilde{\mathbf{w}}_{i-1}) = -R_u \tilde{\mathbf{w}}_{i-1}$, so that

$$\tilde{\mathbf{w}}_i^{av,\text{LMS}} = (I - \mu R_u) \tilde{\mathbf{w}}_{i-1}^{av,\text{LMS}} \tag{6}$$

$$\tilde{\mathbf{w}}_i^{pav,\text{LMS}} = (I - \mu R_u) \tilde{\mathbf{w}}_{i-1}^{pav,\text{LMS}} - \mu \mathbf{u}_i^* v(i) \tag{7}$$

Comparing (7) with (3), we see that the random matrix $\mathbf{u}_i^* \mathbf{u}_i$ is replaced by its ensemble average R_u. Averaging analysis [1], [2] usually justifies substituting such random matrices by their ensemble averages when the step-sizes are sufficiently small.

The next result is from [2]. Its proof requires that the regressor sequence $\{\mathbf{u}_i\}$ be bounded and that it satisfy a certain mixing condition (in loose terms, the correlation between $\mathbf{u}_i$ and $\mathbf{u}_j$ should "die out" as the time difference $|i-j|$ increases).

Theorem 1 [**Averaging result**] *Assume that the following conditions hold:*

1. $\tilde{\mathbf{w}}^{av} = 0$ *is an exponentially-stable equilibrium point of the averaged system* (4) *with decay rate* $O(\mu)$.

2. $\nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(i, \tilde{\mathbf{w}})$ *exists and is continuous at the origin.*

3. The gradient vector satisfies the Lipschitz condition $\|\nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(i,\mathbf{a}) - \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(i,\mathbf{b})\| \le c\|\mathbf{a} - \mathbf{b}\|$, *for some* $c > 0$.

Under these conditions, $\tilde{\mathbf{w}}_i$ *obtained from* (2) *satisfies*

$$\lim_{\mu \to 0} \sup_{i \ge 0} E\|\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_i^{pav}\| = 0, \quad \lim_{\mu \to 0} \lim_{i \to \infty} E\|\tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_i^{pav}\|^2 = 0$$

$$\lim_{\mu \to 0} \lim_{i \to \infty} \left(\frac{1}{\mu} E\tilde{\mathbf{w}}_i \tilde{\mathbf{w}}_i^*\right) = \lim_{\mu \to 0} \lim_{i \to \infty} \left(\frac{1}{\mu} E\tilde{\mathbf{w}}_i^{pav} \tilde{\mathbf{w}}_i^{pav,*}\right)$$

□

The above result indicates that if the step-size μ is sufficiently small, and for fairly general adaptive schemes, the weight-error vector $\tilde{\mathbf{w}}_i$ of (2) remains close to the *partially averaged* weight-error vector $\tilde{\mathbf{w}}_i^{pav}$ of (5). Consequently, for small step-sizes, we may evaluate the performance of an adaptive filter by examining the performance of its partially averaged recursion (5), as we shall proceed to do.

This work was supported in part by the National Science Foundation under grants ECS-0401188 and CCF-0208573. This work was performed while H. Shin was a visiting graduate student at the UCLA Adaptive Systems Laboratory. His work was 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, Korea.

Hyun-Chool Shin and Woo-Jin Song are with Division of Electronics and Computer Engineering, Pohang University of Science and Technology (POSTECH), Korea.

Ali H. Sayed is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095.

0-7803-8622-1/04/$20.00

III. Mean-Square Performance

Our first objective is to evaluate the steady-state mean-square error of an adaptive filter, i.e., to compute MSE $= E|e(i)|^2$ as $i \to \infty$, where $e(i) = d(i) - \mathbf{u}_i \mathbf{w}_{i-1}$ is the output estimation error at time i. To do so, we shall rely on the energy-conservation approach of [7][Ch. 6] and on the *partially averaged system* (5). Table I lists $\mathbf{f}(i, \tilde{\mathbf{w}}_{i-1})$, $\nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)$, and $(\mathbf{f}(i,0) - \mathbf{f}_{av}(i,0))$ for several adaptive filters,[1] with $\mathbf{f}_\triangle(i,0)$ defined by

$$\mathbf{f}(i,0) - \mathbf{f}_{av}(i,0) = \mathbf{u}_i^* \mathbf{f}_\triangle(i,0)$$

so that (5) becomes

$$\boxed{\tilde{\mathbf{w}}_i^{pav} = (I + \mu \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)) \tilde{\mathbf{w}}_{i-1}^{pav} + \mu \mathbf{u}_i^* \mathbf{f}_\triangle(i,0)} \tag{8}$$

A. Energy Conservation Relation

Let Σ denote any $M \times M$ positive-definite matrix (which we are free to choose), and define the weighted *a-priori* and *a-posteriori* error signals

$$\boxed{e_a^\Sigma(i) \triangleq \mathbf{u}_i \Sigma (I + \mu \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)) \tilde{\mathbf{w}}_{i-1}^{pav}, \quad e_p^\Sigma(i) \triangleq \mathbf{u}_i \Sigma \tilde{\mathbf{w}}_i^{pav}} \tag{9}$$

If we multiply both sides of (8) by $\mathbf{u}_i \Sigma$ from the left we find that the *a-priori* and *a-posteriori* estimation errors $\{e_p^\Sigma(i), e_a^\Sigma(i)\}$ are related via:

$$\boxed{e_p^\Sigma(i) = e_a^\Sigma(i) + \mu \mathbf{u}_i \Sigma \mathbf{u}_i^* \mathbf{f}_\triangle(i,0)} \tag{10}$$

