Transient Analysis of Adaptive Filters With Error Nonlinearities

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

***Abstract*—This paper develops a unified approach to the transient analysis of adaptive filters with error nonlinearities. In addition to deriving earlier results in a unified manner, the approach also leads to new performance results without restricting the regression data to being Gaussian or white. The framework is based on energy-conservation arguments and avoids the need for explicit recursions for the covariance matrix of the weight-error vector.**

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

I. INTRODUCTION

THIS paper describes a unifying framework for the study of the transient performance of adaptive filters that involve error nonlinearities in their update equations (e.g., [1]–[3]). This class of algorithms is among the most difficult to analyze, and it is not uncommon to resort to different methods and assumptions with the intent of performing tractable analyses. Before discussing the features of the approach proposed herein and its contributions, we provide, as a motivation, a summary of selected techniques that have been employed earlier in the literature for the study of such algorithms.

a) *Linearization* (e.g., [4]–[7]). In this method of analysis, the error nonlinearity is linearized around an operating point, and higher order terms are discarded. Analyses that are based on this technique fail to accurately describe the adaptive filter performance for large values of the error, e.g., at early stages of adaptation.

b) *Restricted classes of nonlinearities* (e.g., [8]–[14]). Here, the analysis is restricted to particular classes of algorithms such as the sign-LMS algorithm, the least-mean mixed-norm (LMMN) algorithm, the least-mean fourth (LMF) algorithm, and error saturation nonlinearities. By limiting the study to a specific nonlinearity or to a class of nonlinearities, it is possible to avoid linearization, and the analysis results become more accurate.

c) *Assumptions on the statistics of the errors*. While it is common to impose statistical assumptions on the regression and noise sequences, similar conditions can also be imposed on error quantities. For example, in studying the sign-LMS algorithm, it was assumed in [15] that the elements of the weight-error vector are jointly Gaussian. This assumption was shown in [16] to be valid asymptotically. More accurate is the assumption that the residual error is Gaussian [4], [10] or that its conditional value is [8], [9]. By central limit arguments, this assumption is justified for long adaptive filters [4], [10]. More importantly, this assumption is as valid in the early stages as in the final stages of adaptation. For shorter filters, exact expectation analysis can be employed as in [17]–[19].

d) *Restricted class of inputs*. It is common to assume that the input sequence is white and/or has a Gaussian distribution (e.g., [4], [6], [8]–[12], [20]–[22]).

e) *Independence assumption*. It is even more common to assume that the successive regressors are independent in what is widely known as the independence assumptions [1], [23]. Despite being unrealistic, the independence assumptions are among the most heavily used assumptions in adaptive filtering analysis.

f) *Gaussian noise*. Noise is sometimes restricted to be iid Gaussian as in [4], [8], [15], and [24], although Gaussianity is not as common as the previous assumptions. Surprisingly perhaps, the iid assumption on the noise is almost indispensable, even for the analysis of the simplest of adaptive algorithms.

A. Approach of This Paper

In this paper, we develop an approach that applies to arbitrary error nonlinearities. The arguments assume that the adaptive filter is long enough to justify the following approximations.

i) The residual error $e_a(i)$, to be later defined in (6), can be assumed to be Gaussian.

ii) The norm of the input regressor can be assumed to be uncorrelated with $f^2[e(i)]$, which is the square of the error nonlinearity to be defined later in (2).

Both of these assumptions are realistic for longer adaptive filters (see, e.g., the simulation results in Section V-A). Fortunately, they are also realistic in all stages of adaptation (including the early stages).

The approach we adopt is based on the works [25]–[28], where a unified approach to the steady-state and tracking performances of adaptive filters has been developed that makes it possible not only to treat various 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 a fundamental energy conservation relation that holds for a large

Manuscript received March 12, 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 supported in part by a fellowship from King Fahd University of Petroleum and Minerals, Dharan, Saudi Arabia. The associate editor coordinating the review of this paper and approving it for publication was Dr. Dennis R. Morgan.

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.2002.808108

class of adaptive filters. This relation has been originally developed in [29]–[32] in the context of robustness analysis of adaptive filters within a deterministic framework. It has since then been used in [25]–[28] as a convenient tool for studying the steady-state performance of adaptive filters within a stochastic framework as well.

In this work, we show how to extend the same energy-based approach to the transient analysis of adaptive filters with error nonlinearities. Such an extension is desirable since it allows us to bring forth benefits such as the convenience of a unified treatment, the derivation of stability and convergence results, and the weakening of some assumptions. The main contributions, and an outline of this paper, are as follows.

1) We set the stage in the next section by introducing our notation. We proceed by defining the adaptive filtering problem and some associated error quantities. The energy of these errors are finally related through a fundamental energy relation, which will be the starting point for much of the subsequent analysis. This result is summarized in Theorem 1.
2) The energy relation is used in Section III to derive a general recursion that describes the mean-square evolution (i.e., learning curve) of an adaptive filter with error nonlinearity. To achieve this result, we rely on the long filter assumptions, which are formally introduced in this section. The independence assumption turns out to be useful in constructing the dynamical relation. The main contribution of this section is summarized in Theorem 2, which essentially states that the mean-square behavior of an adaptive filter with error nonlinearity is equivalent to that of a nonlinear *time-invariant* state-space model. The statement of the theorem describes this model.
3) In Section IV, we show that the excess mean-square error (EMSE) of an adaptive filter with error nonlinearity can be obtained as the fixed point of a nonlinear function. The main result here is Theorem 3 and Corollary 2, which hold with a weaker form of the independence assumption.

In a companion paper [33], we similarly extend the energy-conservation approach to study the transient behavior of adaptive filters with data normalization.

B. Notation

We focus on real-valued data, although the extension to complex-valued data is immediate. Small boldface letters are used to denote vectors, e.g., $\mathbf{w}$, and the symbol T denotes transposition. The notation $\|\mathbf{w}\|^2$ denotes the squared Euclidean norm of a vector $\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w}$, whereas $\|\mathbf{w}\|^2_{\boldsymbol{\Sigma}}$ denotes the weighted squared Euclidean norm $\|\mathbf{w}\|^2_{\boldsymbol{\Sigma}} = \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$. 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., $\mathbf{w}_i$ and $e(i)$.

C. Weighted-Norms

We will make substantial use of weighted-norms in this paper. Thus, for for ease of reference, we summarize below some of their properties. Thus, let a_1 and a_2 be scalars, $\tilde{\mathbf{w}}_i$ a column vector, and $\mathbf{u}_i$ a row vector, and let $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ be symmetric matrices. Then, the following properties hold.

1) **Superposition.**

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

2) **Polarization**. Since

$$\begin{aligned}(\mathbf{u}_i\boldsymbol{\Sigma}_1\tilde{\mathbf{w}}_i)(\mathbf{u}_i\boldsymbol{\Sigma}_2\tilde{\mathbf{w}}_i) &= \left(\tilde{\mathbf{w}}_i^T\boldsymbol{\Sigma}_1\mathbf{u}_i^T\right)(\mathbf{u}_i\boldsymbol{\Sigma}_2\tilde{\mathbf{w}}_i)\\ &= \tilde{\mathbf{w}}_i^T\left(\boldsymbol{\Sigma}_1\mathbf{u}_i^T\mathbf{u}_i\boldsymbol{\Sigma}_2\right)\tilde{\mathbf{w}}_i\\ &= \tilde{\mathbf{w}}_i^T\left(\boldsymbol{\Sigma}_2\mathbf{u}_i^T\mathbf{u}_i\boldsymbol{\Sigma}_1\right)\tilde{\mathbf{w}}_i\end{aligned}$$

we can write

$$(\mathbf{u}_i\boldsymbol{\Sigma}_1\tilde{\mathbf{w}}_i)(\mathbf{u}_i\boldsymbol{\Sigma}_2\tilde{\mathbf{w}}_i) = \|\tilde{\mathbf{w}}_i\|^2_{\Sigma_1 u_i^T u_i\Sigma_2} = \|\tilde{\mathbf{w}}_i\|^2_{\Sigma_2 u_i^T u_i\Sigma_1}. \tag{1}$$

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

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

II. Adaptive Algorithms With Error Nonlinearity

An adaptive filter attempts to identify a weight vector $\mathbf{w}^o$, of length M, by using a sequence of row regressors $\{\mathbf{u}_i\}$, of length M, and output samples $\{d(i)\}$ that are related via

