Transient Analysis of Data-Normalized Adaptive Filters

Tareq Y. Al-Naffouri and Ali H. Sayed, *Fellow, IEEE*

***Abstract*—This paper develops an approach to the transient analysis of adaptive filters with data normalization. Among other results, the derivation characterizes the transient behavior of such filters in terms of a linear time-invariant state-space model. The stability of the model then translates into the mean-square stability of the adaptive filters. Likewise, the steady-state operation of the model provides information about the mean-square deviation and mean-square error performance of the filters. In addition to deriving earlier results in a unified manner, the approach leads to stability and performance results without restricting the regression data to being Gaussian or white. The framework is based on energy-conservation arguments and does not require an explicit recursion for the covariance matrix of the weight-error vector.**

***Index Terms*—Adaptive filter, data nonlinearity, energy-conservation, feedback analysis, mean-square-error, stability, steady-state analysis, transient analysis.**

I. Introduction

ADAPTIVE filters are, by design, time-variant and nonlinear systems that adapt to variations in signal statistics and that learn from their interactions with the environment. The success of their learning mechanism can be measured in terms of how fast they adapt to changes in the signal characteristics and how well they can learn given sufficient time (e.g., [1]–[3]). It is therefore typical to measure the performance of an adaptive filter in terms of both its transient performance and its steady-state performance. The former is concerned with the stability and convergence rate of an adaptive scheme, whereas the latter is concerned with the mean-square error that is left in steady state.

There have been extensive works in the literature on the performance of adaptive filters with many ingenious results and approaches (e.g., [1]–[11]). However, it is generally observed that most works study individual algorithms separately. This is because different adaptive schemes have different nonlinear update equations, and the particularities of each case tend to require different arguments and assumptions.

In recent works [12]–[15], a unified energy-based approach to the steady-state and tracking performance of adaptive filters has been developed that makes it possible not only to treat algorithms uniformly but also to arrive at new performance results. This approach is based on studying the energy flow through each iteration of an adaptive filter, and it relies on an exact energy conservation relation that holds for a large class of adaptive filters. This relation has been originally developed in [16]–[19] in the context of robustness analysis of adaptive filters within a deterministic framework. It has since then been used in [12]–[15] as a convenient tool for studying the steady-state performance of adaptive filters within a stochastic framework as well. In this paper, we show how to extend the energy-based approach to the *transient* analysis (as opposed to the *steady-state* analysis) of adaptive filters. Such an extension is desirable since it would allow us, just as in the steady-state case, to bring forth similar benefits such as the convenience of a unified treatment, the derivation of stability and convergence results, and the weakening of some assumptions.

In a companion article [20], we similarly extend the energy-conservation approach to study the transient behavior of adaptive filters with error nonlinearities.

A. Contributions of the Work

The main contributions of the paper are as follows.

a) In the next section, we introduce weighted estimation errors as well as weighted energy norms and relate these quantities through a fundamental energy relation. The main results of this section are summarized in Theorem 1.
b) In Sections III and IV, we illustrate the mechanism of our approach for transient analysis by applying it to the LMS algorithm and its normalized version for Gaussian regressors.
c) In Section V, we study the general case of adaptive algorithms with data nonlinearities without imposing restrictions on the color of the regression data (i.e., without requiring the regression data to be Gaussian or white). The analysis leads to stability results and closed-form expressions for the MSE and MSD. The main results are summarized in Theorem 2.
d) In Section VI, we extend our study to include adaptive filters that employ matrix data nonlinearities. We again derive stability results and closed-form expressions for the MSE and MSD. The main results are summarized in Theorem 3.

The statements of Theorems 1–3 constitute the contributions of this work.

B. Notation

We focus on real-valued data, although the extension to complex-valued data is immediate. Small boldface letters are

Manuscript received February 28, 2001; revised October 23, 2002. This work was supported in part by the National Science Foundation under Grants CCR -9732376, ECS-9820765, and CCR -0208573. The work of T. Y. Al-Naffouri was also partially supported by a fellowship from King Fahd University of Petroleum and Minerals, Saudi Arabia. The associate editor coordinating the review of this paper and approving it for publication was Dr. Xiang-Gen Xia.

T. Y. Al-Naffouri is with the Electrical Engineering Department, Stanford University, Stanford, CA 94305 USA (e-mail: naffouri@stanford.edu).

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/TSP.2002.808106

TABLE I
EXAMPLES OF DATA NONLINEARTIES $g[\cdot]$ OR $\mathbf{H}[\cdot]$

ALGORITHM	$g[\cdot]$ OR $\mathbf{H}[\cdot]$
LMS	1
NLMS	$\\|\mathbf{u}_i\\|^2$
ϵ-NLMS	$\epsilon + \\|\mathbf{u}_i\\|^2$
NLMS family	$\frac{1}{\\|\mathbf{u}_i\\|_q}\operatorname{diag}\left(\|u_{i_1}\|^{q-1}\operatorname{sgn}(u_{i_1}), \|u_{i_2}\|^{q-1}\operatorname{sgn}(u_{i_2}), \ldots, \|u_{i_M}\|^{q-1}\operatorname{sgn}(u_{i_M})\right)$
Power normalized LMS	$\operatorname{diag}(p_1(i), p_2(i), \ldots, p_M(i))$ $p_k(i+1) = \beta p_k(i) + (1-\beta)\|u_{i_k}\|^2, \quad 0 \ll \beta < 1$
Sign regressor	$\operatorname{diag}\left(\frac{\operatorname{sgn}(u_{i_1})}{u_{i_1}}, \frac{\operatorname{sgn}(u_{i_2})}{u_{i_2}}, \ldots, \frac{\operatorname{sgn}(u_{i_M})}{u_{i_M}}\right)$
Multiple step-sizes	$\operatorname{diag}(\mu_1, \mu_2, \ldots, \mu_M)$

used to denote vectors, e.g., $\boldsymbol{w}$, and the symbol T denotes transposition. The notation $\|\boldsymbol{w}\|^2$ denotes the squared Euclidean norm of a vector $\|\boldsymbol{w}\|^2 = \boldsymbol{w}^T\boldsymbol{w}$, whereas $\|\boldsymbol{w}\|^2_{\boldsymbol{\Sigma}}$ denotes the weighted squared Euclidean norm $\|\boldsymbol{w}\|^2_{\boldsymbol{\Sigma}} = \boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w}$. All vectors are column vectors except for a single vector, namely, the input data vector denoted by $\mathbf{u}_i$, which is taken to be a row vector. The time instant is placed as a subscript for vectors and between parentheses for scalars, e.g., $\boldsymbol{w}_i$ and $e(i)$.

C. Adaptive Filters With Data Nonlinearities

Consider noisy measurements $\{d(i)\}$ that arise from the model

$$d(i) = \mathbf{u}_i\boldsymbol{w}^o + v(i)$$

for some $M \times 1$ unknown vector $\boldsymbol{w}^o$ that we wish to estimate, and where $v(i)$ accounts for measurement noise and modeling errors, and $\mathbf{u}_i$ denotes a *row* regression vector. Both $\mathbf{u}_i$ and $v(i)$ are stochastic in nature. Many adaptive schemes have been developed in the literature for the estimation of $\boldsymbol{w}^o$ in different contexts. Most of these algorithms fit into the general description

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu f[e(i), \mathbf{u}_i]\mathbf{u}_i^T, \qquad i \geq 0 \tag{1}$$

where $\boldsymbol{w}_i$ is an estimate for $\boldsymbol{w}^o$ at iteration i, μ is the step-size

$$e(i) = d(i) - \mathbf{u}_i\boldsymbol{w}_i = \mathbf{u}_i\boldsymbol{w}^o - \mathbf{u}_i\boldsymbol{w}_i + v(i) \tag{2}$$

is the estimation error, and $f[e(i), \mathbf{u}_i]$ denotes a generic function of $e(i)$ and the regression vector $\mathbf{u}_i$.

In terms of the weight-error vector $\tilde{\boldsymbol{w}}_i = \boldsymbol{w}^o - \boldsymbol{w}_i$, the adaptive filter (1) and (2) can be equivalently rewritten as

$$\tilde{\boldsymbol{w}}_{i+1} = \tilde{\boldsymbol{w}}_i - \mu f[e(i), \mathbf{u}_i]\mathbf{u}_i^T \tag{3}$$

and

$$e(i) = \mathbf{u}_i\tilde{\boldsymbol{w}}_i + v(i). \tag{4}$$

We restrict our attention in this paper to nonlinearities $f[\cdot]$ that can be expressed in the *separable* form

$$f[e(i), \mathbf{u}_i] = \frac{e(i)}{g[\mathbf{u}_i]} \tag{5}$$

for some positive scalar-valued function $g[\mathbf{u}_i]$. In the latter part of this paper (see Section VI), matrix nonlinearities $\mathbf{H}[\mathbf{u}_i]$ will also be considered, i.e., functions $f[\cdot]$ of the form

$$f[e(i), \mathbf{u}_i] = \mathbf{H}[\mathbf{u}_i]e(i).$$

Table I lists some examples of data nonlinearities $\{g[\cdot], \mathbf{H}[\cdot]\}$ that appear in the literature. In the table, the notation $\{u_{i_1}, u_{i_2}, \ldots, u_{i_M}\}$ refers to the entries of the regressor vector $\mathbf{u}_i$.

II. WEIGHTED ENERGY RELATION

The adaptive filter analysis in future sections is based on an energy-conservation relation that relates the energies of several error quantities. To derive this relation, we first define some useful weighted errors. Thus, let $\boldsymbol{\Sigma}$ denote any symmetric positive definite $M \times M$ weighting matrix and define the weighted *a priori* and *a posteriori* error signals

$$e_a^{\Sigma}(i) \triangleq \mathbf{u}_i\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}_i, \qquad e_p^{\Sigma}(i) \triangleq \mathbf{u}_i\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}_{i+1}. \tag{6}$$

For $\boldsymbol{\Sigma} = \boldsymbol{I}$, we use the more standard notation

$$e_a(i) \triangleq e_a^I(i) = \mathbf{u}_i\tilde{\boldsymbol{w}}_i, \quad e_p(i) \triangleq e_p^I(i) = \mathbf{u}_i\tilde{\boldsymbol{w}}_{i+1}.$$

The freedom in selecting $\boldsymbol{\Sigma}$ will enable us to perform different kinds of analyses. For now, $\boldsymbol{\Sigma}$ will simply denote an arbitrary weighting matrix.

A. Energy-Conservation Relation

The energy relation that we seek is one that relates the energies of the following error quantities:

$$\{\ \tilde{\boldsymbol{w}}_i, \tilde{\boldsymbol{w}}_{i+1}, e_a^{\Sigma}(i), e_p^{\Sigma}(i)\ \}. \tag{7}$$

To arrive at the desired relation, we premultiply both sides of the adaptation equation (3) by $\mathbf{u}_i\boldsymbol{\Sigma}$ and incorporate the definitions (6). This results in an equality that relates the estimation errors $e_a^{\Sigma}(i)$, $e_p^{\Sigma}(i)$, and $e(i)$, namely

$$e_p^{\Sigma}(i) = e_a^{\Sigma}(i) - \frac{\mu}{\bar{\mu}_{\Sigma}(i)} f[e(i), \mathbf{u}_i] \tag{8}$$

where we introduced, for compactness of notation, the scalar quantity

$$\overline{\mu}_{\Sigma}(i) \triangleq \begin{cases} \frac{1}{\mathbf{u}_i \mathbf{\Sigma} \mathbf{u}_i^T}, & \text{if } \mathbf{u}_i \mathbf{\Sigma} \mathbf{u}_i^T \neq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Using (8), the nonlinearity $f[e(i), \mathbf{u}_i]$ can be eliminated from (3), yielding the following relation between the errors in (7):

$$\tilde{\boldsymbol{w}}_{i+1} = \tilde{\boldsymbol{w}}_i - \overline{\mu}_{\Sigma}(i)\mathbf{u}_i^T[e_a^{\Sigma}(i) - e_p^{\Sigma}(i)].$$

From this equation, it follows that the weighted energies of these errors are related by

$$\tilde{\boldsymbol{w}}_{i+1}^T \mathbf{\Sigma} \tilde{\boldsymbol{w}}_{i+1} = \left(\tilde{\boldsymbol{w}}_i - \overline{\mu}_{\Sigma}(i)\mathbf{u}_i^T[e_a^{\Sigma}(i) - e_p^{\Sigma}(i)]\right)^T \cdot \mathbf{\Sigma}\left(\tilde{\boldsymbol{w}}_i - \overline{\mu}_{\Sigma}(i)\mathbf{u}_i^T[e_a^{\Sigma}(i) - e_p^{\Sigma}(i)]\right)$$

or, more compactly, after expanding and grouping terms, by the following *energy-conservation* identity

$$\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2 + \overline{\mu}_{\Sigma}(i)\left|e_a^{\Sigma}(i)\right|^2 = \|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 + \overline{\mu}_{\Sigma}(i)\left|e_p^{\Sigma}(i)\right|^2. \tag{10}$$

This result is *exact* for any adaptive algorithm described by (3), i.e., for any nonlinearity $f[\cdot,\cdot]$, and it has been derived without any approximations, and no restrictions have been imposed on the symmetric weighting matrix $\mathbf{\Sigma}$.