We can use (10) to solve for $\mathbf{f}_\triangle(i,0)$,

$$\mathbf{f}_\triangle(i,0) = -\frac{e_a^\Sigma(i) - e_p^\Sigma(i)}{\mu \|\mathbf{u}_i\|_\Sigma^2} \tag{11}$$

where the notation $\|\mathbf{u}_i\|_\Sigma^2$ stands for $\mathbf{u}_i \Sigma \mathbf{u}_i^*$. Substituting into (8) we get

$$\tilde{\mathbf{w}}_i^{pav} = (I + \mu \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)) \tilde{\mathbf{w}}_{i-1}^{pav} - \frac{\mathbf{u}_i^* (e_a^\Sigma(i) - e_p^\Sigma(i))}{\|\mathbf{u}_i\|_\Sigma^2} \tag{12}$$

which can be rearranged as

$$\tilde{\mathbf{w}}_i^{pav} + \frac{\mathbf{u}_i^*}{\|\mathbf{u}_i\|_\Sigma^2} e_a^\Sigma(i) = (I + \mu \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)) \tilde{\mathbf{w}}_{i-1}^{pav} + \frac{\mathbf{u}_i^*}{\|\mathbf{u}_i\|_\Sigma^2} e_p^\Sigma(i) \tag{13}$$

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

$$\boxed{\begin{aligned} &\|\tilde{\mathbf{w}}_i^{pav}\|_\Sigma^2 + \frac{|e_a^\Sigma(i)|^2}{\|\mathbf{u}_i\|_\Sigma^2} = \\ &\|\tilde{\mathbf{w}}_{i-1}^{pav}\|^2_{(I+\mu\nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0))^*\Sigma(I+\mu\nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0))} + \frac{|e_p^\Sigma(i)|^2}{\|\mathbf{u}_i\|_\Sigma^2} \end{aligned}} \tag{14}$$

The important fact to emphasize is that *no approximations* are used to establish the energy relation (14); 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 (14) is the extension to *partially averaged systems* of the energy-conservation relation described in [7].

[1] For the sign algorithm (SA), we assume that $e(i)$ and $v(i)$ are jointly Gaussian and use Price's theorem [20]. See [21] for details. Also, for the complex case, terms of the form a^3 should be replaced by $a|a|^2$.

B. Weighted Variance Relation

The relevance of (14) to mean-square analysis can be seen as follows. Replacing $e_p^\Sigma(i)$ by its equivalent expression (10) in terms of $e_a^\Sigma(i)$ and $\mathbf{f}_\triangle(i,0)$ we get

$$\begin{aligned} &\|\mathbf{u}_i\|_\Sigma^2 \cdot \|\tilde{\mathbf{w}}_i^{pav}\|_\Sigma^2 + |e_a^\Sigma(i)|^2 = \\ &\|\mathbf{u}_i\|_\Sigma^2 \cdot \|\tilde{\mathbf{w}}_{i-1}^{pav}\|_{\Sigma'}^2 + |e_a^\Sigma(i) + \mu \|\mathbf{u}_i\|_\Sigma^2 \mathbf{f}_\triangle(i,0)|^2 \end{aligned} \tag{15}$$

where $\Sigma' = [I + \mu \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)]^* \Sigma [I + \mu \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0)]$. After expanding the rightmost term of (15), the right-hand side (RHS) becomes:

$$\begin{aligned} \text{RHS} \;=\; & \|\mathbf{u}_i\|_\Sigma^2 \cdot \|\tilde{\mathbf{w}}_{i-1}^{pav}\|_{\Sigma'}^2 + |e_a^\Sigma(i)|^2 \\ & + \mu \|\mathbf{u}_i\|_\Sigma^2 e_a^{\Sigma*}(i) \mathbf{f}_\triangle(i,0) + \mu \|\mathbf{u}_i\|_\Sigma^2 \mathbf{f}_\triangle^*(i,0) e_a^\Sigma(i) \\ & + \mu^2 |\mathbf{f}_\triangle(i,0)|^2 (\|\mathbf{u}_i\|_\Sigma^2)^2 \end{aligned} \tag{16}$$

Normally, the event $\|\mathbf{u}_i\|_\Sigma^2 = 0$ occurs with probability zero. We can eliminate $\|\mathbf{u}_i\|_\Sigma^2$ from both sides of (15) and take expectations to find that

$$\begin{aligned} E\|\tilde{\mathbf{w}}_i^{pav}\|_\Sigma^2 \;=\; & E\|\tilde{\mathbf{w}}_{i-1}^{pav}\|_{\Sigma'}^2 \\ & + \mu E[e_a^{\Sigma*}(i) \mathbf{f}_\triangle(i,0)] + \mu E[\mathbf{f}_\triangle^*(i,0) e_a^\Sigma(i)] \\ & + \mu^2 E[|\mathbf{f}_\triangle(i,0)|^2 \|\mathbf{u}_i\|_\Sigma^2] \end{aligned} \tag{17}$$

Under the often realistic assumption that

A.1 *The noise $\{v(i)\}$ is i.i.d. and statistically independent of $\mathbf{u}_j$ for all j,*

and using $E[\mathbf{f}_\triangle(i,0)] = 0$ for the algorithms in Table I, we find that, except for NLMS, the variance relation (17) reduces to

$$\boxed{E\|\tilde{\mathbf{w}}_i^{pav}\|_\Sigma^2 = E\|\tilde{\mathbf{w}}_{i-1}^{pav}\|_{\Sigma'}^2 + \mu^2 E[|\mathbf{f}_\triangle(i,0)|^2] E\|\mathbf{u}_i\|_\Sigma^2} \tag{18}$$

For NLMS, we shall use the approximation

$$E[|\mathbf{f}_\triangle(i,0)|^2 \|\mathbf{u}_i\|_\Sigma^2] = \sigma_v^2 E\left(\frac{\|\mathbf{u}_i\|_\Sigma^2}{\|\mathbf{u}_i\|^4}\right) \approx \frac{\sigma_v^2}{[\text{Tr}(R_u)]^2} \cdot E\|\mathbf{u}_i\|_\Sigma^2$$

so that (18) can be also used for NLMS with $E[|\mathbf{f}_\triangle(i,0)|^2]$ replaced by $\sigma_v^2/[\text{Tr}(R_u)]^2$.