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

Here, $v(i)$ accounts for measurement noise and modeling errors. Many adaptive schemes have been proposed in the literature for this purpose (see, e.g., [1]–[3]). In this paper, we focus on the class of algorithms

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu\mathbf{u}_i^T f[e(i)], \quad i \geq 0 \tag{2}$$

where $\mathbf{w}_i$ is the estimate of $\mathbf{w}$ at time i, μ is the step size

$$e(i) \triangleq d(i) - \mathbf{u}_i\mathbf{w}_i = \mathbf{u}_i\mathbf{w}^o - \mathbf{u}_i\mathbf{w}_i + v(i) \tag{3}$$

is the estimation error, and $f[e(i)]$ is a scalar function of the error $e(i)$. Table I lists some common adaptive algorithms and their corresponding error nonlinearities.[1]

A. Error Measures

Given an adaptive filter of the family (2), we are interested in studying the time-evolution and the steady-state values of the variances

$$E|e(i)|^2 \quad \text{and} \quad E\|\tilde{\mathbf{w}}_i\|^2 \tag{4}$$

where $\tilde{\mathbf{w}}_i$ stands for the weight-error vector

$$\tilde{\mathbf{w}}_i = \mathbf{w}^o - \mathbf{w}_i.$$

The steady-state values of the above variances represent the mean-square-error and the mean-square-deviation performances of the filter, respectively, whereas their time-evolution relate to the learning or the transient behavior of the filter.

[1]In this table, LMF stands for the least-mean fourth algorithm [5], whereas LMMN stands for the least-mean mixed-norm algorithm [13], [14].

TABLE I
EXAMPLES FOR $f[e(i)]$

ALGORITHM	ERROR NONLINEARITIES $f[e(i)]$
LMS	$e(i)$
LMF	$e^3(i)$
LMF family	$e^{2k+1}(i)$
LMMN	$ae(i)+be^3(i)$
Sign error	$\mathrm{sgn}[e(i)]$
Sat. nonlin.	$\int_0^{e(i)} \exp\left(-\frac{z^2}{2\sigma_{sat}^2}\right) dz$

In order to study the variances (4), the framework of this paper relies on introducing the weighted *a priori* and *a posteriori* errors defined by

$$e_a^{\Sigma}(i) \triangleq \mathbf{u}_i \mathbf{\Sigma} \tilde{\mathbf{w}}_i, \quad e_p^{\Sigma}(i) \triangleq \mathbf{u}_i \mathbf{\Sigma} \tilde{\mathbf{w}}_{i+1} \tag{5}$$

for some symmetric positive definite weighting matrix $\mathbf{\Sigma}$ to be specified later; it will be seen that different choices for $\mathbf{\Sigma}$ allow us to evaluate different performance measures of an adaptive filter. We will use a more standard notation for the usual case $\mathbf{\Sigma} = \mathbf{I}$, namely

$$e_a(i) \triangleq e_a^I(i) = \mathbf{u}_i \tilde{\mathbf{w}}_i, \quad e_p(i) \triangleq e_p^I(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i+1}. \tag{6}$$

With the error quantities $\{\tilde{\mathbf{w}}_i, e_p^{\Sigma}(i), e_a^{\Sigma}(i)\}$ so defined, we can rewrite the adaptation and filtering (2) and (3) in terms of them. Specifically, by subtracting $\mathbf{w}^o$ from both sides of (2), we get

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - \mu f[e(i)] \mathbf{u}_i^T \tag{7}$$

and by combining the defining expressions (3) and (6), we obtain

$$e(i) = e_a(i) + v(i). \tag{8}$$

The estimation errors $e_a^{\Sigma}(i)$, $e_p^{\Sigma}(i)$, and $e(i)$ can be related by premultiplying both sides of the adaptation (7) by $\mathbf{u}_i \mathbf{\Sigma}$

$$\mathbf{u}_i \mathbf{\Sigma} \tilde{\mathbf{w}}_{i+1} = \mathbf{u}_i \mathbf{\Sigma} \tilde{\mathbf{w}}_i - \mu f[e(i)] \|\mathbf{u}_i\|_{\Sigma}^2$$

and incorporating the defining expressions (5), which yield

$$e_p^{\Sigma}(i) = e_a^{\Sigma}(i) - \frac{\mu}{\bar{\mu}_{\Sigma}(i)} f[e(i)] \tag{9}$$

where

$$\bar{\mu}_{\Sigma}(i) \triangleq \begin{cases} \frac{1}{\|\mathbf{u}_i\|_{\Sigma}^2}, & \text{if } \|\mathbf{u}_i\|_{\Sigma}^2 \neq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

B. Weighted-Energy Relation

We are now in a position to derive a weighted-energy relation that relates the energy of the error quantities $\{\tilde{\mathbf{w}}_i, \tilde{\mathbf{w}}_{i+1}, e_a^{\Sigma}(i), e_p^{\Sigma}(i)\}$. This relation will be instrumental in achieving our stated objective of studying the steady-state and transient performances of adaptive filters of the form (2).

First, we determine a relation between the errors. This is obtained by combining (7) and (9) to eliminate the nonlinearity $f[e(i)]$:

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i + \bar{\mu}_{\Sigma}(i) \left(e_p^{\Sigma}(i) - e_a^{\Sigma}(i)\right) \mathbf{u}_i^T. \tag{11}$$

Both sides of (11) should have the same weighted energy, namely

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

which, after some straightforward manipulations, yields the desired energy relation

$$\|\tilde{\mathbf{w}}_{i+1}\|_{\Sigma}^2 + \bar{\mu}_{\Sigma}(i) \left|e_a^{\Sigma}(i)\right|^2 = \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 + \bar{\mu}_{\Sigma}(i) \left|e_p^{\Sigma}(i)\right|^2. \tag{12}$$

This relation shows how the weighted energies of the error quantities evolve in time. Observe that it is an exact relation and no approximations or assumptions are used to derive it. The result, for $\mathbf{\Sigma} = I$, has been originally developed in [29]–[32] in the context of robustness analysis of adaptive filters within a deterministic framework. It has since then been used in [25]–[28] as a convenient tool for studying the steady-state performance of adaptive filters within a stochastic framework as well. We will now show its relevance to the transient analysis of adaptive filters with error nonlinearities.

III. DYNAMICAL BEHAVIOR OF THE WEIGHT-ERROR VECTOR

Our first step is to examine how the energy relation (12) can be used to characterize the time-evolution of the weighted variance $E\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2$ for any Σ. Thus, consider (12) and replace the *a posteriori* error $e_p^{\Sigma}(i)$ by its equivalent expression (9). This yields

$$\|\tilde{\mathbf{w}}_{i+1}\|_{\Sigma}^2 = \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 - 2\mu e_a^{\Sigma}(i) f[e(i)] + \mu^2 \|\mathbf{u}_i\|_{\Sigma}^2 f^2[e(i)]$$

or, upon taking the expectation of both sides

$$E\left[\|\tilde{\mathbf{w}}_{i+1}\|_{\Sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right] - 2\mu \overbrace{E\left[e_a^{\Sigma}(i) f[e(i)]\right]}^{①} + \mu^2 \overbrace{E\left[\|\mathbf{u}_i\|_{\Sigma}^2 f^2[e(i)]\right]}^{②}. \tag{13}$$

Now, two expectations call for evaluation. This is facilitated by the following assumption on the noise sequence.

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

A. Evaluating Term ①

To evaluate the first expectation

$$① = E\left[e_a^{\Sigma}(i) f[e(i)]\right]$$

we will assume that the adaptive filter is long enough such that the random variables $e_a(i)$ and $e_a^{\Sigma}(i)$ can be assumed to be jointly Gaussian.

AG: For any constant matrix $\mathbf{\Sigma}$ and for all i, $e_a(i)$ and $e_a^{\Sigma}(i)$ are jointly Gaussian.

As mentioned in the introduction, this assumption is reasonable for longer filters by central limit arguments (see also the simulation results in Section V-A). A similar assumption was adopted in [4], [9], and [10], and its usefulness can be understood from the following result and from the subsequent discussion (see, e.g., [9] and [10]).