The result (10) with $\mathbf{\Sigma} = \boldsymbol{I}$ was developed in [16]–[18] in the context of robustness analysis of adaptive filters, and it was later used in [12]–[15] in the context of steady-state and tracking analysis. The incorporation of a weighting matrix $\mathbf{\Sigma}$ allows us to perform transient analyzes as well, as we will discuss in future sections.

B. Algebra of Weighted Norms

Before proceeding, it is convenient for the subsequent discussion to list some algebraic properties of weighted norms. Therefore, let a_1 and a_2 be scalars, and let $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ be symmetric matrices of size M. Then, the following properties hold.

1) **Superposition.**

$$a_1\|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1}^2 + a_2\|\tilde{\boldsymbol{w}}_i\|_{\Sigma_2}^2 = \|\tilde{\boldsymbol{w}}_i\|_{a_1\Sigma_1 + a_2\Sigma_2}^2. \tag{11}$$

2) **Polarization.**

$$\begin{aligned}(\mathbf{u}_i\mathbf{\Sigma}_1\tilde{\boldsymbol{w}}_i)(\mathbf{u}_i\mathbf{\Sigma}_2\tilde{\boldsymbol{w}}_i) &= \|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1 u_i^T u_i \Sigma_2}^2 \\ &= \|\tilde{\boldsymbol{w}}_i\|_{\Sigma_2 u_i^T u_i \Sigma_1}^2.\end{aligned} \tag{12}$$

3) **Independence.** If $\mathbf{u}_i$ and $\tilde{\boldsymbol{w}}_i$ are independent random vectors, then the polarization property gives

$$\begin{aligned}E\left[(\mathbf{u}_i\mathbf{\Sigma}_1\tilde{\boldsymbol{w}}_i)(\mathbf{u}_i\mathbf{\Sigma}_2\tilde{\boldsymbol{w}}_i)\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1 u_i^T u_i \Sigma_2}^2\right] \\ &= E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1 E[u_i^T u_i]\Sigma_2}^2\right]\end{aligned}$$

where the last equality is true when $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are constant matrices.

4) **Linear transformation.** For any $N \times M$ matrix $\mathbf{A}$

$$\|\mathbf{A}\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 = \|\tilde{\boldsymbol{w}}_i\|_{A^T\Sigma A}^2.$$

5) **Orthogonal transformation.** If $\boldsymbol{Q}$ is orthogonal, it is easy to see that

$$\|\boldsymbol{Q}^T\tilde{\boldsymbol{w}}_i\|^2 = \|\tilde{\boldsymbol{w}}_i\|^2. \tag{13}$$

6) **Blindness to asymmetry.** The weighted sum of squares is blind to any asymmetry in the weight $\mathbf{A}$, i.e.,

$$\|\tilde{\boldsymbol{w}}_i\|_A^2 = \|\tilde{\boldsymbol{w}}_i\|_{A^T}^2 = \|\tilde{\boldsymbol{w}}_i\|_{A/2 + A^T/2}^2. \tag{14}$$

7) **Notational convention.** We will often write

$$\|\tilde{\boldsymbol{w}}_i\|_{\text{vec}(\Sigma_1)}^2 \triangleq \|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1}^2$$

where $\text{vec}(\Sigma_1)$ is obtained by stacking all the columns of $\mathbf{\Sigma}_1$ into a vector. For the special case when $\mathbf{\Sigma}_1$ is diagonal, it suffices to collect the diagonal entries of $\mathbf{\Sigma}_1$ into a vector, and we thus write

$$\|\tilde{\boldsymbol{w}}_i\|_{\text{diag}(\Sigma_1)}^2 \triangleq \|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1}^2.$$

C. Data-Normalized Filters

We now examine the simplifications that occur when $f[\cdot,\cdot]$ is restricted to the form (5). Upon replacing $e_p^{\Sigma}(i)$ in (10) by its equivalent expression (8) and expanding, we get

$$\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2 = \|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 - 2\mu\frac{e(i)e_a^{\Sigma}(i)}{g[\mathbf{u}_i]} + \frac{\mu^2}{\overline{\mu}_{\Sigma}(i)}\frac{e^2(i)}{g^2[\mathbf{u}_i]}. \tag{15}$$

To proceed, we replace $e(i)$, as defined in (4), by

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

Then, (15) becomes

$$\begin{aligned}\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2 = {}& \|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 - 2\mu\frac{e_a(i)e_a^{\Sigma}(i)}{g[\mathbf{u}_i]} + \frac{\mu^2}{\overline{\mu}_{\Sigma}(i)}\frac{e_a^2(i)}{g^2[\mathbf{u}_i]} \\ & -2\mu\left(\frac{e_a^{\Sigma}(i)}{g[\mathbf{u}_i]} - \frac{\mu}{\overline{\mu}_{\Sigma}(i)}\frac{e_a(i)}{g^2[\mathbf{u}_i]}\right)v(i) + \frac{\mu^2}{\overline{\mu}_{\Sigma}(i)}\frac{v^2(i)}{g^2[\mathbf{u}_i]}.\end{aligned} \tag{16}$$

Now, note that $e_a^{\Sigma}(i)e_a(i)$ and $e_a^2(i)$ can be expressed as some weighted norms of $\tilde{\boldsymbol{w}}_i$. Indeed, from (12), we have

$$e_a(i)e_a^{\Sigma}(i) = (\mathbf{u}_i\tilde{\boldsymbol{w}}_i)(\mathbf{u}_i\mathbf{\Sigma}\tilde{\boldsymbol{w}}_i) = \|\tilde{\boldsymbol{w}}_i\|_{u_i^T u_i \Sigma}^2 \tag{17}$$

and, subsequently

$$e_a^2(i) = e_a(i)e_a^I(i) = \|\tilde{\boldsymbol{w}}_i\|_{u_i^T u_i}^2. \tag{18}$$

Upon substituting (17) and (18) into (16), we get

$$\begin{aligned}\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2 = {}& \|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 - 2\mu\frac{1}{g[\mathbf{u}_i]}\|\tilde{\boldsymbol{w}}_i\|_{u_i^T u_i \Sigma}^2 + \mu^2\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]}\|\tilde{\boldsymbol{w}}_i\|_{u_i^T u_i}^2 \\ & -2\mu\left(\frac{e_a^{\Sigma}(i)}{g[\mathbf{u}_i]} - \frac{\mu}{\overline{\mu}_{\Sigma}(i)}\frac{e_a(i)}{g^2[\mathbf{u}_i]}\right)v(i) + \mu^2 v^2(i)\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]}.\end{aligned}$$

This relation can be written more compactly by using the superposition property (11) to group the various weighted norms of

$\tilde{\boldsymbol{w}}_i$ into one term, namely

$$\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma} = \|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma'} - 2\mu\left(\frac{e_a^{\Sigma}(i)}{g[\mathbf{u}_i]} - \frac{\mu}{\overline{\mu}_{\Sigma}(i)}\frac{e_a(i)}{g^2[\mathbf{u}_i]}\right) \cdot v(i) + \mu^2 v^2(i)\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2[\mathbf{u}_i]} \quad (19)$$

where

$$\boldsymbol{\Sigma}' \triangleq \boldsymbol{\Sigma} - 2\mu\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\boldsymbol{\Sigma} + \mu^2\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2[\mathbf{u}_i]}\mathbf{u}_i^T\mathbf{u}_i. \quad (20)$$

The only role that $\boldsymbol{\Sigma}'$ plays is a weight in the quadratic form $\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma'}$. Hence, and in view of (14), we can replace the defining expression (20) for $\boldsymbol{\Sigma}'$ by its symmetric part

$$\boldsymbol{\Sigma}' \triangleq \boldsymbol{\Sigma} - \mu\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\boldsymbol{\Sigma} - \mu\boldsymbol{\Sigma}\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]} + \mu^2\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2[\mathbf{u}_i]}\mathbf{u}_i^T\mathbf{u}_i. \quad (21)$$

Finally, it is straightforward to conclude from the weight-error recursion

$$\tilde{\boldsymbol{w}}_{i+1} = \tilde{\boldsymbol{w}}_i - \mu\frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]}[e_a(i) + v(i)]$$

and from $e_a(i) = \mathbf{u}_i\tilde{\boldsymbol{w}}_i$ that

$$\tilde{\boldsymbol{w}}_{i+1} = \left(\boldsymbol{I} - \mu\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right)\tilde{\boldsymbol{w}}_i - \mu\frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]}v(i). \quad (22)$$

D. Weighted Variance Relation

A few comments are in place.

1) First, the pair (19) and (21) is equivalent to the energy relation (10) and, hence, is exact.
2) This pair represents the starting point for various types of analyzes of adaptive filters with data normalization.
3) As it stands, the energy relation (19)–(21) cannot be propagated in time since it requires a recursion describing the evolution of $e_a(i)$. However, this complication can be removed by introducing the following reasonable assumption on the noise sequence:

   AN. The noise sequence $v(i)$ is zero-mean, iid and is independent of $\mathbf{u}_i$.

   This assumption renders the third term of (19) zero-mean, and (19) simplifies under expectation to

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma}\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma'}\right] + \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2(\mathbf{u}_i)}\right]. \quad (23)$$

   Likewise, (22) simplifies to

$$E\tilde{\boldsymbol{w}}_{i+1} = E\left[\left(\boldsymbol{I} - \mu\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right)\tilde{\boldsymbol{w}}_i\right]. \quad (24)$$

   While the iterated relation (23) is compact, it is still hard to propagate since $\boldsymbol{\Sigma}'$ is dependent on the data $\mathbf{u}_i$ so that the evaluation of the expectation $E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma'}\right]$ is not trivial in general.

d) For this reason, we shall contend ourselves with the independence assumption.

AI. The sequence of vectors $\mathbf{u}_i$ is iid.

This condition enables us to the split the expectation in (23) as

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma}\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{E[\Sigma']}\right] + \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2(\mathbf{u}_i)}\right]. \quad (25)$$

Observe that the weighting matrix for $\tilde{\boldsymbol{w}}_i$ is now given by the expectation $E[\boldsymbol{\Sigma}']$. As we will soon see, the above equality renders the issue of transient and stability analyses of an adaptive filter equivalent to a multivariate computation of certain moments.

In order to emphasize the fact that the weighting matrix changes from $\boldsymbol{\Sigma}$ to $E[\boldsymbol{\Sigma}']$ according to (21), we will attach a time index to the weighting matrices and use (21) and (25) to write more explicitly

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma_{i+1}}\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_i}\right] + \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{\Sigma_{i+1}}}{g^2[\mathbf{u}_i]}\right]$$

where we replaced $\boldsymbol{\Sigma}$ by $\boldsymbol{\Sigma}_{i+1}$ and $E[\boldsymbol{\Sigma}']$ by $\boldsymbol{\Sigma}_i$, which is now defined by

$$\boldsymbol{\Sigma}_i \triangleq \boldsymbol{\Sigma}_{i+1} - \mu E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right]\boldsymbol{\Sigma}_{i+1} - \mu\boldsymbol{\Sigma}_{i+1}E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right] + \mu^2 E\left[\frac{\|\mathbf{u}_i\|_{\Sigma_{i+1}}}{g^2[\mathbf{u}_i]}\mathbf{u}_i^T\mathbf{u}_i\right].$$

Note that this recursion runs backward in time, and its boundary condition will therefore be specified at ∞. Moreover, Σ_i can be verified to be positive definite.