The subsequent analysis is simplified if we introduce a convenient change of coordinates by appealing to the eigendecomposition $R_u = Q\Lambda Q^*$, where Λ is a diagonal matrix with the eigenvalues of R_u and Q is a unitary matrix (i.e., it satisfies $QQ^* = Q^*Q = I$). Define the transformed quantities:

$$\begin{aligned} \bar{\mathbf{w}}_i^{pav} &\triangleq Q^* \tilde{\mathbf{w}}_i^{pav} & \nabla_{\tilde{\mathbf{w}}} \bar{\mathbf{f}}_{av}(0) &\triangleq Q^* \nabla_{\tilde{\mathbf{w}}} \mathbf{f}_{av}(0) Q \\ \bar{\mathbf{u}}_i &\triangleq \mathbf{u}_i Q & \bar{\Sigma} &\triangleq Q^* \Sigma Q \end{aligned}$$

Under this change of variables the variance relation (18) retains a similar form:

$$\boxed{E\|\bar{\mathbf{w}}_i^{pav}\|_{\bar{\Sigma}}^2 = E\|\bar{\mathbf{w}}_{i-1}^{pav}\|_{\bar{\Sigma}'}^2 + \mu^2 E[|\mathbf{f}_\triangle(i,0)|^2] E\|\bar{\mathbf{u}}_i\|_{\bar{\Sigma}}^2} \tag{19}$$

where

$$\boxed{\bar{\Sigma}' = \bar{\Sigma} + 2\mu \bar{\Sigma} \nabla_{\tilde{\mathbf{w}}} \bar{\mathbf{f}}_{av}(0) + \mu^2 \nabla_{\tilde{\mathbf{w}}} \bar{\mathbf{f}}_{av}(0) \bar{\Sigma} \nabla_{\tilde{\mathbf{w}}} \bar{\mathbf{f}}_{av}(0)} \tag{20}$$

Table II lists $\nabla_{\tilde{\mathbf{w}}} \bar{\mathbf{f}}_{av}(0)$ and $E[|\mathbf{f}_\triangle(i,0)|^2]$ for the adaptive filters of Table I. Note that $\nabla_{\tilde{\mathbf{w}}} \bar{\mathbf{f}}_{av}(0)$ is a diagonal matrix.

TABLE I

Examples of $\{\mathbf{f}(i,\tilde{\mathbf{w}}_{i-1}), \nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0), \mathbf{f}(i,0)-\mathbf{f}_{av}(i,0)\}$ for various adaptive filters.

Algorithm	$\mathbf{f}(i,\tilde{\mathbf{w}}_{i-1})$	$\nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0)$	$\mathbf{f}(i,0)-\mathbf{f}_{av}(i,0)$	$\mathbf{f}_{\triangle}(i,0)$
LMS	$-\mathbf{u}_i^*(\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}+v(i))$	$-R_u$	$-\mathbf{u}_i^*v(i)$	$-v(i)$
NLMS	$\frac{-\mathbf{u}_i^*(\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}+v(i))}{\lVert\mathbf{u}_i\rVert^2}$	$-E\left[\frac{\mathbf{u}_i^*\mathbf{u}_i}{\lVert\mathbf{u}_i\rVert^2}\right]\approx-\frac{R_u}{\text{Tr}(R_u)}$	$-\frac{\mathbf{u}_i^*v(i)}{\lVert\mathbf{u}_i\rVert^2}$	$-\frac{v(i)}{\lVert\mathbf{u}_i\rVert^2}$
LMF	$-\mathbf{u}_i^*(\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}+v(i))^3$	$-3R_u\sigma_v^2(\text{real})$ $-2R_u\sigma_v^2(\text{complex})$	$-\mathbf{u}_i^*v^3(i)$	$-v^3(i)$
LMMN	$-\delta\mathbf{u}_i^*(\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}+v(i))$ $-(1-\delta)\mathbf{u}_i^*(\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}+v(i))^3$	$-\delta R_u-3(1-\delta)R_u\sigma_v^2(\text{real})$ $-\delta R_u-2(1-\delta)R_u\sigma_v^2(\text{complex})$	$-\delta\mathbf{u}_i^*v(i)$ $-(1-\delta)\mathbf{u}_i^*v^3(i)$	$-\delta v(i)$ $-(1-\delta)v^3(i)$
SA	$-\mathbf{u}_i^*\text{sgn}(\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}+v(i))$	$-\sqrt{\frac{2}{\pi}}\frac{R_u}{\sigma_v}$	$-\mathbf{u}_i^*\text{sgn}(v(i))$	$-\text{sgn}(v(i))$

TABLE II

EMSE and Examples of $\{\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0), E[|\mathbf{f}_{\triangle}(i,0)|^2\}$ for various adaptive filters.