TABLE II
$h_G[\cdot]$ FOR THE ERROR NONLINEARITIES OF TABLE I ($\sigma_{e_a}^2 \triangleq E[e_a^2(i)]$)

ALGORITHM	$h_G[\sigma_{e_a}^2]$ ($v(i)$ Gaussian)	$h_G[\sigma_{e_a}^2]$ (general case)
LMS	1	1
LMF	$3(\sigma_{e_a}^2+\sigma_v^2)$	$3(\sigma_{e_a}^2+\sigma_v^2)$
LMF family	$\frac{(2k+2)!}{2^{k+1}(k+1)!}(\sigma_{e_a}^2+\sigma_v^2)^k$	$\sum_{j=0}^{k}\binom{2k+1}{j}\sigma_{e_a}^{2j}E\left[v^{2(k-j)}(i)\right]$
LMMN	$a+3b(\sigma_v^2+\sigma_{e_a}^2)$	$a+3b(\sigma_v^2+\sigma_{e_a}^2)$
Sign error	$\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{\sigma_{e_a}^2+\sigma_v^2}}$	$\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_{e_a}}E\left[e^{-\frac{v^2(i)}{2\sigma_{e_a}^2}}\right]$
Sat. nonlin.	$\frac{\sigma_{\text{sat}}}{\sqrt{\sigma_{e_a}^2+\sigma_v^2+\sigma_{\text{sat}}^2}}$	$\frac{\sigma_{\text{sat}}}{\sqrt{\sigma_{e_a}^2+\sigma_{\text{sat}}^2}}E\left[e^{-\frac{v^2(i)}{2(\sigma_{e_a}^2+\sigma_{\text{sat}}^2)}}\right]$

Lemma 1 (Price's Result): Let x and y be jointly Gaussian random variables that are independent from a third random variable z. Then

$$E\left[xf[y+z]\right]=\frac{E[xy]}{E[y^2]}E\left[yf[y+z]\right]. \qquad \diamondsuit$$

With Price's theorem at hand, we can use assumption AG together with the standing assumption on the noise AN and (8) to write ① as

$$\begin{aligned}E\left[e_a^{\Sigma}f[e(i)]\right]&=E\left[e_a^{\Sigma}f[e_a(i)+v(i)]\right]\\&=E\left[e_a^{\Sigma}(i)e_a(i)\right]\frac{E\left[e_a(i)f[e_a(i)+v(i)]\right]}{E[e_a^2(i)]}.\end{aligned} \tag{14}$$

At first glance, it would appear that we have replaced the expectation $E\left[e_a^{\Sigma}(i)f[e(i)]\right]$ with a similar one $E[e_a(i)f(e(i))]$. However, this second form is more tractable. Indeed, the expectation $E[e_a(i)f(e(i))]$ depends on $e_a(i)$ through *the second moment* $E[e_a^2(i)]$ *only*. This can be further seen by expanding it as (where we suppress the time index on the right-hand side)

$$E\left[e_a(i)f[e_a(i)+v(i)]\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e_af[e_a+v]\frac{1}{\sqrt{2\pi E[e_a^2]}}e^{-e_a^2/2E[e_a^2]}p_v(v)de_adv \tag{15}$$

where p_v is the pdf of the additive noise. The contribution of $e_a(i)$ to the result of the integration will depend solely on $E[e_a^2(i)]$. Therefore, the ratio $E\left[e_a(i)f[e(i)]\right]/E[e_a^2(i)]$, which appears in (14), is a function of $E[e_a^2(i)]$. This fact motivates the following definition:[2]

$$h_G\left[E[e_a^2(i)]\right]\triangleq\frac{E\left[e_a(i)f[e(i)]\right]}{E[e_a^2(i)]}. \tag{16}$$

For future reference, h_G is evaluated for the algorithms of Table I, and the results are shown in Table II (for general noise distribution and for the Gaussian noise case as well). Combining (14) and (16) yields

$$E\left[e_a^{\Sigma}(i)f(e(i))\right]=E\left[e_a^{\Sigma}(i)e_a(i)\right]h_G\left[E[e_a^2(i)]\right]. \tag{17}$$

[2]The Gaussianity assumption AG is the main assumption leading to the defining expression (16) for h_G, hence, the subscript G. The subscript U for h_U, which is defined later in (20), is similarly motivated.

We finally use the polarization property (1) to write the first expectation in (17) as a weighted-norm of $\tilde{\mathbf{w}}_i$, yielding

$$① = E\left[e_a^{\Sigma}(i)f(e(i))\right]=E\left[\|\tilde{\mathbf{w}}_i\|^2_{\Sigma u_i^Tu_i}\right]h_G\left[E[e_a^2(i)]\right]. \tag{18}$$

B. Evaluating Term ②

We turn our attention now to the second expectation in (13) $② = E\left[\|\mathbf{u}_i\|^2_{\Sigma}f^2[e(i)]\right]$, which is easier to handle. The long filter assumption is also useful here.

AU: The adaptive filter is long enough such that $\mu\|\mathbf{u}_i\|^2_{\Sigma}$ and $f^2[e(i)]$ are uncorrelated.

The unweighted version of this assumption was used in [25]–[27]. It becomes more realistic as the filter gets longer. The assumption enables us to split the expectation ② as

$$E\left[\|\mathbf{u}_i\|^2_{\Sigma}f^2[e(i)]\right]=E\left[\|\mathbf{u}_i\|^2_{\Sigma}\right]E\left[f^2[e(i)]\right]. \tag{19}$$

Moreover, since $e_a(i)$ is Gaussian and independent of the noise, we can show [as in (15)] that $E\left[f^2[e(i)]\right]$ depends on $e_a(i)$ through its *second moment only*. This prompts us to define

$$h_U\left[E[e_a^2(i)]\right]\triangleq E\left[f^2[e(i)]\right] \tag{20}$$

which together with (19) yields

$$E\left[\|\mathbf{u}_i\|^2_{\Sigma}f^2[e(i)]\right]=E\left[\|\mathbf{u}_i\|^2_{\Sigma}\right]h_U\left[E\left[e_a^2(i)\right]\right]. \tag{21}$$

The function h_U is evaluated for the algorithms of Table I, and the results are shown in Table III for general noise and for the Gaussian noise special case (the last entry in the table is derived in the Appendix).

C. Weight-Error Recursion

By substituting (18) and (21) into (13), we obtain

$$\begin{aligned}E\left[\|\tilde{\mathbf{w}}_{i+1}\|^2_{\Sigma}\right]&=E\left[\|\tilde{\mathbf{w}}_i\|^2_{\Sigma}\right]-2\mu h_G\left[E[e_a^2(i)]\right]\\&\quad\times E\left[\|\tilde{\mathbf{w}}_i\|^2_{\Sigma u_i^Tu_i}\right]+\mu^2E\left[\|\mathbf{u}_i\|^2_{\Sigma}\right]h_U\left[E[e_a^2(i)]\right].\end{aligned}$$

Upon replacing the mean-square error $E\left[e_a^2(i)\right]$ with the equivalent expression $E\left[\|\tilde{\mathbf{w}}_i\|^2_{u_i^Tu_i}\right]$, the recursion takes the more homogeneous form shown in the statement below.