Likewise, applying the independence assumption AI to the right-hand side of (24), we find that

$$E\tilde{\boldsymbol{w}}_{i+1} = E\left(\boldsymbol{I} - \mu\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right)\cdot E\tilde{\boldsymbol{w}}_i$$

with the expectation on the right-hand side of (24) split into the product of two expectations.

e) Inspection of recursions (19) and (23) reveals that the iid assumption (AN) on the noise sequence is critical. Indeed, while (23) can be propagated in time without the independence assumption AI, it is not possible to do the same for (19). Fortunately, assumption AN is, in general, reasonable.

We summarize in the following statement the variance and mean recursions that will form the basis of our transient analysis.

Theorem 1 (Weighted-Variance Relation): Consider an adaptive filter of the form

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu\frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]}e(i), \quad i \geq 0 \quad (26)$$

where $e(i) = d(i) - \mathbf{u}_i\boldsymbol{w}_i$, and $d(i) = \mathbf{u}_i\boldsymbol{w}^o + v(i)$. Assume that the sequences $\{v(i), \mathbf{u}_i\}$ are iid and mutually independent. For any given $\boldsymbol{\Sigma}_{i+1}$, it holds that

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma_{i+1}}\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_i}\right] + \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{\Sigma_{i+1}}}{g^2[\mathbf{u}_i]}\right] \quad (27)$$

where $\boldsymbol{\Sigma}_i$ is constructed from $\boldsymbol{\Sigma}_{i+1}$ via

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i+1} - \mu E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right]\boldsymbol{\Sigma}_{i+1} - \mu\boldsymbol{\Sigma}_{i+1}E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right] + \mu^2 E\left[\frac{\|\mathbf{u}_i\|_{\Sigma_{i+1}}}{g^2[\mathbf{u}_i]}\mathbf{u}_i^T\mathbf{u}_i\right]. \quad (28)$$

It also holds that the mean weight-error vector satisfies

$$E\tilde{\boldsymbol{w}}_{i+1} = E\left(\boldsymbol{I} - \mu\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right)\cdot E\tilde{\boldsymbol{w}}_i. \tag{29}$$

$\diamond$

The purpose of the sections that follow is to show how the above variance and mean recursions can be used to study the transient performance of adaptive schemes with data nonlinearities. In particular, we will show how the freedom in selecting the weighting matrix $\boldsymbol{\Sigma}_{i+1}$ can be used advantageously to derive several performance measures.

First, however, we shall illustrate the mechanism of our analysis by considering two special cases for which results are already available in the literature. More specifically, we will start with the transient analysis of LMS and normalized LMS algorithms for Gaussian regression data in Sections III and IV. Once the main ideas have been illustrated in this manner, we will then describe our general procedure in Section V, which applies to adaptive filters with more general data normalizations, as well as to regression data that are not restricted to being Gaussian or white.

E. Change of Variables

In the meantime, we remark that sometimes it is useful to employ a convenient change of coordinates, especially when dealing with Gaussian regressors. Thus, let $\boldsymbol{R} = E\mathbf{u}_i^T\mathbf{u}_i$ denote the covariance matrix of $\mathbf{u}_i$ and introduce its eigendecomposition

$$\boldsymbol{R} = \boldsymbol{Q}^T\boldsymbol{\Lambda}\boldsymbol{Q}$$

where $\boldsymbol{Q}$ is orthogonal, and $\boldsymbol{\Lambda}$ is a positive diagonal matrix with entries $\{\lambda_k\}$. Define further

$$\overline{\boldsymbol{w}}_i \triangleq \boldsymbol{Q}\tilde{\boldsymbol{w}}_i, \qquad \overline{\mathbf{u}}_i \triangleq \mathbf{u}_i\boldsymbol{Q}^T, \qquad \overline{\boldsymbol{\Sigma}}_i \triangleq \boldsymbol{Q}\boldsymbol{\Sigma}_i\boldsymbol{Q}^T. \tag{30}$$

In view of the orthogonal transformation property (13), we have

$$\|\tilde{\boldsymbol{w}}_i\|^2_{\boldsymbol{\Sigma}_i} = \|\overline{\boldsymbol{w}}_i\|^2_{\overline{\boldsymbol{\Sigma}}_i} \quad \text{and} \quad \|\mathbf{u}_i\|^2_{\boldsymbol{\Sigma}_i} = \|\overline{\mathbf{u}}_i\|^2_{\overline{\boldsymbol{\Sigma}}_i}.$$

Moreover, assuming that the nonlinearity $g[\cdot]$ is invariant under orthogonal transformations, i.e., $g[\mathbf{u}_i] = g[\overline{\mathbf{u}}_i]$ (e.g., $g[\mathbf{u}_i] = 1$ or $g[\mathbf{u}_i] = \|\mathbf{u}_i\|^2$), we find that the variance relation (27) retains the same form, namely

$$E\left[\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{\Sigma}_{i+1}}\right] = E\left[\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\Sigma}_i}\right] + \mu^2\sigma_v^2 E\left[\frac{\|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}_{i+1}}}{g^2[\overline{\mathbf{u}}_i]}\right]. \tag{31}$$

By premultiplying both sides of (28) by $\boldsymbol{Q}$ and post-multiplying by $\boldsymbol{Q}^T$, we similarly see that (28) also retains the same form

$$\overline{\boldsymbol{\Sigma}}_i = \overline{\boldsymbol{\Sigma}}_{i+1} - \mu E\left[\frac{\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i}{g[\overline{\mathbf{u}}_i]}\right]\overline{\boldsymbol{\Sigma}}_{i+1} - \mu\overline{\boldsymbol{\Sigma}}_{i+1}E\left[\frac{\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i}{g[\overline{\mathbf{u}}_i]}\right] + \mu^2 E\left[\frac{\|\overline{\mathbf{u}}_i\|_{\overline{\Sigma}_{i+1}}}{g^2[\overline{\mathbf{u}}_i]}\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i\right]. \tag{32}$$

Likewise, (29) becomes

$$E\overline{\boldsymbol{w}}_{i+1} = E\left(\boldsymbol{I} - \mu\frac{\overline{\mathbf{u}_i}^T\overline{\mathbf{u}_i}}{g[\overline{\mathbf{u}_i}]}\right)\cdot E\overline{\boldsymbol{w}}_i. \tag{33}$$

III. LMS With Gaussian Regressors

Consider the LMS algorithm for which $g[\mathbf{u}_i] = 1$ and assume the following.

AG. The regressors $\{\mathbf{u}_i\}$ arise from a Gaussian distribution with covariance matrix $\boldsymbol{R}$.

In this case, the data dependent moments that appear in (31)–(33) are given by

$$E\left[\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i\right] = \boldsymbol{\Lambda}, \quad E\left[\|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}_i}\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i\right] = 2\overline{\boldsymbol{\Sigma}}_i\boldsymbol{\Lambda}^2 + \text{Tr}\left(\overline{\boldsymbol{\Sigma}}_i\boldsymbol{\Lambda}\right)\boldsymbol{\Lambda}.$$

Therefore, for LMS, recursions (31) and (32) simplify to

$$E\left[\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{\Sigma}_{i+1}}\right] = E\left[\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\Sigma}_i}\right] + \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}_{i+1}}\right] \tag{34}$$

and

$$\overline{\boldsymbol{\Sigma}}_i = \overline{\boldsymbol{\Sigma}}_{i+1} - \mu\boldsymbol{\Lambda}\overline{\boldsymbol{\Sigma}}_{i+1} - \mu\overline{\boldsymbol{\Sigma}}_{i+1}\boldsymbol{\Lambda} + 2\mu^2\overline{\boldsymbol{\Sigma}}_{i+1}\boldsymbol{\Lambda}^2 + \mu^2\text{Tr}\left(\overline{\boldsymbol{\Sigma}}_{i+1}\boldsymbol{\Lambda}\right)\boldsymbol{\Lambda} \tag{35}$$

while (33) becomes

$$E\overline{\boldsymbol{w}}_{i+1} = E\left(\boldsymbol{I} - \mu\boldsymbol{\Lambda}\right)\cdot E\overline{\boldsymbol{w}}_i. \tag{36}$$

Now, observe that in recursion (35), $\overline{\boldsymbol{\Sigma}}_i$ will be diagonal if $\overline{\boldsymbol{\Sigma}}_{i+1}$ is. Therefore, in order for all successive $\overline{\boldsymbol{\Sigma}}_i$s to be diagonal it is sufficient to assume that the boundary condition for the recursion for $\boldsymbol{\Sigma}_i$ is taken as diagonal. In this way, the $\overline{\boldsymbol{\Sigma}}_i$s will be completely characterized by their diagonal entries. This prompts us to define the column vectors

$$\overline{\boldsymbol{\sigma}}_i \triangleq \text{diag}(\overline{\boldsymbol{\Sigma}}_i) \quad \text{and} \quad \boldsymbol{\lambda} \triangleq \text{diag}(\boldsymbol{\Lambda}).$$

In terms of these vectors, the matrix recursion (35) can be replaced by the more compact vector recursion

$$\overline{\boldsymbol{\sigma}}_i = \left(\boldsymbol{I} - 2\mu\boldsymbol{\Lambda} + 2\mu^2\boldsymbol{\Lambda}^2\right)\overline{\boldsymbol{\sigma}}_{i+1} + \mu^2\left(\boldsymbol{\lambda}^T\overline{\boldsymbol{\sigma}}_{i+1}\right)\boldsymbol{\lambda}$$

or

$$\overline{\boldsymbol{\sigma}}_i \triangleq \overline{\boldsymbol{F}}\overline{\boldsymbol{\sigma}}_{i+1} \tag{37}$$

where

$$\overline{\boldsymbol{F}} \triangleq \left(\boldsymbol{I} - 2\mu\boldsymbol{\Lambda} + 2\mu^2\boldsymbol{\Lambda}^2\right) + \mu^2\boldsymbol{\lambda}\boldsymbol{\lambda}^T.$$

The matrix $\overline{\boldsymbol{F}}$ describes the dynamics by which the weighting matrices $\overline{\boldsymbol{\Sigma}}_i$ evolve in time, and its eigenstructure turns out to be essential for filter stability. Using the fact that $\overline{\boldsymbol{\sigma}}_i = \overline{\boldsymbol{F}}\overline{\boldsymbol{\sigma}}_{i+1}$, we can rewrite (34) using a compact vector weighting notation

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{\sigma}_{i+1}} = E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}\overline{\sigma}_{i+1}} + \mu^2\sigma_v^2 E\|\overline{\mathbf{u}}_i\|^2_{\overline{\sigma}_{i+1}}. \tag{38}$$

Recursions (36)–(38) describe the transient behavior of LMS, and conclusions about mean-square stability and mean-square performance are now possible.

In transient analysis, we are interested in the time evolution of the expectations $\{E\tilde{\boldsymbol{w}}_i, E\|\tilde{\boldsymbol{w}}_i\|^2\}$ or, equivalently, $\{E\overline{\boldsymbol{w}}_i, E\|\overline{\boldsymbol{w}}_i\|^2\}$ since $\overline{\boldsymbol{w}}_i$ and $\tilde{\boldsymbol{w}}_i$ are related via the orthogonal matrix $\boldsymbol{Q}$. We start with the mean behavior.

A. Mean Behavior and Mean Stability

From (36) we find that the filter is convergent in the mean if, and only if, the step-size μ satisfies

$$\mu < \frac{2}{\lambda_{\max}} \tag{39}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{R}$.