Algorithms	$\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)$	$E[\lvert\mathbf{f}_{\triangle}(i,0)\rvert^2]$	EMSE (small μ)	EMSE (with I. A.)
LMS	$-\Lambda$	σ_v^2	$\frac{\mu\sigma_v^2\text{Tr}(R_u)}{2}$	$\frac{\mu\sigma_v^2\text{Tr}(R_u)}{2}$
NLMS	$-\frac{\Lambda}{\text{Tr}(R_u)}$	$\frac{\sigma_v^2}{[\text{Tr}(R_u)]^2}$	$\frac{\mu\sigma_v^2}{2}$	$\frac{\mu\sigma_v^2}{2-\mu}$
LMF	$-3\Lambda\sigma_v^2(\text{real})$ $-2\Lambda\sigma_v^2(\text{complex})$	$\xi_v^6=E[v^6(i)]$	$\frac{\mu}{2}\left(\frac{\xi_v^6}{3\sigma_v^2}\right)\text{Tr}(R_u)$	$\frac{\mu}{2}\left(\frac{\xi_v^6}{3\sigma_v^2}\right)\text{Tr}(R_u)$
LMMN	$-\delta\Lambda-3(1-\delta)\Lambda\sigma_v^2(\text{real})$ $-\delta\Lambda-2(1-\delta)\Lambda\sigma_v^2(\text{complex})$	$\delta\sigma_v^2+(1-\delta)^2\xi_v^6$ $+2\delta(1-\delta)\xi_v^4$	$\frac{\mu a}{2b}\text{Tr}(R_u)$	$\frac{\mu a}{2b}\text{Tr}(R_u)$
SA	$-\sqrt{\frac{2}{\pi}}\frac{\Lambda}{\sigma_v}$	1	$\frac{\alpha}{2}\sqrt{4\sigma_v^2}$	$\frac{\alpha}{2}(\alpha+\sqrt{\alpha^2+4\sigma_v^2})$

C. Steady-State Mean-Square Performance

Relation (19) can now be used to deduce an approximate expression for the filter excess mean-square error (EMSE), which is defined as

$$\text{EMSE}=\lim_{i\to\infty}E|\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}|^2$$

Using $\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}\approx\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}^{pav}$, we set

$$\text{EMSE}\approx\lim_{i\to\infty}E|\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}^{pav}|^2 \tag{21}$$

To evaluate this EMSE we note from (8) that the dyanmics of the weight-error vector $\tilde{\mathbf{w}}_i^{pav}$ is determined by the eigenvalues of the coefficient matrix

$$I+\mu\nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0) \tag{22}$$

where, from Table I, the matrix $\nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0)$ is negative-definite. For very small step-sizes, the eigenvalues of the matrix (22) are inside the unit circle but close to unity so that the variations in $\tilde{\mathbf{w}}_i^{pav}$ occur slowly. Moreover, from (8), we also have that

$$\begin{aligned}\tilde{\mathbf{w}}_{i-1}^{pav} &= [I+\mu\nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0)]^i\,\tilde{\mathbf{w}}_{-1}^{pav}+\\ &\quad \mu\sum_{j=0}^{i-2}[I+\mu\nabla_{\tilde{\mathbf{w}}}\mathbf{f}_{av}(0)]^j\mathbf{u}_{i-1-j}^*\mathbf{f}_{\triangle}(i-1-j,0)\end{aligned}$$

where the first term on the right-hand side is independent of $\mathbf{u}_i$, while the second term is a function of the regressors $\{\mathbf{u}_{i-1-j}\}$. Since we are assuming sufficiently small step-sizes ($\mu\to0$), we may ignore the dependence between $\mathbf{u}_i$ and this second term. It follows that we may assume that $\mathbf{u}_i$ and $\tilde{\mathbf{w}}_{i-1}^{pav}$ are essentially independent of each other for small step-sizes. Then (21) gives

$$\text{EMSE}\approx\lim_{i\to\infty}E\|\tilde{\mathbf{w}}_{i-1}^{pav}\|_{R_u}^2=\lim_{i\to\infty}E\|\bar{\mathbf{w}}_{i-1}^{pav}\|_{\Lambda}^2$$

Taking the limit of (19) as $i\to\infty$, and using the steady-state condition $E\|\bar{\mathbf{w}}_i^{pav}\|_{\bar{\Sigma}}^2=E\|\bar{\mathbf{w}}_{i-1}^{pav}\|_{\bar{\Sigma}}^2$, we obtain

$$\boxed{\begin{aligned}E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{-2\bar{\Sigma}\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)-\mu\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)\bar{\Sigma}\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)}=\\ \mu E[|\mathbf{f}_{\triangle}(i,0)|^2]\text{Tr}(\Lambda\bar{\Sigma})\end{aligned}} \tag{23}$$

In order to evaluate the EMSE, we need to select $\bar{\Sigma}$ such that the weighting matrix on the left-hand side is equal to Λ, i.e.,

$$-2\bar{\Sigma}\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)-\mu\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)\bar{\Sigma}\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)=\Lambda \tag{24}$$

We illustrate this procedure for several algorithms.