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

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

TABLE III
$h_U[\cdot]$ FOR THE ERROR NONLINEARITIES OF TABLE I ($\sigma_{e_a}^2 \triangleq E[e_a^2(i)]$)

ALGORITHM	$h_U[\sigma_{e_a}^2]$ ($v(i)$ Gaussian)	$h_U[\sigma_{e_a}^2]$ (general case)
LMS	$\sigma_{e_a}^2 + \sigma_v^2$	$\sigma_{e_a}^2 + \sigma_v^2$
LMF	$15(\sigma_{e_a}^2 + \sigma_v^2)^3$	$15\sigma_{e_a}^6 + 45\sigma_{e_a}^4\sigma_v^2 + 15\sigma_{e_a}^2 E\left[v^4(i)\right] + E\left[v^6(i)\right]$
LMF family	$\frac{(4k+2)!}{2^{2k+1}(2k+1)!}(\sigma_{e_a}^2 + \sigma_v^2)^{2k+1}$	$\sum_{j=0}^{2k+1} \binom{4k+2}{2j} \frac{(2j)!}{2^j j!}\sigma_{e_a}^{2j} E\left[v^{2(2k-j+1)}(i)\right]$
LMMN	$a^2(\sigma_{e_a}^2 + \sigma_v^2) + 6ab(\sigma_{e_a}^2 + \sigma_v^2)^2 + 15b^2(\sigma_{e_a}^2 + \sigma_v^2)^3$	$15b^2\sigma_{e_a}^6 + (45b^2\sigma_v^2 + 6ab)\sigma_{e_a}^4 + (15b^2 E\left[v^4(i)\right] + 12ab\sigma_v^2 + a^2)\sigma_{e_a}^2 + E\left[(bv^2(i) + a)^2 v^2(i)\right]$
Sign error	1	1
Sat. nonlin.	$\sigma_{\text{sat}}^2 \sin^{-1}\left(\frac{\sigma_{e_a}^2 + \sigma_v^2}{\sigma_{e_a}^2 + \sigma_v^2 + \sigma_{\text{sat}}^2}\right)$	$2\pi\sigma_{\text{sat}}^2\left(\frac{1}{4} - \frac{1}{\pi}\int_{\pi/4}^{\pi/2}\sqrt{\frac{\sigma_{\text{sat}}^2\sin^2(\theta)}{\sigma_{e_a}^2 + \sigma_{\text{sat}}^2\sin^2(\theta)}} E[e^{-\frac{v^2(i)}{2(\sigma_{e_a}^2 + \sigma_{\text{sat}}^2\sin^2(\theta))}}]\right)$

where $e(i) = d(i) - \mathbf{u}_i\mathbf{w}_i$ and $d(i) = \mathbf{u}_i\mathbf{w}^o + v(i)$. Assume the noise sequence $v(i)$ is iid and independent of $\mathbf{u}_i$ and that the filter is long enough so that $e_a(i)$ and $e_a^{\Sigma}(i)$ are jointly Gaussian and that $\mu\|\mathbf{u}_i\|_{\Sigma}^2$ and $f^2[e(i)]$ are uncorrelated. Then, the following recursion holds for the weighted weight-error variance $E\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2$:

$$E\left[\|\tilde{\mathbf{w}}_{i+1}\|_{\Sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right] - 2\mu h_G\left[E\left[\|\tilde{\mathbf{w}}_i\|_{u_i^T u_i}^2\right]\right] \times E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma u_i^T u_i}^2\right] + \mu^2 E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right] h_U\left[E\left[\|\tilde{\mathbf{w}}_i\|_{u_i^T u_i}^2\right]\right] \quad (22)$$

where the functions $h_G[\cdot]$ and $h_U[\cdot]$ are defined by

$$h_U = E\left[f^2[e(i)]\right], \quad h_G = \frac{E[e_a(i)f[e(i)]]}{E[e_a^2(i)]}. \qquad \diamondsuit$$

Remarks:

1) What we have achieved so far is to transform recursion (13) into (22), which depends on various weighted Euclidean norms of the weight-error vector, thanks to assumptions AG and AU.
2) Assumptions AG and AU eventually get translated into some mixing conditions on the signal statistics. In particular, the Gaussian assumption AG on $e_a(i) = \mathbf{u}_i\tilde{\mathbf{w}}_i$ requires that the process of individual summands $u_i(l)w_i(l)$ are mixed [35, Th. 27.4]. Similarly, the AU assumption is justified by the law of large numbers, which in turn requires that the input $\mathbf{u}_i$ is mixed [37].
3) The independence assumption on the noise AN is equally essential in developing (18) and (21) and, hence, (22). It is a reasonable assumption that allows us to express the expectations in (13) in terms of the weight-error energy.
4) Recursion (22) as it stands is difficult to propagate in time. The reason is that the recursion is not self-contained as the right-hand side is dependent on $E\|\tilde{\mathbf{w}}_i\|_{\Sigma u_i^T u_i}^2$ and $E\|\tilde{\mathbf{w}}_i\|_{u_i^T u_i}^2$, in addition to $E\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2$.
5) Note that only a weak form of the independence assumption, namely AU, has been used so far. Contrast this with the standard (stronger)[3] independence assumption:

   AI: The sequence $\mathbf{u}_i$ is zero-mean, iid, with autocorrelation matrix $\mathbf{R} = E\left[\mathbf{u}_i^T\mathbf{u}_i\right]$.

   In this case, recursion (22) reduces to the following.

[3]For example, when the input is of constant modulus, assumption AU is true, whereas AI is not.

Corollary 1 (Energy Recursion With Independence): Consider the same setting of Theorem 1. If, in addition, the sequence $\mathbf{u}_i$ is zero-mean, iid, and has covariance matrix $\mathbf{R}$, then (22) becomes

$$E\left[\|\tilde{\mathbf{w}}_{i+1}\|_{\Sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right] - 2\mu h_G\left[E[\|\tilde{\mathbf{w}}_i\|_R^2]\right] E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma R}^2\right] + \mu^2 E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right] h_U\left[E[\|\tilde{\mathbf{w}}_i\|_R^2]\right]. \quad (23)$$

$\diamondsuit$

D. Constructing the Learning Curves

The learning curve of the filter refers to the time-evolution of the variance $Ee^2(i)$; its steady-state value is the mean-square error (MSE). Clearly, in view of (8), we have that

$$Ee^2(i) = Ee_a^2(i) + \sigma_v^2$$

so that studying the evolution of $Ee^2(i)$ is equivalent to studying the evolution of $Ee_a^2(i)$; the steady-state value of the latter is called the excess mean-square error (EMSE).

Now, under the independence assumption, we have

$$Ee_a^2(i) = E|\mathbf{u}_i\tilde{\mathbf{w}}_i|^2 = E\left[\|\tilde{\mathbf{w}}_i\|_R^2\right].$$

This suggests that the learning curve can be evaluated by computing $E[\|\tilde{\mathbf{w}}_i\|_R^2]$ for each i. This task can be accomplished recursively from (23) by essentially choosing $\mathbf{\Sigma} = \mathbf{R}$, as we now verify.

1) Case of White Regression Data: Consider first the case of white input data for which $\mathbf{R} = \sigma_u^2\mathbf{I}$ so that $E[e_a^2(i)] = \sigma_u^2 E\|\tilde{\mathbf{w}}_i\|^2$. Restricting the input in this manner is a common practice in the literature (e.g., as in [4], [10], [12], [36], and [38]).

Thus, setting $\mathbf{\Sigma} = \mathbf{I}$ in (23), we get

$$E\left[\|\tilde{\mathbf{w}}_{i+1}\|^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|^2\right] - 2\mu\sigma_u^2 h_G\left[\sigma_u^2 E\left[\|\tilde{\mathbf{w}}_i\|^2\right]\right] \times E\left[\|\tilde{\mathbf{w}}_i\|^2\right] + \mu^2\sigma_u^2 M h_U\left[\sigma_u^2 E\left[\|\tilde{\mathbf{w}}_i\|^2\right]\right]. \quad (24)$$

Note that the right-hand side now depends on $E\|\tilde{\mathbf{w}}_i\|^2$ only, and (24) can be propagated in time. We have thus obtained a recursion for the evolution of the variance $E\|\tilde{\mathbf{w}}_i\|^2$ for adaptive filters with error nonlinearities and white input regression data.