B. Mean-Square Behavior

The evolution of $E\|\tilde{\boldsymbol{w}}_i\|^2 = E\|\overline{\boldsymbol{w}}_i\|^2$ can be deduced from the variance recursion (34) if $\overline{\boldsymbol{\Sigma}}_{i+1}$ is chosen as $\overline{\boldsymbol{\Sigma}}_{i+1} = \boldsymbol{I}$ (or, equivalently, $\boldsymbol{\Sigma}_{i+1} = \boldsymbol{I}$). This corresponds to choosing $\overline{\boldsymbol{\sigma}}_{i+1}$ in (38) as a column vector with unit entries, which is denoted by

$$\overline{\boldsymbol{\sigma}}_{i+1} = \mathbf{1} \stackrel{\Delta}{=} \mathrm{col}\{1, 1, \ldots, 1\}.$$

Now, we can see from (38) that

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2 = E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}\mathbf{1}} + \mu^2\sigma_v^2\left(\sum_{k=1}^{M}\lambda_k\right) \tag{40}$$

which shows that in order to evaluate $E\|\overline{\boldsymbol{w}}_{i+1}\|^2$, we need $E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}\mathbf{1}}$ with a weighting matrix equal to $\overline{\boldsymbol{F}}\mathbf{1}$. Now, $E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}\mathbf{1}}$ can be deduced from (38) by setting $\overline{\boldsymbol{\sigma}}_{i+1} = \overline{\boldsymbol{F}}\mathbf{1}$, i.e.,

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}\mathbf{1}} = E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}^2\mathbf{1}} + \mu^2\sigma_v^2\left(\boldsymbol{\lambda}^T\overline{F}\mathbf{1}\right). \tag{41}$$

Again, in order to evaluate $E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}\mathbf{1}}$, we need $E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}^2\mathbf{1}}$ with weighting $\overline{\boldsymbol{F}}^2\mathbf{1}$. This term can be deduced from (38) by choosing $\overline{\boldsymbol{\sigma}}_{i+1} = \overline{\boldsymbol{F}}^2\mathbf{1}$

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}^2\mathbf{1}} = E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}^3\mathbf{1}} + \mu^2\sigma_v^2\left(\boldsymbol{\lambda}^T\overline{F}^2\mathbf{1}\right) \tag{42}$$

and a new term with weighting matrix $\overline{\boldsymbol{F}}^3\mathbf{1}$ appears. Fortunately, this procedure terminates in view of the Cayley–Hamilton theorem. Thus, let $p(x) = \det(x\boldsymbol{I} - \overline{\boldsymbol{F}})$ denote the characteristic polynomial of $\overline{\boldsymbol{F}}$; it is an Mth-order polynomial in x

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

with coefficients $\{p_k, p_M = 1\}$. The Cayley–Hamilton theorem states that every matrix satisfies its characteristic equation, i.e., $p(\overline{\boldsymbol{F}}) = 0$, which allows us to conclude that

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}^M\mathbf{1}} = \sum_{k=0}^{M-1} -p_k E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}^k\mathbf{1}}. \tag{43}$$

We can now collect the above results into a single recursion by writing (40)–(43) as

$$\underbrace{\begin{bmatrix} E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\mathbf{1}} \\ E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}\mathbf{1}} \\ E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}^2\mathbf{1}} \\ \vdots \\ E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}^{(M-2)}\mathbf{1}} \\ E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\overline{F}^{(M-1)}\mathbf{1}} \end{bmatrix}}_{=\mathcal{W}_{i+1}} = \underbrace{\begin{bmatrix} 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ \vdots & & & & \\ 0 & 0 & 0 & & 1 \\ -p_0 & -p_1 & -p_2 & \ldots & -p_{M-1} \end{bmatrix}}_{=\mathcal{F}}$$

$$\times \begin{bmatrix} E\|\overline{\boldsymbol{w}}_i\|^2_{\mathbf{1}} \\ E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}\mathbf{1}} \\ E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}^2\mathbf{1}} \\ \vdots \\ E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}^{(M-2)}\mathbf{1}} \\ E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{F}^{(M-1)}\mathbf{1}} \end{bmatrix} + \mu^2\sigma_v^2 \underbrace{\begin{bmatrix} \boldsymbol{\lambda}^T\mathbf{1} \\ \boldsymbol{\lambda}^T\overline{\boldsymbol{F}}\mathbf{1} \\ \boldsymbol{\lambda}^T\overline{\boldsymbol{F}}^2\mathbf{1} \\ \vdots \\ \boldsymbol{\lambda}^T\overline{\boldsymbol{F}}^{M-1}\mathbf{1} \end{bmatrix}}_{\mathcal{Y}}.$$

If we define the vector and matrix quantities $\{\mathcal{W}_i, \mathcal{F}, \mathcal{Y}\}$ as indicated above, then the recursion can be rewritten more compactly as

$$\mathcal{W}_{i+1} = \mathcal{F}\mathcal{W}_i + \mu^2\sigma_v^2\mathcal{Y}. \tag{44}$$

We therefore find that the transient behavior of LMS is described by the M-dimensional state-space recursion (44) with coefficient matrix $\mathcal{F}$.[1] The evolution of the top entry of $\mathcal{W}_i$ corresponds to the mean-square deviation of the filter. Observe further that the eigenvalues of $\mathcal{F}$ coincide with those of $\overline{\boldsymbol{F}}$.

It is worth remarking that the same derivation that led to (44) with $\mathcal{W}_i$ defined in terms of the unity vector $\mathbf{1}$ can be repeated for any other choice of $\overline{\boldsymbol{\sigma}}_{i+1}$, say $\overline{\boldsymbol{\sigma}}_{i+1} = \overline{\boldsymbol{\sigma}}$ for some $\overline{\boldsymbol{\sigma}}$, to conclude that the same recursion (44) still holds with $\mathbf{1}$ replaced by $\overline{\boldsymbol{\sigma}}$. For instance, if we choose $\overline{\boldsymbol{\sigma}} = \boldsymbol{\lambda}$, then the top entry of the resulting state vector $\mathcal{W}_i$ will correspond to the learning curve of the adaptive filter. In Section V-B we will use this remark to describe more fully the learning behavior of adaptive filters with data normalizations.

C. Mean-Square Stability

From the results in the above two sections, we conclude that the LMS filter will be stable in the mean and mean-square senses if, and only if, μ satisfies (39) and guarantees the stability of the matrix $\overline{\boldsymbol{F}}$ (i.e., all the eigenvalues of $\overline{\boldsymbol{F}}$ should lie inside the unit circle). Since $\overline{\boldsymbol{F}}$ is easily seen to be non-negative definite in this case, we only need to worry about guaranteeing that its eigenvalues be smaller than unity.

Let us write $\overline{\boldsymbol{F}}$ in the form

$$\overline{\boldsymbol{F}} = \boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\boldsymbol{B}$$

where the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are both positive-definite and given by

$$\boldsymbol{A} \stackrel{\Delta}{=} 2\boldsymbol{\Lambda}, \qquad \boldsymbol{B} \stackrel{\Delta}{=} 2\boldsymbol{\Lambda}^2 + \boldsymbol{\lambda}\boldsymbol{\lambda}^T. \tag{45}$$

It follows from the argument in Appendix A that the eigenvalues of $\overline{\boldsymbol{F}}$ will be upper bounded by one if, and only if, the parameter μ satisfies

$$0 < \mu < \frac{1}{\lambda_{\max}\left(\boldsymbol{A}^{-1}\boldsymbol{B}\right)} \tag{46}$$

in terms of the maximum eigenvalue of $\boldsymbol{A}^{-1}\boldsymbol{B}$ (all eigenvalues of $\boldsymbol{A}^{-1}\boldsymbol{B}$ are real and positive). The above upper bound on μ can also be interpreted as the smallest positive scalar η that makes $(\boldsymbol{I} - \eta\boldsymbol{A}^{-1}\boldsymbol{B})$ singular. Let us denote this value of η by η^o. Combining (46) with (39), we find that μ should satisfy

$$0 < \mu < \min\{\frac{2}{\lambda_{\max}(\boldsymbol{R})}, \eta^o\}.$$

We can be more specific about η^o and show that it is smaller than $1/\lambda_{\max}(\boldsymbol{R})$. Actually, we can characterize η^o in terms of the eigenvalues of $\boldsymbol{R}$ as follows. Using the definitions (45) for $\boldsymbol{A}$ and $\boldsymbol{B}$, it can be verified that for all $\eta \in (0, 1/\lambda_{\max})$

$$\det\left(\boldsymbol{I} - \eta\boldsymbol{A}^{-1}\boldsymbol{B}\right) = \left(1 - \frac{\boldsymbol{\lambda}^T}{2}\left[\eta^{-1}\boldsymbol{I} - \boldsymbol{\Lambda}\right]^{-1}\mathbf{1}\right) \cdot \det\left(\boldsymbol{I} - \eta\boldsymbol{\Lambda}\right).$$

The values of $\eta \in (0, 1/\lambda_{\max})$ that result in $\det\left(\boldsymbol{I} - \eta A^{-1}B\right) = 0$ should therefore satisfy

$$\frac{\boldsymbol{\lambda}^T}{2}\left(2\eta^{-1}\boldsymbol{I} - \boldsymbol{\Lambda}\right)^{-1}\mathbf{1} = 1$$

[1] To be more precise, the transient behavior of LMS is described by the combination of both (44) and recursion (36).

i.e.,

$$\frac{1}{2}\sum_{k=1}^{M}\frac{\lambda_k\eta}{1-\eta\lambda_k}=1.$$

This equality has a unique solution η^o inside the interval $(0, 1/\lambda_{\max})$. This is because the function

$$f(\eta)\stackrel{\Delta}{=}\frac{1}{2}\sum_{k=1}^{M}\frac{\lambda_k\eta}{1-\eta\lambda_k}=1$$

is monotonically increasing in the interval $(0, 1/\lambda_{\max})$. Moreover, it evaluates to 0 at $\eta = 0$ and becomes unbounded as $\eta \rightarrow 1/\lambda_{\max}$. We therefore conclude that LMS is stable in the mean- and mean-square senses for all step sizes μ satisfying

$$\frac{1}{2}\sum_{k=1}^{M}\left(\frac{\lambda_k\mu}{1-\mu\lambda_k}\right)<1.$$

D. Steady-State Performance

Once filter stability has been guaranteed, we can proceed to derive expressions for the steady-state value of the mean-square error (MSE) and the mean-square deviation (MSD). To this end, note that in steady state, we have that for any vector $\boldsymbol{\sigma}$

$$\lim_{i\to\infty}E\|\overline{\boldsymbol{w}}_{i+1}\|^2_{\sigma}=\lim_{i\to\infty}E\|\overline{\boldsymbol{w}}_i\|^2_{\sigma}.$$

Thus, in the limit, (38) leads to

$$\lim_{i\to\infty}E\|\overline{\boldsymbol{w}}_i\|^2_{(I-F)\overline{\sigma}_\infty}=\mu^2\sigma_v^2E\|\overline{\mathbf{u}}_i\|^2_{\overline{\sigma}_\infty}. \tag{47}$$

Here, $\overline{\boldsymbol{\Sigma}}_\infty = \text{diag}(\overline{\boldsymbol{\sigma}}_\infty)$ denotes the boundary condition of the recursion (32), which we are free to choose.

Now, in order to evaluate the MSE, we first recall that it is defined by

$$\text{MSE}=\lim_{i\to\infty}Ee_a^2(i)$$

which, in view of the independence assumption AI, is also given by

$$\text{MSE}=\lim_{i\to\infty}E\|\overline{\boldsymbol{w}}_i\|^2_{\lambda}.$$

This is because

$$Ee_a^2(i)=E\|\tilde{\boldsymbol{w}}_i\|_{u_i^Tu_i}=E\|\tilde{\boldsymbol{w}}_i\|_R=E\|\overline{\boldsymbol{w}}_i\|^2_{\lambda}.$$

Therefore, to obtain the MSE, we should choose $\overline{\boldsymbol{\sigma}}_\infty$ in (47) so that $\left(\boldsymbol{I}-\overline{\boldsymbol{F}}\right)\overline{\boldsymbol{\sigma}}_\infty=\boldsymbol{\lambda}$, in which case, we get

$$\text{MSE}=\mu^2\sigma_v^2E\left[\|\overline{\mathbf{u}}_i\|^2_{(I-\overline{F})^{-1}\lambda}\right]. \tag{48}$$

A more explicit expression for the MSE can be obtained by using the matrix inversion lemma to evaluate the matrix inverse that appears in (48). Doing so leads to the well-known result

$$\text{MSE}=\frac{\sigma_v^2\sum_{i=1}^{M}\frac{\mu\lambda_i}{2-2\mu\lambda_i}}{1-\sum_{i=1}^{M}\frac{\mu\lambda_i}{2-2\mu\lambda_i}}.$$

The MSD can be calculated along the same lines by noting that

$$\text{MSD}=\lim_{i\to\infty}E\|\tilde{\boldsymbol{w}}_i\|^2=\lim_{i\to\infty}E\|\overline{\boldsymbol{w}}_i\|^2_I=\lim_{i\to\infty}E\|\overline{\boldsymbol{w}}_i\|^2_{\mathbf{1}}.$$

The above means that in order to obtain an expression for the MSD, we should now choose $\overline{\boldsymbol{\sigma}}_\infty$ in (47) such that $\overline{\boldsymbol{\sigma}}_\infty=\left(\boldsymbol{I}-\overline{\boldsymbol{F}}\right)^{-1}\mathbf{1}$, which yields

$$\text{MSD}=\mu^2\sigma_v^2E\left[\|\overline{\mathbf{u}}_i\|^2_{(I-\overline{F})^{-1}1}\right].$$

Just like the expression for the MSE, we can use the matrix inversion lemma to get an explicit expression for $(\boldsymbol{I}-\boldsymbol{F})^{-1}\mathbf{1}$ and, subsequently, for the MSD

$$\text{MSD}=\frac{\sigma_v^2\sum_{i=1}^{M}\frac{\mu}{2-2\mu\lambda_i}}{1-\sum_{i=1}^{M}\frac{\mu\lambda_i}{2-2\mu\lambda_i}}.$$

Both of these steady-state expressions were derived in [5]. Here, we arrived at the expressions as a byproduct of a framework that can also handle a variety of data-normalized adaptive filters (see Section V). In addition, observe how the expressions for MSE and MSD can be obtained simply by conveniently choosing different values for the boundary condition $\overline{\boldsymbol{\sigma}}_\infty$.