1. **LMS**. We substitute $\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)$ and $E|\mathbf{f}_{\triangle}(i,0)|^2$ by $-\Lambda$ and σ_v^2, respectively. Then (24) becomes $2\bar{\Sigma}\Lambda-\mu\Lambda\bar{\Sigma}\Lambda=\Lambda$, i.e., $\bar{\Sigma}$ should be selected as the diagonal matrix $\bar{\Sigma}=(2I-\mu\Lambda)^{-1}$. Then the LHS of (23) becomes the filter EMSE and (23) gives

$$\text{EMSE}=\mu\sigma_v^2\text{Tr}[\Lambda(2I-\mu\Lambda)^{-1}]=\mu\sigma_v^2\sum_{i=0}^{M-1}\frac{\lambda_i}{2-\mu\lambda_i} \tag{25}$$

When $\mu\lambda_i\ll2$, the EMSE expression reduces to

$$\boxed{\text{EMSE}=\frac{\mu\sigma_v^2\text{Tr}(R_u)}{2}}$$

2. **NLMS** For NLMS, by substituting $\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)$ and $E|\mathbf{f}_{\triangle}(i,0)|^2$ by $-\frac{\Lambda}{\text{Tr}(R_u)}$ and $\frac{\sigma_v^2}{\text{Tr}^2(R_u)}$, respectively, relation (23) becomes

$$E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{2\bar{\Sigma}\frac{\Lambda}{\text{Tr}(R_u)}-\mu\frac{\Lambda\bar{\Sigma}\Lambda}{\text{Tr}^2(R_u)}} = \mu\frac{\sigma_v^2}{\text{Tr}^2(R_u)}\text{Tr}(\Lambda\bar{\Sigma}) \tag{26}$$

Then we choose $\bar{\Sigma}$ so that

$$2\bar{\Sigma}\frac{\Lambda}{\text{Tr}(R_u)} - \mu\frac{\Lambda\bar{\Sigma}\Lambda}{\text{Tr}^2(R_u)} = \Lambda$$

i.e.,

$$\bar{\Sigma} = (2\text{Tr}(R_u)I - \mu\Lambda)^{-1}\text{Tr}^2(R_u)$$

and the LHS of (26) becomes the filter EMSE:

$$\begin{aligned}\text{EMSE} &= \mu\sigma_v^2\text{Tr}[\Lambda(2\text{Tr}(R_u)I-\mu\Lambda)^{-1}] \\ &= \mu\sigma_v^2\sum_{i=0}^{M-1}\frac{\lambda_i}{2\text{Tr}(R_u)-\mu\lambda_i}\end{aligned} \tag{27}$$

When $\mu\lambda_i \ll 2\text{Tr}(R_u)$, this EMSE expression reduces to

$$\boxed{\text{EMSE} = \frac{\mu\sigma_v^2}{2}}$$

3. **LMF**. To obtain the EMSE for LMF with real-valued data, we substitute $\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)$ and $E|\mathbf{f}_{\triangle}(i,0)|^2$ by $-3\sigma_v^2\Lambda$ and $\xi_v^6 = E[v^6(i)]$, respectively. Then relation (24) becomes

$$6\sigma_v^2\bar{\Sigma}\Lambda - 9\mu\xi_v^6\Lambda\bar{\Sigma}\Lambda = \Lambda$$

so that

$$\bar{\Sigma} = (6\sigma_v^2 I - 9\mu\xi_v^6\Lambda)^{-1}$$

Then the LHS of (23) becomes the filter EMSE and it leads to

$$\begin{aligned}\text{EMSE} &= \mu\xi_v^6\text{Tr}[\Lambda(6\sigma_v^2 I - 9\mu\xi_v^6\Lambda)^{-1}] \\ &= \mu\xi_v^6\sum_{i=0}^{M-1}\frac{\lambda_i}{6\sigma_v^2 - 9\mu\xi_v^6\lambda_i}\end{aligned} \tag{28}$$

When $9\mu\xi_v^6\lambda_i \ll 6\sigma_v^2$, this EMSE expression reduces to

$$\boxed{\text{EMSE} = \frac{\mu}{2}\left(\frac{\xi_v^6}{3\sigma_v^2}\right)\text{Tr}(R_u)}$$

4. **LMMN**. For LMMN with real-valued data, we have

$$\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0) = \delta\Lambda - 3\bar{\delta}\sigma_v^2\Lambda$$

and

$$E|\mathbf{f}_{\triangle}(i,0)|^2 = \delta\sigma_v^2 + \bar{\delta}^2\xi_v^6 + \delta\bar{\delta}\xi_v^4$$

where $\bar{\delta} = 1-\delta$ and $\xi_v^4 = E[v^4(i)]$. After substitution, relation (24) becomes

$$2(\delta+3\bar{\delta}\sigma_v^2)\bar{\Sigma}\Lambda - \mu(\delta+3\bar{\delta}\sigma_v^2)^2\Lambda\bar{\Sigma}\Lambda = \Lambda$$

so that

$$\bar{\Sigma} = (2(\delta+3\bar{\delta}\sigma_v^2)I - \mu(\delta+3\bar{\delta}\sigma_v^2)^2\Lambda)^{-1}$$

Then the LHS of (23) becomes the EMSE and it leads to

$$\text{EMSE} = \mu a\sum_{i=0}^{M-1}\frac{\lambda_i}{2b-\mu c\lambda_i} \tag{29}$$

where we introduced the constants

$$\begin{aligned} a &= \delta\sigma_v^2 + 2\delta\bar{\delta}\xi_v^4 + \bar{\delta}^2\xi_v^6 \\ b &= \delta + 3\bar{\delta}\sigma_v^2 \\ c &= \delta^2 + 6\delta\bar{\delta}\sigma_v^2 + 9\bar{\delta}^2\xi_v^4\end{aligned}$$

When $\mu c\lambda_i \ll 2b$, this EMSE expression reduces to

$$\boxed{\text{EMSE} = \frac{\mu a}{2b}\text{Tr}(R_u)}$$

For complex valued data, we replace $e^3(i)$ by $e(i)|e(i)|^2$ and assume the noise is circular, i.e., $Ev^2(i) = 0$. Then repeating the above arguments we find that expression (29) is still valid but with b and c replaced by

$$b' = \delta + 2\bar{\delta}\sigma_v^2, \quad c' = \delta^2 + 4\delta\bar{\delta}\sigma_v^2 + 4\bar{\delta}^2\xi_v^4$$