2) Case of Correlated Regression Data: The result (23), however, allows us to evaluate the time evolution of $E\|\tilde{\mathbf{w}}_i\|^2$ and $Ee_a^2(i)$, even without the whiteness assumption on the regression data (i.e., for general matrices $\mathbf{R}$). The key idea is to take advantage of the free parameter $\boldsymbol{\Sigma}$. Let us, in particular, write (23) for the choices $\boldsymbol{\Sigma} = \mathbf{I}, \mathbf{R}, \ldots, \mathbf{R}^{M-1}$ (the arguments of the functions h_G and h_U remain the same (i.e., $E\left[\|\tilde{\mathbf{w}}_i\|_R^2\right]$), regardless of the choice of $\boldsymbol{\Sigma}$ and are therefore suppressed for convenience of notation): See equation (25) at the bottom of the page. The problem now is that the left-hand side of (25) is always one variable short of the number of variables on the right-hand side. Fortunately, we do not have to continue in this manner indefinitely since the additional variable $E\left[\|\tilde{\mathbf{w}}_i\|_{R^M}^2\right]$ can be expressed in terms of the "lower order" variables. Using the Cayley–Hamilton theorem, we have

$$\mathbf{R}^M = -p_0\mathbf{I} - p_1\mathbf{R} - \cdots - p_{M-1}\mathbf{R}^{M-1}$$

where

$$\begin{aligned} p(x) &\stackrel{\Delta}{=} \det(x\mathbf{I} - \mathbf{R}) \\ &= p_0 + p_1 x + \cdots + p_{M-1}x^{M-1} + x^M \end{aligned}$$

is the characteristic polynomial of $\mathbf{R}$. This induces the desired relation

$$\|\tilde{\mathbf{w}}_i\|_{R^M}^2 = -p_0\|\tilde{\mathbf{w}}_i\|^2 - p_1\|\tilde{\mathbf{w}}_i\|_R^2 - \cdots - p_{M-1}\|\tilde{\mathbf{w}}_i\|_{R^{M-1}}^2$$

and enables us to rewrite the last equation in (25) as

$$\begin{aligned} E\left[\|\tilde{\mathbf{w}}_{i+1}\|_{R^{M-1}}^2\right] =& E\left[\|\tilde{\mathbf{w}}_i\|_{R^{M-1}}^2\right] \\ &+ 2\mu\left(p_0\|\tilde{\mathbf{w}}_i\|^2 + p_1\|\tilde{\mathbf{w}}_i\|_R^2 \right. \\ &\left. + \cdots + p_{M-1}\|\tilde{\mathbf{w}}_i\|_{R^{M-1}}^2\right) h_G \\ &+ \mu^2 E\left[\|\mathbf{u}_i\|_{R^{M-1}}^2\right] h_U. \end{aligned}$$

The system (25) now becomes truly self-contained and, as such, can be put into the state-space form shown in the following theorem.

Theorem 2 (Transient Behavior With Independence): Consider an adaptive filter of the form

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

where $e(i) = d(i) - \mathbf{u}_i w_i$ and $d(i) = \mathbf{u}_i\mathbf{w}^o + v(i)$. Assume that $\{v(i), \mathbf{u}_i\}$ are iid and mutually independent, that the filter is long enough so that $e_a(i)$ and $e_a^{\Sigma}(i)$ are jointly Gaussian, and that $\mu\|\mathbf{u}_i\|_{\Sigma}^2$ and $f^2[e(i)]$ are uncorrelated. Then, regardless of the statistics of the regression data, the transient behavior of the filter is characterized by the state-space recursion

$$\mathcal{W}_{i+1} = \mathcal{A}\mathcal{W}_i + \mu^2\mathcal{Y} \tag{26}$$

where the state-vector $\mathcal{W}_i$ and the input vector $\mathcal{Y}$ are defined by

$$\mathcal{W}_i = \begin{bmatrix} E\left[\|\tilde{\mathbf{w}}_i\|^2\right] \\ E\left[\|\tilde{\mathbf{w}}_i\|_R^2\right] \\ \vdots \\ E\left[\|\tilde{\mathbf{w}}_i\|_{R^{M-1}}^2\right] \end{bmatrix}, \quad \mathcal{Y} = h_U \cdot \begin{bmatrix} E\left[\|\mathbf{u}_i\|^2\right] \\ E\left[\|\mathbf{u}_i\|_R^2\right] \\ \vdots \\ E\left[\|\mathbf{u}_i\|_{R^{M-1}}^2\right] \end{bmatrix}$$

and the coefficient matrix $\mathcal{A}$ is given by the equation at the bottom of the page in terms of $\{h_G, h_U\}$ and the $\{p_i\}$. ◇

Remarks:

1) Since $\mathcal{A}$ and $\mathcal{Y}$ depend on $\{h_U, h_G\}$, they are also functions of $E\left[\|\tilde{\mathbf{w}}_i\|_R^2\right]$ and, hence, of the state vector $\mathcal{W}_i$. Thus, the state-space model (26) is generally nonlinear, yet time invariant.
2) Stability and steady-state analysis of the adaptive filter can now be characterized by studying the properties of the state-space model (26).
3) The top entry of the state-vector $\mathcal{W}_i$ characterizes the evolution of $E\|\tilde{\mathbf{w}}_i\|^2$ (mean-square deviation curve), whereas the second entry of $\mathcal{W}_i$ characterizes the evolution of $Ee_a^2(i)$ (learning curve).

IV. Steady-State Analysis

Now that the transient behavior of adaptive filters of the class (2) has been characterized, we move on to show how the results so far can be used to evaluate the steady-state performance of this same class of filters. Actually, the discussion that follows does not require the independence assumption AI any longer.

We refer again to the averaged energy relation (22), which we rewrite using (5) and (6) as

$$\begin{aligned} E\left[\|\tilde{\mathbf{w}}_{i+1}\|_{\Sigma}^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right] &- 2\mu h_G\left[E[e_a^2(i)]\right] E\left[e_a^{\Sigma}(i)e_a(i)\right] \\ &+ \mu^2 E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right] h_U\left[E[e_a^2(i)]\right]. \end{aligned} \tag{27}$$

Assuming that the weight-error vector reaches a steady-state mean-square value, i.e.,

$$\lim_{i\to\infty} E\left[\|\tilde{\mathbf{w}}_{i+1}\|_{\Sigma}^2\right] = \lim_{i\to\infty} E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2\right]$$

$$\begin{cases} E\left[\|\tilde{\mathbf{w}}_{i+1}\|_I^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|_I^2\right] - 2\mu h_G E\left[\|\tilde{\mathbf{w}}_i\|_R^2\right] + \mu^2 E\left[\|\mathbf{u}_i\|_I^2\right] h_U \\ E\left[\|\tilde{\mathbf{w}}_{i+1}\|_R^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|_R^2\right] - 2\mu h_G E\left[\|\tilde{\mathbf{w}}_i\|_{R^2}^2\right] + \mu^2 E\left[\|\mathbf{u}_i\|_R^2\right] h_U \\ \quad\vdots \\ E\left[\|\tilde{\mathbf{w}}_{i+1}\|_{R^{M-1}}^2\right] = E\left[\|\tilde{\mathbf{w}}_i\|_{R^{M-1}}^2\right] - 2\mu h_G E\left[\|\tilde{\mathbf{w}}_i\|_{R^M}^2\right] + \mu^2 E\left[\|\mathbf{u}_i\|_{R^{M-1}}^2\right] h_U \end{cases} \tag{25}$$

$$\mathcal{A} = \begin{bmatrix} 1 & -2\mu h_G & 0 & \cdots & 0 & 0 \\ 0 & 1 & -2\mu h_G & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -2\mu h_G \\ 2\mu p_0 h_G & 2\mu p_1 h_G & 2\mu p_2 h_G & \cdots & 2\mu p_{M-2} h_G & 1 + 2\mu p_{M-1} h_G \end{bmatrix}$$

the energy relation (27) becomes, in the limit

$$\lim_{i\to\infty} h_G\left[E[e_a^2(i)]\right] \lim_{i\to\infty} E\left[e_a^{\Sigma}(i)e_a(i)\right] = \frac{\mu}{2}E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right] \lim_{i\to\infty} h_U\left[E[e_a^2(i)]\right]$$

or

$$\lim_{i\to\infty} E[e_a^{\Sigma}(i)e_a(i)] = \frac{\mu}{2}E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right] \frac{\lim_{i\to\infty} h_U\left[E[e_a^2(i)]\right]}{\lim_{i\to\infty} h_G[E[e_a^2(i)]]}. \tag{28}$$

Now, let ζ denote the EMSE, i.e.,

$$\zeta = \lim_{i\to\infty} E[e_a^2(i)] \tag{29}$$

which, assuming the filter is mean-square stable, exists and is finite. Then

$$\lim_{i\to\infty} h_G[E[e_a^2(i)]] = h_G[\zeta] \text{ and } \lim_{i\to\infty} h_U[E[e_a^2(i)]] = h_U[\zeta]$$

and accordingly, (28) can be written more compactly, as shown below.