IV. Normalized LMS with Gaussian Regressors

We now consider the normalized LMS algorithm, for which $g(\mathbf{u}_i)=\epsilon+\|\mathbf{u}_i\|^2$ with $\epsilon\geq 0$. For this choice of $g(\mathbf{u}_i)$, recursion (32) becomes

$$\overline{\boldsymbol{\Sigma}}_i=\overline{\boldsymbol{\Sigma}}_{i+1}-\mu E\left[\frac{\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i}{\epsilon+\|\overline{\mathbf{u}}_i\|^2}\right]\overline{\boldsymbol{\Sigma}}_{i+1}$$
$$-\mu\overline{\boldsymbol{\Sigma}}_{i+1}E\left[\frac{\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i}{\epsilon+\|\overline{\mathbf{u}}_i\|^2}\right]+\mu^2E\left[\frac{\|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}_{i+1}}}{(\epsilon+\|\overline{\mathbf{u}}_i\|^2)^2}\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i\right]. \tag{49}$$

Progress in the analysis is now pending on the evaluation of the moments

$$\mathbf{A}\stackrel{\Delta}{=}2E\left[\frac{\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i}{\epsilon+\|\overline{\mathbf{u}}_i\|^2}\right],\quad \mathbf{B}'\stackrel{\Delta}{=}E\left[\frac{\|\overline{\mathbf{u}}_i\|^2\overline{\Sigma}_{i+1}\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i}{(\epsilon+\|\overline{\mathbf{u}}_i\|^2)^2}\right]. \tag{50}$$

Although the individual elements of $\overline{\mathbf{u}}_i$ are independent, no closed-form expressions for $\mathbf{A}$ and $\mathbf{B}'$ are available. However, we can carry out the analysis in terms of these matrices as follows. First, we argue in Appendix B that $\mathbf{A}$ is diagonal. We also show that if $\overline{\boldsymbol{\Sigma}}_{i+1}$ is diagonal, then so is $\mathbf{B}'$ and that

$$\text{diag}\left(\mathbf{B}'\right)=\mathbf{B}\ \text{diag}\left(\overline{\boldsymbol{\Sigma}}_{i+1}\right)$$

where $\mathbf{B}$ is the diagonal matrix

$$\mathbf{B}=E\left[\frac{(\overline{\mathbf{u}}_i\odot\overline{\mathbf{u}}_i)^T(\overline{\mathbf{u}}_i\odot\overline{\mathbf{u}}_i)}{(\epsilon+\|\overline{\mathbf{u}}_i\|^2)^2}\right].$$

Here, the notation $\odot$ denotes an element-by-element (Hadamard) product.[2] Thus, the successive $\overline{\boldsymbol{\Sigma}}_i$s in recursion (49) will also be diagonal if the boundary condition is. Subsequently, as in the LMS case, we can again obtain a recursive relation for their diagonal entries of the form $\overline{\boldsymbol{\sigma}}_i=\overline{\boldsymbol{F}}\overline{\boldsymbol{\sigma}}_{i+1}$, where $\overline{\boldsymbol{F}}$ retains the same form, namely

$$\overline{\boldsymbol{F}}=\boldsymbol{I}-\mu\mathbf{A}+\mu^2\mathbf{B}.$$

Mean-square stability now requires that the step-size μ be chosen such that $\overline{\boldsymbol{F}}$ is a stable matrix (i.e., all its eigenvalues should be strictly inside the unit circle). For NLMS, it can be

[2]For two row vectors $\{\mathbf{x},\mathbf{y}\}$, the quantity $\mathbf{x}\odot\mathbf{y}$ is a row vector with the elementwise products; see [24].

verified that $\mu < 2$ is a sufficient condition for this fact to hold, as can be seen from the following argument.

Choosing $\overline{\boldsymbol{\Sigma}}_{i+1} = \boldsymbol{I}$, we have

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2 = E\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\Sigma}_i} + \mu^2\sigma_v^2 E\left[\frac{\|\overline{\mathbf{u}}_i\|^2}{(\epsilon + \|\overline{\mathbf{u}}_i\|^2)^2}\right]$$

and

$$\overline{\boldsymbol{\Sigma}}_i = \boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\boldsymbol{B}'.$$

Obviously, $\boldsymbol{B}' \leq \boldsymbol{A}/2$ so that

$$\boldsymbol{\Sigma}_i \leq \boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\frac{\boldsymbol{A}}{2}$$

and, hence

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2 \leq E\left[\overline{\boldsymbol{w}}_i^T\left(\boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\frac{\boldsymbol{A}}{2}\right)\overline{\boldsymbol{w}}_i\right] + \mu^2\sigma_v^2 E\left[\frac{\|\overline{\mathbf{u}}_i\|^2}{(\epsilon + \|\overline{\mathbf{u}}_i\|^2)^2}\right].$$

Now, it is clear that $0 < \lambda(\boldsymbol{A}/2) < 1$. Moreover, over the interval $0 < \mu < 2$, it holds that

$$\boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\frac{\boldsymbol{A}}{2} \leq \underbrace{\left(1 - 2\mu\lambda_{\min}\left(\frac{\boldsymbol{A}}{2}\right) + \mu^2\lambda_{\min}\left(\frac{\boldsymbol{A}}{2}\right)\right)}_{\alpha}\boldsymbol{I}$$

from which we conclude that

$$E\|\overline{\boldsymbol{w}}_{i+1}\|^2 \leq \alpha E\|\overline{\boldsymbol{w}}_i\|^2 + \mu^2\sigma_v^2 E\left[\frac{\|\overline{\mathbf{u}}_i\|^2}{(\epsilon + \|\overline{\mathbf{u}}_i\|^2)^2}\right]$$

where the scalar coefficient α is positive and strictly less than one for $0 < \mu < 2$. It follows that $E\|\overline{\boldsymbol{w}}_i\|^2$ remains bounded for all i, as desired. It is also straightforward to verify from

$$E\overline{\boldsymbol{w}}_{i+1} = \left[\boldsymbol{I} - \mu E\left(\frac{\overline{\mathbf{u}_i}^T\overline{\mathbf{u}_i}}{\epsilon + \|\overline{\mathbf{u}_i}\|^2}\right)\right]\cdot E\overline{\boldsymbol{w}}_i$$

that $\mu < 2$ guarantees filter stability in the mean as well (just note that $\overline{\mathbf{u}_i}^T\overline{\mathbf{u}_i}/(\epsilon + \|\overline{\mathbf{u}_i}\|^2)$ is a rank-one matrix whose largest eigenvalue is smaller than one).

Finally, repeating the discussion we had for the steady-state performance of LMS, we arrive at the following expressions for the MSE and MSD of normalized LMS:

$$\begin{aligned}\text{MSE} =& \mu^2\sigma_v^2 E\left[\frac{\|\overline{\mathbf{u}}_i\|^2_{(I-\overline{F})^{-1}\lambda}}{\epsilon + \|\overline{\mathbf{u}}_i\|^2}\right]\\ \text{MSD} =& \mu^2\sigma_v^2 E\left[\frac{\|\overline{\mathbf{u}}_i\|^2_{(I-\overline{F})^{-1}1}}{\epsilon + \|\overline{\mathbf{u}}_i\|^2}\right].\end{aligned}$$

These expressions hold for arbitrary colored Gaussian regressors.

The presentation so far illustrates how the energy-conservation approach can be used to perform transient analysis of LMS and its normalized version. Our contribution lies in the ability to perform the analysis in a unified manner. This can be appreciated, for example, by comparing the analysis of the normalized LMS algorithm in [7], [10], [11], [22], and [23] with the analysis in the previous section. A substantial part of prior studies is often devoted to studying the multivariate moments of (50) and, as a result, eventually resort to some whiteness assumption on the data. Our derivation bypasses this requirement. Moreover, earlier approaches do not seem to handle transparently non-Gaussian regression data, which is discussed later in Section V.

V. Data-Normalized Filters

We now consider general data-normalized adaptive filters of the form (26) and drop the Gaussian assumption AG. The analysis that follows shows how to extend the discussions of the previous two sections to this general scenario.

Our starting point is the mean and variance relations (27)–(29).

A. Mean-Square Analysis

For arbitrary regression data, we can no longer guarantee that the data moments

$$E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right], \qquad E\left[\frac{\|\mathbf{u}_i\|^2_\Sigma\mathbf{u}_i^T\mathbf{u}_i}{g^2[\mathbf{u}_i]}\right]$$

are jointly diagonalizable (as we had, for example, in the case of LMS with Gaussian regressors). Consequently, $\boldsymbol{\Sigma}_i$ need not be diagonal even if $\boldsymbol{\Sigma}_{i+1}$ is, i.e., these matrices can no more be fully characterized by their diagonal elements alone. Still, we can perform mean-square analysis by replacing the diag operation with the vec operation, which transforms a matrix into a column vector by stacking all its columns on top of each other.

Let

$$\boldsymbol{\sigma}_{i+1} \triangleq \text{vec}\left(\boldsymbol{\Sigma}_{i+1}\right).$$

Then, using the Kronecker product notation (e.g., [24]) and the following property, for arbitrary matrices $\{\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{\Sigma}\}$:

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

it is straightforward to verify that the recursion (28) for $\boldsymbol{\Sigma}_i$ transforms into the linear vector relation

$$\boldsymbol{\sigma}_i = \boldsymbol{F}\boldsymbol{\sigma}_{i+1}$$

where the coefficient matrix $\boldsymbol{F}$ is now $M^2 \times M^2$ and is given by

$$\boldsymbol{F} \triangleq \boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\boldsymbol{B} \tag{51}$$

with the $M^2 \times M^2$ symmetric matrices $\{\boldsymbol{A}, \boldsymbol{B}\}$ defined by

$$\begin{aligned}\boldsymbol{A} =& \left(E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right] \otimes \boldsymbol{I}_M\right) + \left(\boldsymbol{I}_M \otimes E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right]\right)\\ \boldsymbol{B} =& E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i \otimes \mathbf{u}_i^T\mathbf{u}_i}{g^2[\mathbf{u}_i]}\right].\end{aligned}$$

In particular, $\boldsymbol{A}$ is positive-definite, and $\boldsymbol{B}$ is non-negative-definite. In addition, introduce the $M \times M$ matrix

$$\boldsymbol{P} \triangleq E\left[\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right]$$

which appears in the mean weight-error recursion (29) and in the expression for $\boldsymbol{A}$.