5. **Sign Algorithm**. We substitute $\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0)$ and $E|\mathbf{f}_{\triangle}(i,0)|^2$ by $-\sqrt{\frac{2}{\pi}}\frac{\Lambda}{\sigma_v}$ and 1, respectively. Then relation (24) becomes

$$\sqrt{\frac{2}{\pi}}\bar{\Sigma}\frac{\Lambda}{\sigma_v} - \mu\frac{2}{\pi}\frac{1}{\sigma_v^2}\Lambda\bar{\Sigma}\Lambda = \Lambda$$

so that

$$\bar{\Sigma} = \left(\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_v}I - \mu\frac{2}{\pi}\frac{\Lambda}{\sigma_v^2}\right)^{-1}$$

Then the LHS of (23) becomes the EMSE and it leads to

$$\begin{aligned}\text{EMSE} &= \mu\text{Tr}\left(\Lambda\left(\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_v}I - \mu\frac{2}{\pi}\frac{\Lambda}{\sigma_v^2}\right)^{-1}\right) \\ &= \mu\sum_{i=0}^{M-1}\frac{\pi\sigma_v^2\lambda_i}{\sqrt{2\pi}\sigma_v - 2\mu\lambda_i}\end{aligned} \tag{30}$$

When $2\mu\lambda_i \ll \sqrt{2\pi}\sigma_v$, this EMSE expression reduces to

$$\boxed{\text{EMSE} = \mu\sqrt{\frac{\pi}{2}}\sigma_v\text{Tr}(R_u) \quad \text{or} \quad \frac{\alpha}{2}\sqrt{4\sigma_v^2}}$$

where $\alpha = \sqrt{\frac{\pi}{8}}\mu\text{Tr}(R_u)$.

Table II summarizes the derived EMSE for various adaptive filters for sufficiently small μ. These expressions, which have been derived here using averaging theory, are essentially identical to the ones derived in the literature by means of the independence assumptions.

In Figs. 1– 2 we illustrate the theoretical results presented in this paper for LMS and NLMS by carrying out computer simulations in a channel estimation scenario. Similar results apply to LMMN and SA. In the simulations, the unknown channel has 10 taps and the input is Gaussian of unit variance. The noise is chosen to be white Gaussian of variance $\sigma_v^2 = 10^{-3}$. Each simulation result is obtained by averaging the last 1000 instantaneous squared error over 100 independent trials. The simulation results show good agreement with the theoretical results for small step-sizes, but deviate from the theoretical ones for larger step-sizes. This is because the *partially averaged system* is valid only for small step-sizes.

IV. Transient Analysis

We now study the transient performance of the partially averaged system (8) in terms of its stability and learning curve behavior.

Our analysis starts from the weighted variance relation (19). In addition, taking expectations of both sides of (8) and changing variables into transformed quantities, we obtain the following result for the evolution of the mean of the weight-error vector:

$$E[\tilde{\mathbf{w}}_i^{pav}] = (I + \mu\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0))E[\tilde{\mathbf{w}}_{i-1}^{pav}] \tag{31}$$

since $E[\mathbf{u}_i^*\mathbf{f}_{\Delta}(i,0)] = 0$. Relations (19) and (31) can be used to derive conditions for mean-square stability and to characterize the learning curve of the system.

Following the approach described in [7][Ch. 9] and [9], we observe the interesting fact that $\bar{\Sigma}'$ will be diagonal if $\bar{\Sigma}$ is. In

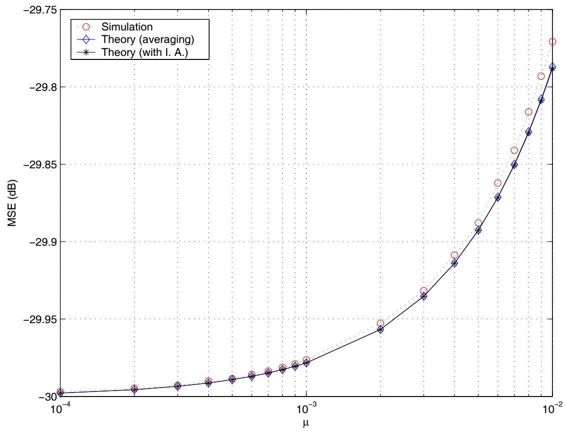

Fig. 1. Experimental and theoretical MSE versus μ for LMS (Input: Gaussian, System: FIR (10), SNR=30dB)

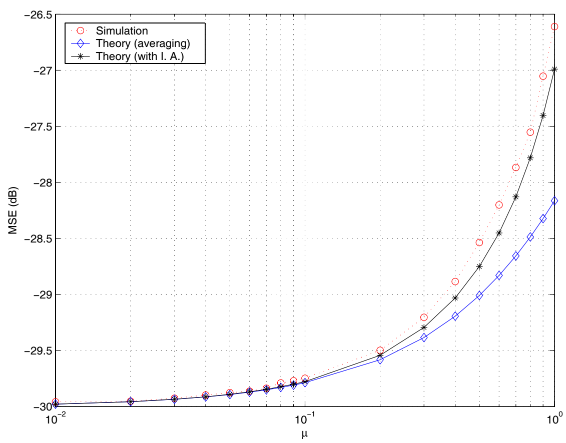

Fig. 2. Experimental and theoretical MSE versus μ for NLMS (Input: Gaussian, System: FIR (10), SNR=30dB)