Theorem 3 (Steady-State Performance): Consider the same setting of Theorem 1. Then, assuming a mean-square stable filter with EMSE denoted by ζ, the following equality holds:

$$\lim_{i\to\infty} E\left[e_a^{\Sigma}(i)e_a(i)\right] = \frac{\mu}{2}E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right] \frac{h_U[\zeta]}{h_G[\zeta]} \tag{30}$$

◊

The above relation has been derived for general memoryless error nonlinearities. We now show how it can be used to evaluate various steady-state quantities such as the excess mean-square error and the mean-square deviation.

A. Excess Mean-Square Error

To calculate the excess mean-square error, we employ (30) with $\mathbf{\Sigma}$ set to the identity matrix

$$\lim_{i\to\infty} E[e_a^2(i)] = \frac{\mu}{2}E\left[\|\mathbf{u}_i\|^2\right] \frac{h_U[\zeta]}{h_G[\zeta]} = \frac{\mu}{2}\mathrm{Tr}(\mathbf{R}) \frac{h_U[\zeta]}{h_G[\zeta]}$$

or since $\zeta = \lim_{i\to\infty} E[e_a^2(i)]$, we arrive at the following statement.

Corollary 2 (EMSE): Consider the same setting of Theorem 1. Then, the EMSE is a positive solution of the equation

$$\zeta = \frac{\mu}{2}\mathrm{Tr}(\mathbf{R}) \frac{h_U[\zeta]}{h_G[\zeta]} \tag{31}$$

i.e., the EMSE is a fixed point of the function $(\mu/2)\mathrm{Tr}(\mathbf{R})\, h_U[\zeta]/h_G[\zeta]$. ◊

Relation (31) is a generalization of the results of [26] to general error functions f. In the following, we show how (31) specializes for some nonlinearities.

1) LMS Algorithm: In the LMS case, (31) reads

$$\zeta = \frac{\mu}{2}\mathrm{Tr}(\mathbf{R})\left(\zeta + \sigma_v^2\right)$$

or, upon solving for ζ, we obtain the well-known result [20]:

$$\zeta = \frac{\mu\sigma_v^2\mathrm{Tr}(\mathbf{R})}{2 - \mu\mathrm{Tr}(\mathbf{R})}$$

TABLE IV
EMSE FOR THE SIGN ALGORITHM FOR VARIOUS NOISE STATISTICS.

Noise	EMSE
Gaussian [8, 26]	$\zeta = \alpha\,\frac{\alpha + \sqrt{\alpha^2 + 4\sigma_v^2}}{2}, \quad \alpha = \mu\sqrt{\frac{\pi}{8}}\mathrm{Tr}(\mathbf{R})$
Binary [12]	$\zeta = \alpha^2 e^{\frac{\sigma_v^2}{\zeta}}, \quad \alpha = \mu\sqrt{\frac{\pi}{8}}\mathrm{Tr}(\mathbf{R})$
Uniform [12]	$\zeta = \frac{\mu}{2}\frac{\sqrt{3\sigma_v^2}}{\mathrm{erf}\left(\sqrt{\frac{3\sigma_v^2}{2\zeta}}\right)}\mathrm{Tr}(\mathbf{R})$

2) Sign Algorithm: We start from (31) again. With the aid of Tables II and III, we see that

$$\zeta = \frac{\mu}{2}\mathrm{Tr}(\mathbf{R}) \frac{h_U[\zeta]}{h_G[\zeta]} = \mu\sqrt{\frac{\pi}{8}}\mathrm{Tr}(\mathbf{R}) \frac{\sqrt{\zeta}}{E\left[e^{-v^2(i)/2\zeta}\right]}. \tag{32}$$

It is worth noting in the sign algorithm case that assumption AU is not needed. In other words, we only need the Gaussian assumption AG to establish (32). This was the same conclusion arrived at in [26], but the study there was limited to the Gaussian noise case. Further progress is pending the evaluation of $E\left[e^{-v^2(i)/2\zeta}\right]$, which calls for specifying the noise statistics. Our findings are summarized in the Table IV. In particular, we arrive at the same EMSE expressions of [12] derived there under the independence assumption for iid input. In the second line of Table IV, the noise is assumed to be equal to $\pm\sigma_v$ with probability 1/2, whereas in the third line, the noise is assumed to be uniformly distributed inside the interval $(-\sqrt{3}\sigma_v, \sqrt{3}\sigma_v)$. The erf function is defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt.$$

3) Error-Saturation Algorithm: Consider the saturation nonlinearity in Table I. The associated expectations h_G and h_U are relatively easy to establish in the Gaussian noise case (see Tables II and III)

$$h_G[\zeta] = \frac{\sigma_{\mathrm{sat}}}{\sqrt{\zeta + \sigma_v^2 + \sigma_{\mathrm{sat}}^2}}$$
$$h_U[\zeta] = \sigma_{\mathrm{sat}}^2 \sin^{-1}\left(\frac{\zeta + \sigma_v^2}{\zeta + \sigma_v^2 + \sigma_{\mathrm{sat}}^2}\right)$$

which upon substitution in (31) yields the following relation for the EMSE:

$$\frac{\zeta}{\sqrt{\zeta + \sigma_v^2 + \sigma_{\mathrm{sat}}^2}} = \sigma_{\mathrm{sat}}\frac{\mu}{2}\mathrm{Tr}(\mathbf{R})\sin^{-1}\left(\frac{\zeta + \sigma_v^2}{\zeta + \sigma_v^2 + \sigma_{\mathrm{sat}}^2}\right).$$

This is the same result arrived at in [10] under the independence assumption for iid input.

In the general noise case, we have

$$h_G[\zeta] = \frac{\sigma_{\mathrm{sat}}}{\sqrt{\zeta + \sigma_{\mathrm{sat}}^2}} E\left[e^{-v^2(i)/2(\zeta + \sigma_{\mathrm{sat}}^2)}\right] \tag{33}$$

which encompasses the binary noise case considered in [10] as a special case. Evaluating h_U is more difficult; this was attempted in [10], and the argument led to a complicated expression involving double integrals and infinite limits. We arrive in the Appendix at the expression

$$h_U[\zeta] = 2\pi\sigma_{\rm sat}^2\left(\frac{1}{4}-\frac{1}{\pi}\right.\\ \left.\times\int_{\pi/4}^{\pi/2}\sqrt{\frac{\sigma_{\rm sat}^2\sin^2(\theta)}{\sigma_{e_a}^2+\sigma_{\rm sat}^2\sin^2(\theta)}}E[e^{-v^2(i)/2(\sigma_{e_a}^2+\sigma_{\rm sat}^2\sin^2(\theta))}]\right) \tag{34}$$

by relying on a convenient expression for the error function introduced [40]. Upon substituting (33) and (34) into (31), we obtain

$$\frac{\zeta}{\sqrt{\zeta+\sigma_{\rm sat}^2}}E\left[e^{-v^2(i)/2(\zeta+\sigma_{\rm sat}^2)}\right] = \mu\pi\sigma_{\rm sat}{\rm Tr}(\mathbf{R})\left(\frac{1}{4}-\frac{1}{\pi}\right.\\ \left.\times\int_{\pi/4}^{\pi/2}\sqrt{\frac{\sigma_{\rm sat}^2\sin^2(\theta)}{\sigma_{e_a}^2+\sigma_{\rm sat}^2\sin^2(\theta)}}E[e^{-v^2(i)/2(\sigma_{e_a}^2+\sigma_{\rm sat}^2\sin^2(\theta))}]\right)$$

which can be numerically solved for ζ, which is the EMSE.

4) LMF Algorithm: For the LMF algorithm, and with the aid of Tables I and II, (31) takes the form

$$\zeta = \frac{\mu}{6}\frac{15\zeta^3+45\sigma_v^2\zeta^2+15m_{v,4}\zeta+m_{v,6}}{\zeta+\sigma_v^2}{\rm Tr}(\mathbf{R}) \tag{35}$$

where $m_{v,4}$ and $m_{v,6}$ denote the fourth and sixth moments of $v(i)$. Finding the EMSE is thus equivalent to finding the roots of a third-order equation, which can be done numerically. We can avoid this in the Gaussian case and obtain a *closed* formula for the EMSE.