It follows that in terms of the vec notation, the variance relation (27) becomes

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\sigma_{i+1}}\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F\sigma_{i+1}}\right] + \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{\sigma_{i+1}}}{g^2[\mathbf{u}_i]}\right]. \tag{52}$$

Now, contrary to the Gaussian LMS case, the matrix $\boldsymbol{F}$ is no longer guaranteed to be non-negative-definite. It is shown in Appendix A that the condition $-1 < \lambda(F) < 1$ can be enforced for values of μ in the range

$$0 < \mu < \min\left\{\frac{1}{\lambda_{\max}(\boldsymbol{A}^{-1}\boldsymbol{B})},\ \frac{1}{\max\left\{\lambda(\boldsymbol{L}) \in \mathbb{R}^{+}\right\}}\right\} \quad (53)$$

where the second condition is in terms of the largest positive real eigenvalue of the following block matrix:

$$\boldsymbol{L} \triangleq \begin{bmatrix} \boldsymbol{A}/2 & -\boldsymbol{B}/2 \\ \boldsymbol{I}_{M^2} & 0 \end{bmatrix}$$

when it exists. Since $\boldsymbol{L}$ is not symmetric, its eigenvalues may not be positive or even real. If $\boldsymbol{L}$ does not have any real positive eigenvalue, then the corresponding condition is removed from (53), and we only require $\mu < 1/\lambda_{\max}(\boldsymbol{A}^{-1}\boldsymbol{B})$. Condition (53) can be grouped together with the requirement $\mu < 2/\lambda_{\max}(\boldsymbol{P})$, which guarantees convergence in the mean, so that

$$\mu < \min\left\{\frac{2}{\lambda_{\max}(\boldsymbol{P})},\ \frac{1}{\lambda_{\max}(\boldsymbol{A}^{-1}\boldsymbol{B})},\ \frac{1}{\max\left\{\lambda(\boldsymbol{L}) \in \mathbb{R}^{+}\right\}}\right\}. \quad (54)$$

Moreover, the same argument that we used in the LMS case in Section III would show that the transient behavior of data-normalized filters is characterized by the M^2-dimensional state-space model:[3]

$$\mathcal{W}_{i+1} = \mathcal{F}\mathcal{W}_i + \mu^2\sigma_v^2\mathcal{Y} \quad (55)$$

where

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

with

$$p(x) \triangleq \det(x\boldsymbol{I} - \boldsymbol{F}) = x^{M^2} + \sum_{k=0}^{M^2-1} p_k x^k$$

denoting the characteristic polynomial of $\boldsymbol{F}$. In addition, $\mathcal{W}_i$ and $\mathcal{Y}$ are the $M^2 \times 1$ vectors

$$\mathcal{W}_i \triangleq \begin{bmatrix} E\|\tilde{\boldsymbol{w}}_i\|^2_{\sigma} \\ E\|\tilde{\boldsymbol{w}}_i\|^2_{F\sigma} \\ E\|\tilde{\boldsymbol{w}}_i\|^2_{F^2\sigma} \\ \vdots \\ E\|\tilde{\boldsymbol{w}}_i\|^2_{F^{(M^2-1)}\sigma} \end{bmatrix}$$

$$\mathcal{Y} = \begin{bmatrix} E\left(\frac{\|\mathbf{u}_i\|^2_{\sigma}}{g^2[\mathbf{u}_i]}\right) \\ E\left(\frac{\|\mathbf{u}_i\|^2_{F\sigma}}{g^2[\mathbf{u}_i]}\right) \\ E\left(\frac{\|\mathbf{u}_i\|^2_{F^2\sigma}}{g^2[\mathbf{u}_i]}\right) \\ \vdots \\ E\left(\frac{\|\mathbf{u}_i\|^2_{F^{(M^2-1)}\sigma}}{g^2[\mathbf{u}_i]}\right) \end{bmatrix}$$

[3]Observe how the order of the model, in the general case, is M^2 and not M, as was the case in the previous two sections with Gaussian regressors.

for any $\boldsymbol{\sigma}$ of interest, e.g., more commonly, $\boldsymbol{\sigma} = \mathbf{1}$ or $\boldsymbol{\sigma} = \mathbf{r}$, where $\mathbf{r} = \text{vec}(\boldsymbol{R})$.

Moreover, steady-state analysis can be carried out along the same lines of Section III-D. Thus, assuming the filter reaches steady-state, recursion (52) becomes, in the limit

$$\lim_{i\to\infty} E\|\tilde{\boldsymbol{w}}_i\|^2_{(I-F)\sigma_\infty} = \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{\sigma_\infty}}{g^2[\mathbf{u}_i]}\right]$$

in terms of the boundary condition $\boldsymbol{\sigma}_\infty$, which we are free to choose. This expression allows us to evaluate the steady-state value of $E\|\tilde{\boldsymbol{w}}_i\|^2_S$ for any symmetric weighting $\mathbf{S}$ by choosing $\boldsymbol{\sigma}_\infty$ such that $(\boldsymbol{I} - \boldsymbol{F})\boldsymbol{\sigma}_\infty = \text{vec}(\mathbf{S})$. In particular, the EMSE corresponds to the choice $\mathbf{S} = \mathbf{R}$, i.e., $\boldsymbol{\sigma}_\infty = (\boldsymbol{I} - \boldsymbol{F})^{-1}\text{vec}(\mathbf{R})$. Likewise, the MSD is obtained by choosing $\mathbf{S} = \boldsymbol{I}$, i.e., $\boldsymbol{\sigma}_\infty = (\boldsymbol{I} - \boldsymbol{F})^{-1}\text{vec}(\boldsymbol{I})$. We summarize these results in the following statement, which holds for arbitrary input distributions and scalar data nonlinearities.

We summarize in the following statement the above results, which hold for arbitrary input distributions and data nonlinearities.

Theorem 2 (Scalar Nonlinearities): Consider an adaptive filter of the form

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu\frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]}e(i), \quad i \geq 0$$

where $e(i) = d(i) - \mathbf{u}_i\boldsymbol{w}_i$, and $d(i) = \mathbf{u}_i\boldsymbol{w}^o + v(i)$. Assume that the sequences $\{v(i), \mathbf{u}_i\}$ are iid and mutually independent. Then, the filter is stable in the mean and mean-square senses if the step-size μ satisfies (54). Moreover, the resulting EMSE and MSD are given by

$$\text{EMSE} = \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{(I-F)^{-1}\text{vec}(R)}}{g^2[\mathbf{u}_i]}\right]$$

$$\text{MSD} = \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{(I-F)^{-1}\text{vec}(I)}}{g^2[\mathbf{u}_i]}\right]$$

where $\boldsymbol{F}$ is defined by (51). ◇

B. Learning Curves

The learning curve of an adaptive filter refers to the time evolution of $Ee_a^2(i)$; its steady-state value is the MSE. Now, since $Ee_a^2(i) = E\|\tilde{\boldsymbol{w}}_i\|^2_R$, the learning curve can be evaluated by computing $E\|\tilde{\boldsymbol{w}}_i\|^2_R$ for each i. This task can be accomplished recursively from (52) by choosing the boundary condition $\boldsymbol{\sigma}_{i+1}$ as $\mathbf{r} = \text{vec}(\mathbf{R})$. Indeed, iterating (52) with this choice of $\boldsymbol{\sigma}_{i+1}$ and assuming $\boldsymbol{w}_o = 0$, we find that

$$E\|\tilde{\boldsymbol{w}}_{i+1}\|^2_r = \|\boldsymbol{w}^o\|^2_{F^{i+1}r} + \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{(I+F+\cdots+F^i)r}}{g^2[\mathbf{u}_i]}\right]$$

that is

$$E\|\tilde{\boldsymbol{w}}_{i+1}\|^2_r = \|\boldsymbol{w}^o\|^2_{a_i} + \mu^2\sigma_v^2 b_i$$

where the vector $\mathbf{a}_i$ and the scalar b_i satisfy the recursions

$$\mathbf{a}_i = \boldsymbol{F}\mathbf{a}_{i-1}, \quad \mathbf{a}_{-1} = \mathbf{r}$$

$$b_i = b_{i-1} + E\left[\frac{\|\mathbf{u}_i\|^2_{a_{i-1}}}{g^2[\mathbf{u}_i]}\right], \quad b_{-1} = 0.$$

Using these definitions for $\{\mathbf{a}_i, b_i\}$, it is easy to verify that

$$Ee_a^2(i+1) = Ee_a^2(i) + \|\boldsymbol{w}^o\|^2_{F^i(F-I)r} + \mu^2\sigma_v^2 E\left[\frac{\|\mathbf{u}_i\|^2_{F^ir}}{g^2[\mathbf{u}_i]}\right]$$

which describes the learning curve of a data-normalized adaptive filter.

VI. Matrix Nonlinearities

In this section, we extend the earlier results to the case in which the function $g[\mathbf{u}_i]$ is matrix-valued rather than scalar-valued. To motivate this extension, consider the sign-regressor algorithm (e.g., [8]):

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu \text{sgn}[\mathbf{u}_i]^T e(i)$$

where the sgn operates on the individual elements of $\mathbf{u}_i$. This is in contrast to the discussions in the previous sections where all the elements of $\mathbf{u}_i$ were normalized by the same data nonlinearity. Other examples of matrix nonlinearities can be found, e.g., in [25]–[27].

The above update is a special case of more general updates of the form

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu \mathbf{H}[\mathbf{u}_i]\mathbf{u}_i^T e(i) \tag{56}$$

where $\mathbf{H}[\mathbf{u}_i]$ denotes an $M \times M$ matrix nonlinearity.

A. Energy Relation

We first show how to extend the energy relation of Theorem 1 to the more general class of algorithms (56) with matrix data nonlinearities. Our starting point is the adaptation equations (56), which can be written in terms of the weight error vector $\tilde{\boldsymbol{w}}_i$ as

$$\tilde{\boldsymbol{w}}_{i+1} = \tilde{\boldsymbol{w}}_i - \mu \mathbf{H}[\mathbf{u}_i]\mathbf{u}_i^T e(i). \tag{57}$$

By premultiplying both sides of (57) by $\mathbf{u}_i\boldsymbol{\Sigma}$, we see that the estimation errors $e_a^{\Sigma}(i)$, $e_p^{\Sigma}(i)$, and $e(i)$ are related by

$$e_p^{\Sigma}(i) = e_a^{\Sigma}(i) - \mu\|\mathbf{u}_i\|^2_{\Sigma H} e(i). \tag{58}$$

Moreover, the two sides of (57) should have the same weighted-energy, i.e.,

$$\tilde{\boldsymbol{w}}_{i+1}^T \boldsymbol{\Sigma} \tilde{\boldsymbol{w}}_{i+1} = \left(\tilde{\boldsymbol{w}}_i - \mu\mathbf{H}[\mathbf{u}_i]\mathbf{u}_i^T e(i)\right)^T \times \boldsymbol{\Sigma}\left(\tilde{\boldsymbol{w}}_i - \mu\mathbf{H}[\mathbf{u}_i]\mathbf{u}_i^T e(i)\right)$$

so that

$$\begin{aligned}\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma} =& \|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} - 2\mu e(i)\mathbf{u}_i\mathbf{H}[\mathbf{u}_i]\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}_i \\ &+ \mu^2 e^2(i)\mathbf{u}_i\mathbf{H}[\mathbf{u}_i]\boldsymbol{\Sigma}\mathbf{H}[\mathbf{u}_i]\mathbf{u}_i^T \\ =& \|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} - 2\mu e_a^{H\Sigma}(i)e(i) \\ &+ \mu^2 e^2(i)\|\mathbf{u}_i\|^2_{H\Sigma H}.\end{aligned} \tag{59}$$

This form of the energy relation is analogous to (15). As it stands, (59) is just what we need for mean-square analysis. For completeness, however, we develop a cleaner form of (59): a form similar to (10). To this end, notice that upon replacing $\boldsymbol{\Sigma}$ by $\mathbf{H}\boldsymbol{\Sigma}$ in (58), we get

$$\mu\|\mathbf{u}_i\|^2_{H\Sigma H} e(i) = e_a^{H\Sigma}(i) - e_p^{H\Sigma}(i)$$

or, by incorporating the defining expression (9) of $\overline{\mu}_{(\cdot)}(i)$

$$\mu e(i) = \overline{\mu}_{H\Sigma H}(i)\left(e_a^{H\Sigma}(i) - e_p^{H\Sigma}(i)\right). \tag{60}$$

Substituting (60) into (59) produces the desired energy relation form

$$\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma} + \overline{\mu}_{H\Sigma H}(i)\left|e_a^{H\Sigma}(i)\right|^2 = \|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} + \overline{\mu}_{H\Sigma H}(i)\left|e_p^{H\Sigma}(i)\right|^2.$$

B. Mean-Square Analysis

To perform mean-square analysis, we start with form (59) of the energy relation. Bearing in mind the independence assumption on the noise AN and the fact that $e(i) = e_a(i) + v(i)$, (59) reads, under expectation

$$\begin{aligned}E\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma_{i+1}} =& E\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_{i+1}} - 2\mu E\left[e_a^{H\Sigma_{i+1}}(i)e_a(i)\right] + \\ & \mu^2 E\left[e_a^2(i)\|\mathbf{u}_i\|^2_{H\Sigma_{i+1}H}\right] + \mu^2\sigma_v^2 E\left[\|\mathbf{u}_i\|^2_{H\Sigma_{i+1}H}\right]\end{aligned}$$

where the weight $\boldsymbol{\Sigma}$ was replaced by the time-indexed weight $\boldsymbol{\Sigma}_{i+1}$. If we further invoke the polarization identity (12), we get

$$\begin{aligned}e_a^{H\Sigma_{i+1}}(i)e_a(i) =& \|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_{i+1}Hu_i^Tu_i} = \|\tilde{\boldsymbol{w}}_i\|^2_{u_i^Tu_iH\Sigma_{i+1}} \quad \text{and} \\ e_a^2(i) =& \|\tilde{\boldsymbol{w}}_i\|^2_{u_i^Tu_i}.\end{aligned}$$

These equations, together with the linearity property (11) and the independence assumption AI, yield the following result.