this way, rather than propagate the weighting matrices themselves, it is more convenient to rewrite the recursion for $\bar{\Sigma}'$ in terms of its diagonal entries. For this purpose, we let the vectors $\bar{\sigma} = \text{diag}\{\bar{\Sigma}\}$ and $\lambda = \text{diag}\{\Lambda\}$ denote $M \times 1$ columns with the diagonal entries of the corresponding matrices. Actually, we shall use the notation $\text{diag}\{\cdot\}$ in two directions, both of which will be obvious from the context. Writing $\text{diag}\{\bar{\sigma}\}$ for a column vector $\bar{\sigma}$, results in a diagonal matrix whose diagonal entries are obtained from $\bar{\sigma}$. Therefore we also write $\bar{\Sigma} = \text{diag}\{\bar{\sigma}\}$ and $\Lambda = \text{diag}\{\lambda\}$. In terms of the vectors $\{\bar{\sigma}, \bar{\sigma}'\}$, the matrix relation (20) for $\bar{\Sigma}'$ can be replaced by the vector relation:

$$\bar{\sigma}' = \bar{F}\bar{\sigma} \tag{32}$$

in terms of the $M \times M$ coefficient matrix $\bar{F} = (I + \mu\nabla_{\tilde{\mathbf{w}}}\bar{\mathbf{f}}_{av}(0))^2$.

We can rewrite the recursion for $E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{\Sigma}}$ in (19) by using the vectors $\{\bar{\sigma}, \bar{\sigma}'\}$ instead of the matrices $\{\bar{\Sigma}, \bar{\Sigma}'\}$, say as

$$E\left[\|\bar{\mathbf{w}}_i^{pav}\|^2_{\text{diag}\{\bar{\sigma}\}}\right] = E\left[\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{\text{diag}\{\bar{\sigma}'\}}\right] + \mu^2 E|\mathbf{f}_\Delta(i,0)|^2\lambda^T\bar{\sigma}$$

where, for the last term, we used the fact that $E\|\bar{\mathbf{u}}_i\|^2_{\bar{\Sigma}} = \text{Tr}(E[\bar{\mathbf{u}}_i^*\bar{\mathbf{u}}_i]\bar{\Sigma}) = \lambda^T\bar{\sigma}$. For compactness of notation, we drop the $\text{diag}\{\cdot\}$ notation from the subscripts and keep the vectors, so that the above is simply rewritten as

$$E\left[\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{\sigma}}\right] = E\left[\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{\bar{\sigma}'}\right] + \mu^2 E|\mathbf{f}_\Delta(i,0)|^2\lambda^T\bar{\sigma} \tag{33}$$

Recursion (33) shows that in order to evaluate $E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{\sigma}}$ we need to know $E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{\bar{F}\bar{\sigma}}$, with a weighting matrix whose diagonal entries are $\bar{F}\bar{\sigma}$. Now the quantity $E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{F}\bar{\sigma}}$ can be inferred from (33) by writing the recursion for $\bar{F}\bar{\sigma}$, i.e.,

$$E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{F}\bar{\sigma}} = E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{\bar{F}^2\bar{\sigma}} + \mu^2 E|\mathbf{f}_\Delta(i,0)|^2(\lambda^T\bar{F}\bar{\sigma})$$

We again find that in order to evaluate $E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{F}\bar{\sigma}}$ we need to know $E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{\bar{F}^2\bar{\sigma}}$. The natural question is whether this procedure terminates. Fortunately, the procedure terminates. This is because once we write (33) by substituting $\bar{\sigma}$ by $\bar{F}^{M-1}\bar{\sigma}$ we get

$$E\|\mathbf{w}_i^{pav}\|^2_{\bar{F}^{M-1}\bar{\sigma}} = E\|\mathbf{w}_{i-1}^{pav}\|^2_{\bar{F}^M\bar{\sigma}} + \mu^2 E|\mathbf{f}_\Delta(i,0)|^2(\lambda^T\bar{F}^{M-1}\bar{\sigma})$$

where the weighting matrix on the RHS is $\bar{F}^M\bar{\sigma}$. This term can be deduced from the prior weighting factors. Indeed, let $p(x)$ denote the characteristic polynomial of $\bar{F}$, i.e., $p(x) = \det(xI - \bar{F})$. It is a polynomial of order M in x,

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

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

$$E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{F}^{M-1}\bar{\sigma}} = -\sum_{m=0}^{M-1} p_m E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\bar{F}^m\bar{\sigma}} \tag{34}$$

Theorem 2 [**Transient performance**] *Under assumption* A.1, *the transient performance of the partially averaged system (8) for sufficiently small step-sizes can be approximated by the $M-$dimensional state recursion*

$$\boxed{\mathcal{W}_i = \mathcal{F}\mathcal{W}_{i-1} + \mu^2 E|\mathbf{f}_\Delta(i,0)|^2\mathcal{Y}} \tag{35}$$

where

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

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

□

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

A. Learning Curve

The learning curve of the partially averaged system (8) describes the time evolution of $E|\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}^{pav}|^2$. Now, for very small step-sizes, we use the approximation

$$E|\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}|^2 \approx E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{R_u} \approx E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{\Lambda}$$

to evaluate the learning curve by computing $E\|\bar{\mathbf{w}}_{i-1}^{pav}\|^2_{\Lambda}$ for each i. This task can be accomplished recursively from relation (33) by iterating it and choosing $\bar{\sigma} = \lambda = \text{vec}(\Lambda)$. This yields

$$\begin{aligned} E\|\bar{\mathbf{w}}_i^{pav}\|^2_{\lambda} \quad = \quad & E\|\bar{\mathbf{w}}_{-1}^{pav}\|^2_{\bar{F}^i\lambda} + \\ & \mu^2 E|\mathbf{f}_\Delta(i,0)|^2(\lambda^T(I + \cdots + \bar{F}^{i-1})\lambda) \end{aligned} \tag{36}$$