Gaussian Noise: In the Gaussian noise case, (35) simplifies to

$$\zeta = \frac{5\mu}{2}\frac{(\zeta+\sigma_v^2)^3}{\zeta+\sigma_v^2}{\rm Tr}(\mathbf{R}) = \frac{\alpha}{2}(\zeta+\sigma_v^2)^2$$

where $\alpha = 5\mu{\rm Tr}(\mathbf{R})$. This is a quadratic equation in ζ with *two* positive roots

$$\zeta = \frac{(1-\alpha\sigma_v^2)\pm\sqrt{1-2\alpha\sigma_v^2}}{\alpha}. \tag{36}$$

Simulations show that only the smaller root is meaningful.

It appears that calculating the steady-state error for super nonlinearities (e.g., the LMF algorithm, the LMF family, and the LMMN algorithm) has always involved some form of linearization (e.g., [5], [13], [26], [36], [38], [39]). The LMF derivation above demonstrates how the EMSE can be obtained for such algorithms without having to employ linearization arguments.

B. Mean-Square Deviation

The mean-square deviation (MSD), which is defined as

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

can be related to the EMSE by invoking the independence assumption in the limit. More specifically, by combining (30) and (31), we obtain

$$\lim_{i\to\infty} E\left[e_a^{\Sigma}(i)e_a(i)\right] = \zeta\cdot\frac{E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right]}{E\left[\|\mathbf{u}_i\|^2\right]}.$$

Assuming AI holds in the limit, we have

$$\lim_{i\to\infty} E\left[e_a^{\Sigma}(i)e_a(i)\right] = \lim_{i\to\infty} E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma R}^2\right]$$

so that

$$\lim_{i\to\infty} E\left[\|\tilde{\mathbf{w}}_i\|_{\Sigma R}^2\right] = \zeta\cdot\frac{E\left[\|\mathbf{u}_i\|_{\Sigma}^2\right]}{E\left[\|\mathbf{u}_i\|^2\right]}. \tag{37}$$

Since we are interested in $E\|\tilde{\mathbf{w}}_i\|^2$, we choose $\mathbf{\Sigma}$ in (37) as $\mathbf{R}^{-1}$, which leads us to the following conclusion.

Corollary 3 (MSD): Consider the same setting of Theorem 1, and assume, in addition, that the sequence $\mathbf{u}_i$ is zero-mean iid. Then, the MSD is given by

$$\text{MSD} = \frac{M\zeta}{\text{Tr}(\mathbf{R})}$$

where ζ denotes the filter EMSE. ◊

Other steady-state measures can be similarly evaluated. Thus, for any symmetric matrix $\mathbf{A}$, we have

$$\lim_{i\to\infty} E\left[\|\tilde{\mathbf{w}}_i\|_A^2\right] = \frac{E\left[\|\mathbf{u}_i\|_{AR^{-1}}^2\right]}{E\left[\|\mathbf{u}_i\|^2\right]}\text{EMSE}$$

V. Simulations

Throughout this section, the system to be identified is an FIR channel of length 16. The input $u(i)$ is generated by passing an iid (uniform or Gaussian) process $x(i)$ through a first-order model

$$u(i) = au(i-1)+x(i). \tag{38}$$

By varying the value of a, we obtain processes $u(i)$ of different colors. Here, we set $a=0.3$. The output is contaminated by an iid (uniform or Gaussian) additive noise at an SNR level of 10 dB.

A. Testing the Gaussianity of $e_a(i)$

We start by running a simulation to test the Gaussian assumption AG on $e_a(i)$ for the sign algorithm. We choose the sign algorithm because it was argued in [41] that $e_a(i)$ can never be Gaussian under the independence assumption. The signals involved are chosen to be non-Gaussian. Thus, the input is generated by (38), and the processes x and v are both taken to be iid uniform.

The Gaussian hypothesis is tested by running the adaptive algorithm 1000 times and plotting the histogram of $e_a(i)$ at the equispaced instants $i = 0, 200, \ldots, 1000$. The histograms, which are depicted in Fig. 1, suggest that the Gaussian assumption on $e_a(i)$ is still a reasonable approximation for practical purposes. The only exception is the histogram for $e_a(0)$, which is almost uniformly distributed (as it should be since $e_a(0)$ is generated

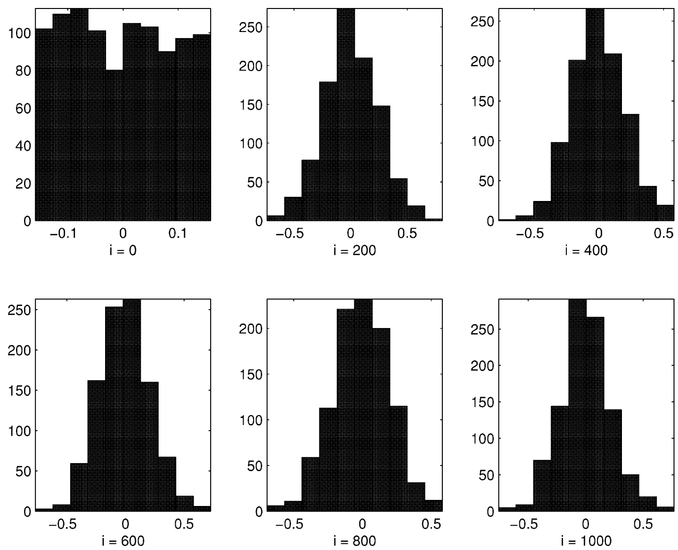


Fig. 1. Histogram of $e_a(i)$ for the sign algorithm at different time instants (uniform noise, uniform input with $a = 0.3$, $\mu = 0.01$, SNR $= 10$ dB).

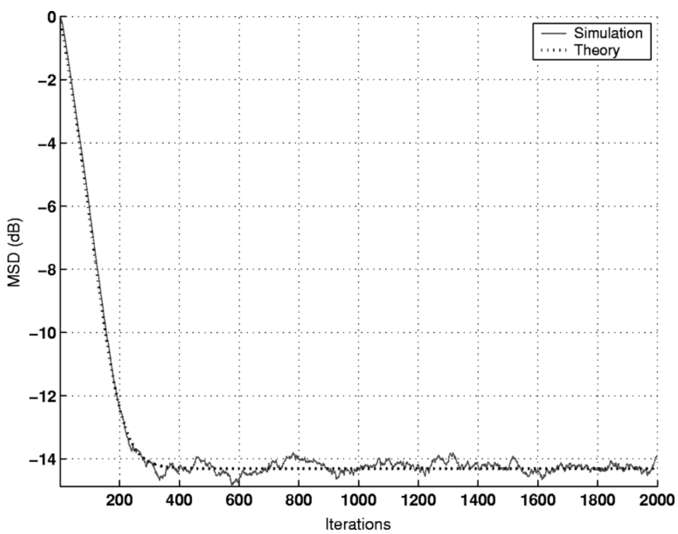


Fig. 2. Theoretical and simulated learning curves for the sign algorithm (Gaussian noise, Gaussian input with $a = .1$, $\mu = .01$, SNR $= 10$ dB).

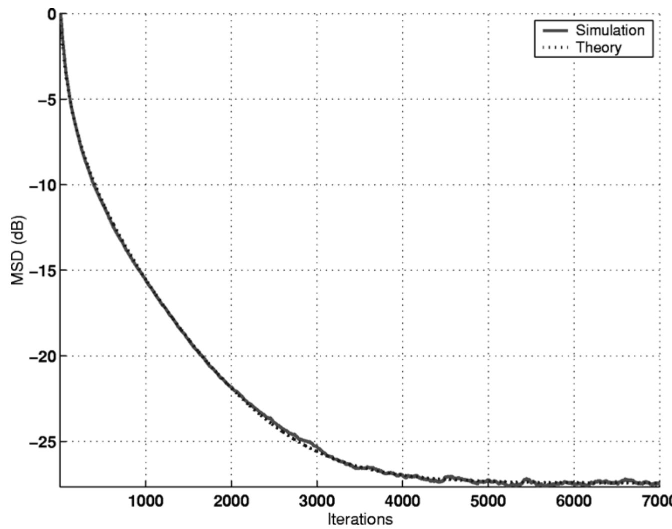


Fig. 3. Theoretical and simulated learning curves for the LMF algorithm (Gaussian noise, Gaussian input with $a = 0.1$, $\mu = .0044$, SNR $= 10$ dB).