Theorem 3 (Matrix Nonlinearities): Consider an adaptive filter of the form

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu\mathbf{H}[\mathbf{u}_i]\mathbf{u}_i^T e(i), \quad i \geq 0$$

where $e(i) = d(i) - \mathbf{u}_i\boldsymbol{w}_i$, and $d(i) = \mathbf{u}_i\boldsymbol{w}^o + v(i)$. Assume that the sequences $\{v(i), \mathbf{u}_i\}$ are iid and mutually independent. Then, it holds that

$$E\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma_{i+1}} = E\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_i} + \mu^2\sigma_v^2 E\left[\|\mathbf{u}_i\|^2_{H\Sigma_{i+1}H}\right] \tag{61}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i+1} - \mu\boldsymbol{\Sigma}_{i+1}E\left[\mathbf{H}\mathbf{u}_i^T\mathbf{u}_i\right] - \mu E\left[\mathbf{u}_i^T\mathbf{u}_i\mathbf{H}\right]\boldsymbol{\Sigma}_{i+1} \\ +\mu^2 E\left[\|\mathbf{u}_i\|^2_{H\Sigma_{i+1}H}\mathbf{u}_i^T\mathbf{u}_i\right].\end{aligned} \tag{62}$$

In addition, the stability condition and the MSE and MSD expressions of Theorem 2 apply here as well with $\{\boldsymbol{A}, \boldsymbol{B}\}$ replaced by

$$\begin{aligned}\boldsymbol{A} =& \left(E[\mathbf{u}_i^T\mathbf{u}_i\mathbf{H}] \otimes \boldsymbol{I}\right) + \left(\boldsymbol{I} \otimes E[\mathbf{u}_i^T\mathbf{u}_i\mathbf{H}]\right) \\ \boldsymbol{B} =& E[\mathbf{u}_i^T\mathbf{u}_i\mathbf{H} \otimes \mathbf{u}_i^T\mathbf{u}_i\mathbf{H}].\end{aligned}$$

Moreover, the construction of the learning curve in Section V-B also extends to this case. ◊

Compared with some earlier studies (e.g., [8], [9], [25]), the above results hold without restricting the regression data to being Gaussian or white.

C. Sign-Regressor Algorithm

To illustrate the application of the above results, we return to the sign-regressor recursion

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu\text{sgn}[\mathbf{u}_i]^T e(i).$$

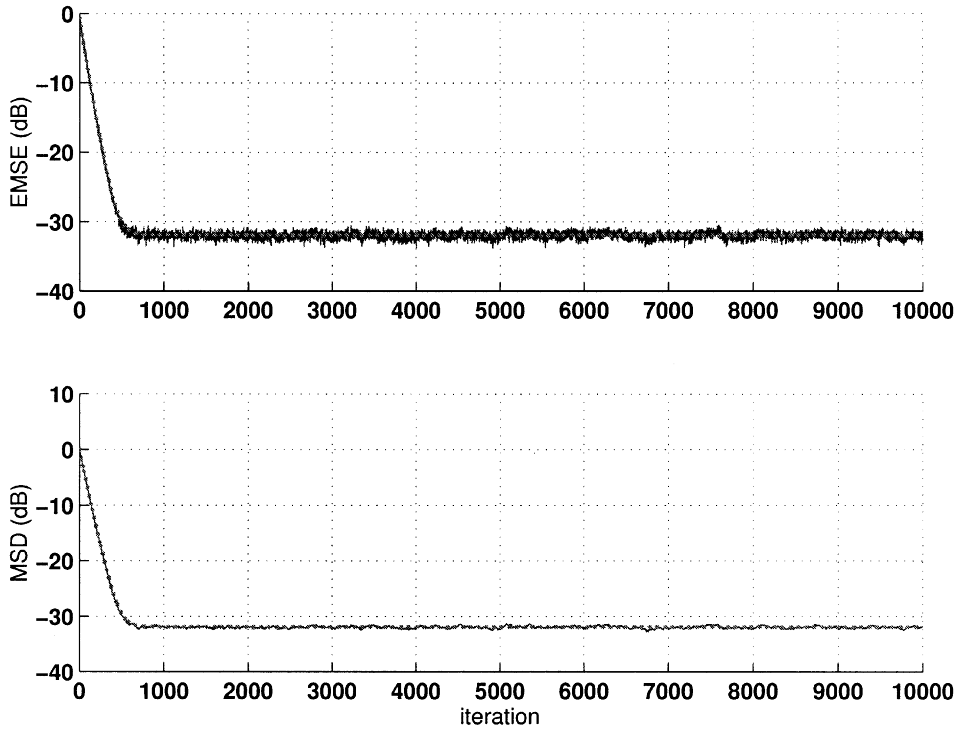


Fig. 1. Theoretical and simulated learning and MSD curves for LMS using correlated uniform input data and $a = 0.2$.

In this case, the matrix nonlinearity $\mathbf{H}[\mathbf{u}_i]$ is implicitly defined by the identity

$$\mathbf{u}_i\mathbf{H}[\mathbf{u}_i] = \text{sgn}[\mathbf{u}_i]$$

which in turn means that relations (61) and (62) become

$$E\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma_{i+1}} = E\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_i} + \mu^2\sigma_v^2 E\left[\ \|\text{sgn}[\mathbf{u}_i]\|^2_{\Sigma_{i+1}}\ \right] \quad (63)$$

and

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i+1} - \mu\boldsymbol{\Sigma}_{i+1}E\left[\text{sgn}[\mathbf{u}_i]^T\mathbf{u}_i\right] - \mu E\left[\mathbf{u}_i^T\text{sgn}[\mathbf{u}_i]\right]\boldsymbol{\Sigma}_{i+1} + \mu^2 E\left[\|\text{sgn}[\mathbf{u}_i]\|^2_{\Sigma_{i+1}}\mathbf{u}_i^T\mathbf{u}_i\right].$$

Assume that the individual entries of the regressor $\mathbf{u}_i$ have variance σ_u^2. In addition, assume that $\mathbf{u}_i$ has a Gaussian distribution. Then, it follows from Price's theorem [29] that[4]

$$E\left[\text{sgn}[\mathbf{u}_i]^T\mathbf{u}_i\right] = \sqrt{\frac{2}{\pi\sigma_u^2}}\mathbf{R}$$

which leads to

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i+1} - \mu\sqrt{\frac{2}{\pi\sigma_u^2}}\boldsymbol{\Sigma}_{i+1}\mathbf{R} - \mu\sqrt{\frac{2}{\pi\sigma_u^2}}\mathbf{R}\boldsymbol{\Sigma}_{i+1} + \mu^2 E\left[\|\text{sgn}[\mathbf{u}_i]\|^2_{\Sigma_{i+1}}\mathbf{u}_i^T\mathbf{u}_i\right]. \quad (64)$$

Now, observe that $\|\text{sgn}[\mathbf{u}_i]\|^2_{\Sigma_{i+1}} = \text{Tr}(\boldsymbol{\Sigma}_{i+1})$ whenever $\boldsymbol{\Sigma}_{i+1}$ is diagonal. Thus, assume we choose $\boldsymbol{\Sigma}_{i+1} = \boldsymbol{I}$. Then, the expression for $\boldsymbol{\Sigma}_i$ becomes

$$\boldsymbol{\Sigma}_i = \boldsymbol{I} + \mu\left(\mu M - 2\sqrt{\frac{2}{\pi\sigma_u^2}}\right)\mathbf{R}$$

[4]The theorem can be used to show that for two jointly zero-mean Gaussian real-valued random variables x and y, it holds that $E\left(x\text{sgn}(y)\right) = \sqrt{2/\pi}1/\sigma_y E\left(xy\right)$.

whereas (63) becomes

$$E\|\tilde{\boldsymbol{w}}_{i+1}\|^2 = E\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_i} + \mu^2\sigma_v^2 M. \quad (65)$$

It is now easy to verify that $E\|\overline{\boldsymbol{w}}_{i+1}\|^2$ converges, provided that $\lambda_{\max}(\boldsymbol{\Sigma}_i) < 1$ or, equivalently

$$\mu < \sqrt{\frac{8}{\pi\sigma_u^2}}\frac{1}{M}.$$

This is the same condition derived in [8].

To evaluate the MSE, we observe from (65) that in steady state

$$\mu\left(2\sqrt{\frac{2}{\pi\sigma_u^2}} - \mu M\right)\lim_{i\to\infty}E\|\boldsymbol{w}_i\|^2_R = \mu^2\sigma_v^2 M$$

so that

$$\text{MSE} = \frac{\mu\sigma_v^2 M}{\sqrt{\frac{8}{\pi\sigma_u^2}} - \mu M}$$

which is again the same expression from [8].

VII. Simulations

Throughout this section, the system to be identified is an FIR channel of length 4. The input $u(i)$ is generated by passing an iid uniform process $x(i)$ through a first-order model

$$u(i) = au(i-1) + x(i). \quad (66)$$

By varying the value of a, we obtain processes $u(i)$ of different colors. We simulate the choices $a = 0.2$ and $a = 0.9$. The input sequence that is feeding the adaptive filter therefore has a correlated uniform distribution. The output of the channel is contaminated by an iid Gaussian additive noise at an SNR level of 30 dB.

Figs. 1 and 2 shows the resulting theoretical and simulated learning and MSD curves for both cases of $a = 0.2$ and $a =$

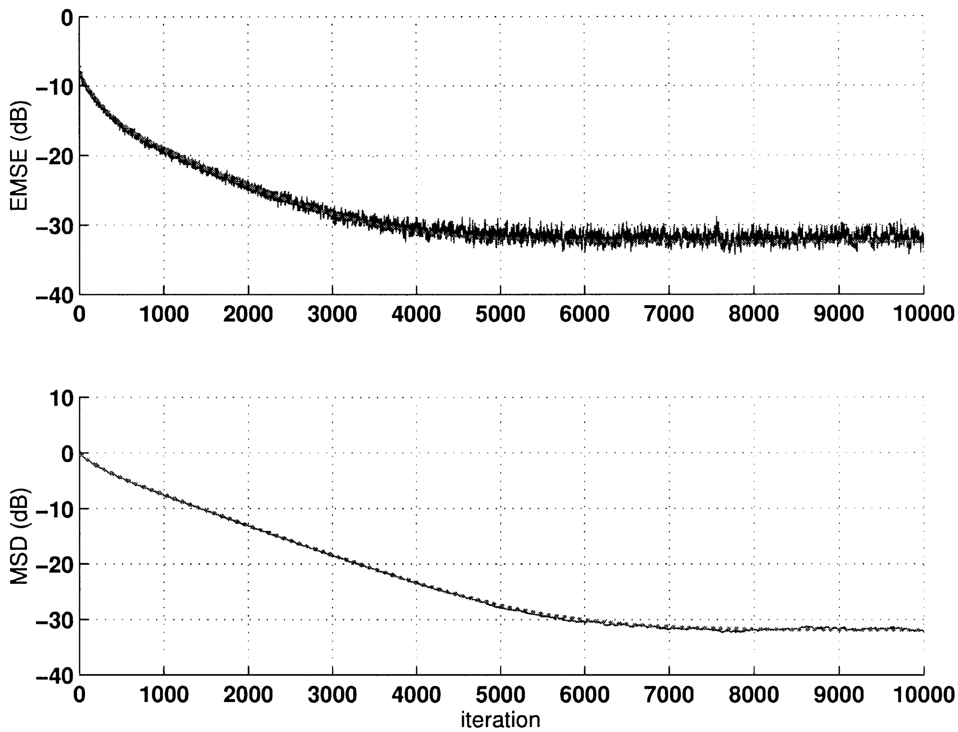


Fig. 2. Theoretical and simulated learning curves for LMS using correlated uniform input data and $a = 0.9$.