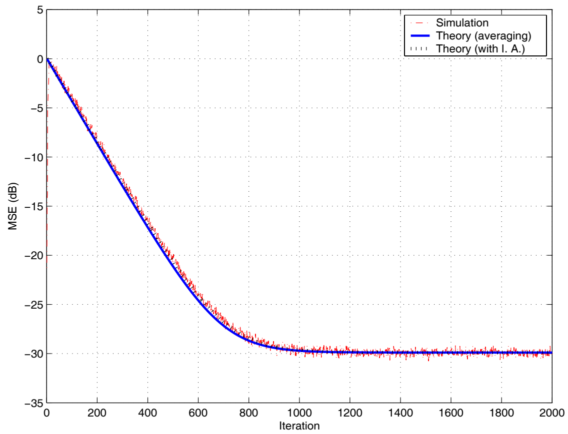

Fig. 3. Learning curves of LMS ($\mu = 0.005$, Input: Gaussian, System: FIR (10), SNR=30dB)

That is,

$$E\|\bar{\mathbf{w}}_{i-1}^{pav}\|_{\lambda}^2 = E\|\bar{\mathbf{w}}_{-1}^{pav}\|_{\mathbf{f}_{i-1}}^2 + \mu^2 E|\mathbf{f}_{\Delta}(i,0)|^2 g(i-1) \tag{37}$$

where the vector $\mathbf{f}_i$ and the scalar $g(i-1)$ satisfy the recursions

$$\mathbf{f}_{i-1} = \bar{F}\mathbf{f}_{i-2}, \qquad \mathbf{f}_0 = \lambda$$

$$g(i-1) = g(i-2) + \lambda^T \mathbf{f}_{i-1}, \qquad g(-1) = 0$$

Fig. 3 shows the learning curves of LMS obtained by using the learning curve of its partially averaged system and by simulation. The step-size is set to 0.005 for LMS. It is seen that there is good match between theory and practice.

B. Steady-State Behavior

In the above we used the variance relation (33) to characterize the transient behavior of the partially averaged system (8) in terms of a state recursion. We can use the same variance relation to shed further light on the mean-square performance of adaptive filters. In particular, we shall re-examine the EMSE, as well as study the mean-square deviation (MSD), which is defined by MSD $= E\|\tilde{\mathbf{w}}_i\|^2$ as $i \to \infty$. Assuming the step-size μ is chosen to guarantee filter stability, recursion (33) becomes in steady-state

$$E\|\bar{\mathbf{w}}_{\infty}^{pav}\|_{\bar{\sigma}}^2 = E\|\bar{\mathbf{w}}_{\infty}^{pav}\|_{\bar{F}\bar{\sigma}}^2 + \mu^2 E|\mathbf{f}_{\Delta}(i,0)|^2(\lambda^T\sigma) \tag{38}$$

which is equivalent to

$$E\|\bar{\mathbf{w}}_{\infty}^{pav}\|_{(I-\bar{F})\bar{\sigma}}^2 = \mu^2 E|\mathbf{f}_{\Delta}(i,0)|^2(\lambda^T\bar{\sigma}) \tag{39}$$

Assume that we select $\bar{\sigma}$ as the solution to the linear system of equations $(I-\bar{F})\bar{\sigma} = \text{diag}\{I\}$. In this case, the weighting quantity that appears in (39) reduces to the vector of unit entries. Then the left-hand side of (39) becomes the filter MSD and (39) leads to

$$\boxed{\text{MSD} = \mu^2 E|\mathbf{f}_{\Delta}(i,0)|^2 \lambda^T (I-\bar{F})^{-1}\text{diag}\{I\}} \tag{40}$$

In a similar way, let us evaluate the EMSE. Now we need to evaluate $E\|\bar{\mathbf{w}}_{\infty}^{pav}\|_{\lambda}^2$, where the weighting factor is $\lambda = \text{diag}\{\Lambda\}$. Assume we select $\bar{\sigma}$ as the solution to the linear system of equations $(I-\bar{F})\bar{\sigma} = \lambda$. In this case, the weighting quantity that appears in (39) reduces to Λ. Then the LHS of (39) becomes the filter EMSE and (39) leads to the desired result

$$\boxed{\text{EMSE} = \mu^2 E|\mathbf{f}_{\Delta}(i,0)|^2 \lambda^T (I-\bar{F})^{-1}\lambda} \tag{41}$$

V. Conclusion

In this paper, we considered general weight-error adaptive updates of the form (2) and studied their mean-square error and mean-square deviation performance measures as well as the transient behavior of their partially averaged systems. The derivation was based on replacing the filter error equation by a partially-averaged system, whose weight-error trajectory stays close to that of the adaptive filter for small step-sizes. By applying energy conservation arguments to the averaged system, we were able to derive performance results as well as characterize the filter transient behavior. The results obtained using averaging are essentially similar to those using independence analysis.

References

[1] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory*, MIT Press, Cambridge, MA, 1984.

[2] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[3] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, 1987.

[4] L. Ljung "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. 22, pp. 551–575, 1977.

[5] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol 58, pp. 963–993, 1979.

[6] S. K. Jones, R. K. Cavin, and W. M. Reed, "Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes," *IEEE Trans. Inform. Theory*, vol. 28, pp. 318–329, 1982.

[7] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, NY, 2003.

[8] 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, no. 2, pp. 314–324, Feb. 2001.

[9] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 639–652, March 2003.