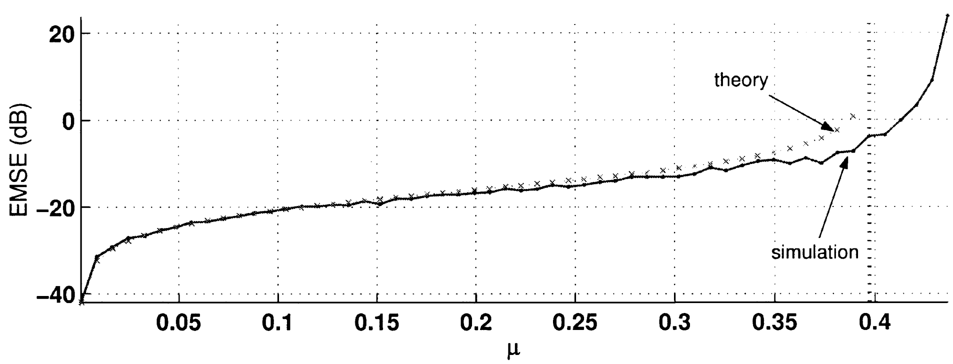


Fig. 3. Theoretical and simulated EMSE for LMS as a function of the step-size for correlated uniform input with $a = 0.2$.

0.9. The simulated curves are obtained by averaging over 200 experiments, whereas the theoretical curves are obtained from the state-space model (55). It is seen that there is a good match between theory and practice.

Fig. 3 examines the stability bound (54); it plots the filter EMSE as a function of the step size using the theoretical expression from Theorem 2, in addition to a simulated EMSE. The bound on the step size is also indicated.

VIII. Concluding Remarks

In this paper, we developed a framework for the transient analysis of adaptive filters with general data nonlinearities (both scalar-valued and matrix-valued). The approach relies on energy conservation arguments. By suitably choosing the boundary condition of the weighting matrix recursion, we can obtain MSE and MSD results and the conditions for mean-square stability. We may add that extensions to leaky algorithms, affine projection algorithms, filters with error nonlinearities, and to tracking analysis are possible and are treated in, e.g., [20], [30], and [31].

Appendix A
Condition for Mean-Square Stability

Consider the matrix form $\boldsymbol{F} = \boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\boldsymbol{B}$ with $\boldsymbol{A} > 0$, $\boldsymbol{B} \geq 0$, and $\mu > 0$. We would like to determine conditions on μ in order to guarantee that the eigenvalues of $\boldsymbol{F}$ satisfy $-1 < \lambda(\boldsymbol{F}) < 1$.

First, in order to guarantee $\lambda(\boldsymbol{F}) < 1$, the step-size μ should be such that $\boldsymbol{F} < \boldsymbol{I}$ or, equivalently, $\boldsymbol{A} - \mu\boldsymbol{B} > 0$. This condition is equivalent to requiring $\boldsymbol{I} - \mu\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-*/2} > 0$, but since the matrices $\boldsymbol{A}^{-1}\boldsymbol{B}$ and $\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-*/2}$ are similar, we conclude that μ should satisfy $\mu < 1/\lambda_{\max}(\boldsymbol{A}^{-1}\boldsymbol{B})$.

In order to enforce $\lambda(\boldsymbol{F}) > -1$, the step-size μ should be such that $\boldsymbol{G}(\mu) = 2\boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\boldsymbol{B} > 0$. When $\mu = 0$, the eigenvalues of $\boldsymbol{G}$ are positive and equal to 2. As μ increases, the eigenvalues of $\boldsymbol{G}$ vary continuously with μ. Therefore, an upper bound on μ that guarantees $\boldsymbol{G}(\mu) > 0$ is determined by the smallest μ that makes $\boldsymbol{G}(\mu)$ singular.

Now, the determinant of $\boldsymbol{G}(\mu)$ is equal to the determinant of the block matrix

$$\boldsymbol{K}(\mu) \triangleq \begin{bmatrix} 2\boldsymbol{I} - \mu\boldsymbol{A} & \mu\boldsymbol{B} \\ -\mu\boldsymbol{I} & \boldsymbol{I} \end{bmatrix}.$$

Moreover, since

$$\boldsymbol{K}(\mu) = \begin{bmatrix} 2\boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{bmatrix} - \mu \begin{bmatrix} \frac{\boldsymbol{A}}{2} & -\frac{\boldsymbol{B}}{2} \\ \boldsymbol{I} & 0 \end{bmatrix} \right)$$

the condition $\det(\boldsymbol{K}(\mu)) = 0$ is equivalent to $\det(\boldsymbol{I} - \mu\boldsymbol{L}) = 0$, where

$$\boldsymbol{L} \triangleq \begin{bmatrix} \frac{\boldsymbol{A}}{2} & -\frac{\boldsymbol{B}}{2} \\ \boldsymbol{I} & 0 \end{bmatrix}.$$

In this way, the smallest positive μ that results in $\det(\boldsymbol{K}(\mu)) = 0$ is given by

$$\mu < \frac{1}{\max\left\{\lambda(\boldsymbol{L}) \in \mathbb{R}^+\right\}}.$$

This condition is in terms of the largest positive real eigenvalue of $\boldsymbol{L}$ when it exists. It follows that the following range of μ guarantees a stable $\boldsymbol{F}$,

$$0 < \mu < \min\left\{ \frac{1}{\lambda_{\max}(\boldsymbol{A}^{-1}\boldsymbol{B})}, \frac{1}{\max\left\{\lambda(\boldsymbol{L}) \in \mathbb{R}^+\right\}} \right\}.$$

APPENDIX B
$\mathbf{A}$ AND $\mathbf{B}'$ OF (50) ARE DIAGONAL

An off-diagonal entry of $\boldsymbol{A}$ has the form

$$A_{jk} = E\left[\frac{2\overline{u}_{i_j}\overline{u}_{i_k}}{\epsilon + \|\overline{\mathbf{u}}_i\|^2} \right].$$

Now, $2\overline{u}_{i_j}^T\overline{u}_{i_k}/\epsilon + \|\overline{\mathbf{u}}_i\|^2$ is an odd function of u_{i_k}, which has an even (Gaussian) pdf and is independent of the other elements of $\overline{\mathbf{u}}_i$. Thus, $E\left[\overline{u}_{i_k}\overline{u}_{i_k}/\epsilon + \|\overline{\mathbf{u}}_i\|^2 \mid \overline{u}_{i_j}\right] = 0$, and hence, A_{jk} is zero as well. Therefore, $\mathbf{A}$ is diagonal. A similar argument can be used to prove that $\mathbf{B}'$ is diagonal. Now, a diagonal entry of $\mathbf{B}'$ can be written as

$$\begin{aligned} B'_{kk} =& E\left[\frac{\overline{u}_{i_k}^2 \|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}_i}}{\left(\epsilon + \|\overline{\mathbf{u}}_i\|^2\right)^2} \right] \\ =& E\left[\frac{\overline{u}_{i_k}^2}{\left(\epsilon + \|\overline{\mathbf{u}}_i\|^2\right)^2} \overline{\mathbf{u}}_i \odot \overline{\mathbf{u}}_i \right] \text{diag}\left(\overline{\mathbf{\Sigma}}_{i+1}\right). \end{aligned}$$

It follows that

$$\text{diag}\left(\mathbf{B}'\right) = \mathbf{B}\ \text{diag}\left(\overline{\mathbf{\Sigma}}_{i+1}\right)$$

where

$$\mathbf{B} = E\left[\frac{\left(\overline{\mathbf{u}}_i \odot \overline{\mathbf{u}}_i\right)^T \left(\overline{\mathbf{u}}_i \odot \overline{\mathbf{u}}_i\right)}{\left(\epsilon + \|\overline{\mathbf{u}}_i\|^2\right)^2} \right].$$

ACKNOWLEDGMENT

The authors are grateful to Dr. V. H. Nascimento for his assistance with the material in Appendix A.

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. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[3] O. Macchi, *Adaptive Processing: The LMS Approach with Applications in Transmission*. New York: Wiley, 1995.

[4] 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. IT-28, pp. 318–329, Mar. 1982.

[5] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, Jan. 1985.

[6] V. Mathews and S. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–35, pp. 450–454, Apr. 1987.

[7] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data," *IEEE Trans. Inform. Theory*, vol. 34, pp. 680–690, July 1988.

[8] E. Eweda, "Analysis and design of signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 1367–1374, Nov. 1990.

[9] S. C. Douglas and T. H. -Y. Meng, "The optimum scalar data nonlinearity in LMS adaptation for arbitrary iid inputs," *IEEE Trans. Signal Processing*, vol. 40, pp. 1566–1570, June 1992.

[10] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Trans. Signal Processing*, vol. 41, pp. 1149–1160, Mar. 1993.

[11] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2811–2825, Sept. 1993.

[12] J. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally-spaced blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 48, pp. 80–91, Jan. 2000.

[13] 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.

[14] ——, "Ability of adaptive filters to track carrier offsets and random channel nonstationarities," *IEEE Trans. Signal Processing*, vol. 50, pp. 1533–1544, July 2002.

[15] ——, "A feedback analysis of the tracking performance of blind adaptive equalization algorithms," in *Proc. Conf. Decision Contr.*, vol. 1, Phoenix, AZ, Dec. 1999, pp. 174–179.

[16] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive algorithms via the small gain theorem," in *Proc. SPIE*, vol. 2563, San Diego, CA, July 1995, pp. 458–69.

[17] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1428–1439, June 1996.

[18] A. H. Sayed and M. Rupp, "Robustness issues in adaptive filtering," in *DSP Handbook*. Boca Raton, FL: CRC, 1998, ch. 20.

[19] M. Rupp and A. H. Sayed, "On the convergence of blind adaptive equalizers for constant-modulus signals," *IEEE Trans. Communications*, vol. 48, pp. 795–803, May 2000.

[20] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Processing*, vol. 51, pp. 653–663, Mar. 2002.

[21] B. Widrow, J. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 46, pp. 1151–1162, Aug. 1976.

[22] N. J. Bershad, "Behavior of the ϵ-normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–35, pp. 636–644, May 1987.

[23] ——, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-34, pp. 793–806, 1986.

[24] H. Lütkepohl, *Handbook of Matrices*. New York: Wiley, 1996.

[25] N. J. Bershad, "On the optimum data nonlinearity in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 69–76, Feb. 1986.

[26] W. B. Mikhael *et al.*, "Adaptive filters with individual adaptation of parameters," *IEEE Trans. Circuits Syst.*, vol. CAS–33, July 1986.

[27] R. W. Harris, D. M. Chabries, and F. A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 309–316, Apr. 1986.

[28] S. C. Douglas and T. H. -Y. Meng, "Normalized data nonlinearities for LMS adaptation," *IEEE Trans. Signal Processing*, vol. 42, pp. 1352–1365, June 1994.

[29] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 69–72, June 1958.

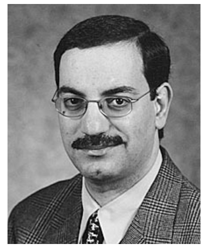
[30] A. H. Sayed and T. Y. Al-Naffouri, "Mean-square analysis of normalized leaky adaptive filters," in *Proc. ICASSP*, vol. VI, Salt Lake City, UT, May 2001.

[31] H.-C. Shin and A. H. Sayed, "Transient behavior of affine projection algorithms," in *Proc. ICASSP*, Hong Kong, Apr. 2003, to be published.

Tareq Y. Al-Naffouri received the B.S. degree in mathematics (with honors) and the M.S. degree in electrical engineering from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 1994 and 1997, respectively, and the M.S. degree in electrical engineering from Georgia Institute of Technology, Atlanta, in 1998. He is currently pursuing the Ph.D. degree with the Electrical Engineering Department, Stanford University, Stanford, CA.

His research interests lie in the area of signal processing for communications. Specifically, he is interested in the analysis and design of algorithms for channel identification and equalization. He has held internship positions at NEC Research Labs, Tokyo, Japan, and at National Semiconductor, Santa Clara, CA.

Mr. Al-Naffouri is the recipient of a 2001 best student paper award at an international meeting for work on adaptive filtering analysis.



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). He has over 180 journal and conference publications, is the author of the forthcoming 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 Procesing 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 and is now serving as Editor-in-Chief of the TRANSACTIONS.