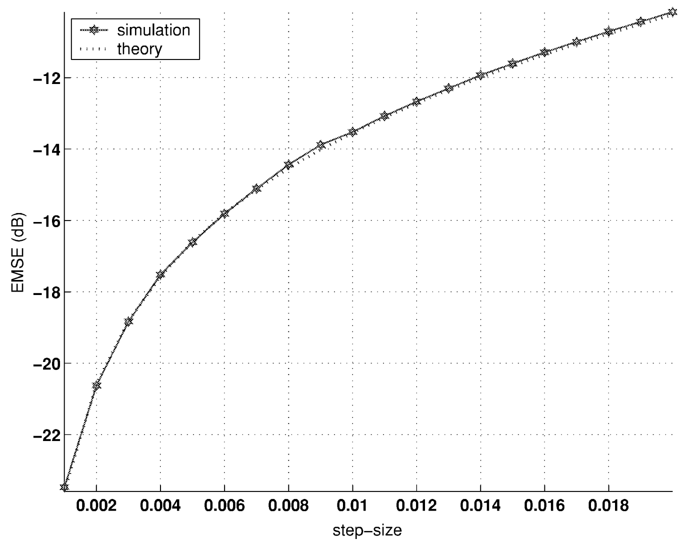


Fig. 4. Theoretical and simulated EMSE versus μ for the sign algorithm (uniform noise, Gaussian input with $a = 0.3$, SNR $= 10$ dB.).

by one data point for which the central limit theorem does not apply).

B. Learning Curves

Next, we study the match between the theoretical (Theorem 2) and simulated learning curves. We test the match for the sign and LMF algorithms. In both cases, the input is assumed to be a Gaussian correlated process with $a = 0.3$. As depicted in Figs. 2 and 3, the experimental and theoretical learning curves agree very well. This agreement occurs despite the fact that large values of the step size are used.

C. Steady-State Behavior

Here, we simulate the steady-state behavior of the sign and LMF algorithms and compare the results to theory. We test the sign algorithm for correlated uniform input (with $a = 0.3$) and uniform noise. Fig. 4 shows an excellent match between the EMSE generated by simulation and that predicted by theory (see Table IV).

The LMF is tested for correlated Gaussian input (with $a = 0.3$) and Gaussian noise. Fig. 5 demonstrates the excellent match between simulation and theoretical values [predicted by (36)]. In this figure, we also plot the value of the steady-state error as predicted by the expression in [26] for small and large μ, which eventually employ some sort of linearization. The predictions of (36) are more accurate.

VI. Concluding Remarks

In this paper, we employed energy-conservation arguments to study the transient performance of adaptive filters with error nonlinearities. The arguments of this work, as well as in [25] and [26], demonstrate the convenience of working with the energy relation. In developing the energy relation, we basically push the algebraic operations to the limit before we undertake any

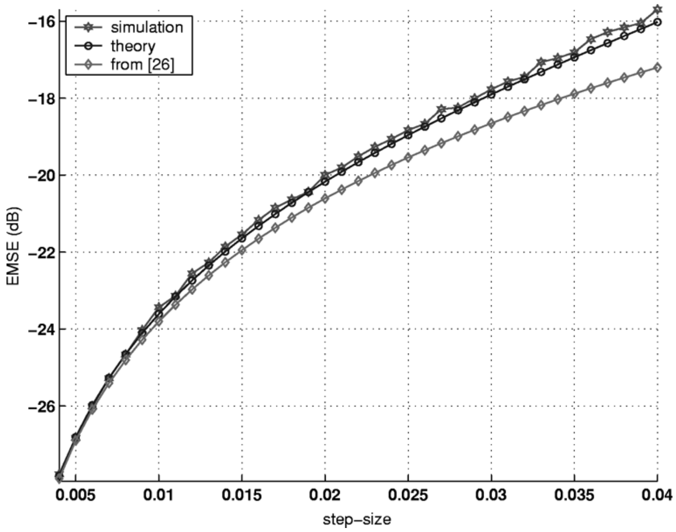


Fig. 5. Theoretical and simulated EMSE versus μ for the LMF algorithm (Gaussian noise, Gaussian input with $a = 0.3$, SNR $= 10$ dB).

averaging operation. We do so because our ability to maneuver algebraically under the expectation operator is usually limited.

The main contributions of this part are Theorems 1–3; the first relates to the energy conservation result, the second relates to the learning curve behavior, and the third relates to a nonlinear equation for EMSE calculation.

APPENDIX
EVALUATING h_U FOR THE ERROR SATURATION NONLINEARITY (34)

To evaluate the expectation

$$h_U\left[E\left[e_a^2(i)\right]\middle| v(i)\right] = E\left[f[e^2(i)]\middle| v(i)\right]$$

for the error saturation nonlinearity $f[e(i)] = \int_0^{e(i)} e^{-z^2/2\sigma_{\rm sat}^2}dz$, we rely on the equivalent representation

$$f[e(i)] = \sqrt{2\pi\sigma_{\rm sat}^2}\left(-\frac{1}{\pi}\int_0^{\pi/2} e^{-e^2(i)/2\sigma_{\rm sat}^2\sin^2(\theta)}d\theta\right) \cdot \text{sign}(e(i)). \quad (39)$$

Powers of f are obtained by changing the integration limits in (39) (in addition to other minor changes, see [40]). Thus

$$f^2[e(i)] = 2\pi\sigma_{\rm sat}^2\left(\frac{1}{4} - \frac{1}{\pi}\int_{\pi/4}^{\pi/2} e^{-e^2(i)/2\sigma_{\rm sat}^2\sin^2(\theta)}d\theta\right). \quad (40)$$

Thanks to (40), in evaluating $Ef^2[e(i)]$ given $v(i)$, the expectation operator can move inside the integral and operate on its integrand, and we can show that

$$E\left[e^{-e^2(i)/2\sigma_{\rm sat}^2\sin^2(\theta)}\middle| v(i)\right] = \sqrt{\frac{\sigma_{\rm sat}^2\sin^2(\theta)}{\sigma_{e_a}^2 + \sigma_{\rm sat}^2\sin^2(\theta)}} \cdot e^{-v^2(i)/2(\sigma_{e_a}^2+\sigma_{\rm sat}^2\sin^2(\theta))} \quad (41)$$

where $\sigma_{e_a}^2 = E\left[e_a^2(i)\right]$. This yields the desired result.

REFERENCES

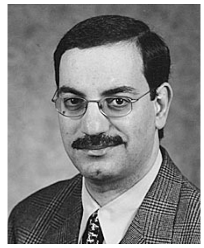
[1] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[3] O. Macchi, *Adaptive Processing: The LMS Approach with Applications in Transmission*. New York: Wiley, 1995.
[4] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 578–586, Aug. 1982.
[5] E. Walach and B. Widrow, "The least-mean fourth (LMF) adaptive algorithm and its family," *IEEE Trans. Inform. Theory*, vol. IT–30, pp. 275–283, Apr. 1984.
[6] J. Gibson and S. Gray, "MVSE adaptive filtering subject to a constraint on MSE," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 603–608, May 1988.
[7] W. Sethares, "Adaptive algorithms with nonlinear data and error functions," *IEEE Trans. Signal Processing*, vol. 40, pp. 2199–2206, Sept. 1992.
[8] 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.
[9] N. J. Bershad, "On error saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 440–452, Apr. 1988.
[10] N. J. Bershad and M. Bonnet, "Saturation effects in LMS adaptive echo cancellation for binary data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1687–1696, Oct. 1990.
[11] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, vol. 42, pp. 2937–2944, Nov. 1994.
[12] T. Claasen and W. Mecklenbräuker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 510–518, June 1981.
[13] O. Tanrikulu and J. A. Chambers, "Convergence and steady-state properties of the least-mean mixed-norm (LMMN) adaptive algorithm," *Proc. Inst. Elect. Eng.—Vision, Image Signal Process.*, vol. 143, no. 3, pp. 137–142, June 1996.
[14] J. A. Chambers, O. Tanrikulu, and A. G. Constantindes, "Least mean mixed-norm adaptive filtering," *Electron. Lett.*, vol. 30, no. 19, pp. 1574–1575, Sept. 1994.
[15] S. Koike, "Convergence analysis of a data echo canceler with a stochastic gradient adaptive FIR filter using the sign algorithm," *IEEE Trans. Signal Processing*, vol. 43, pp. 2852–2861, Dec. 1995.
[16] R. Sharma, W. Sethares, and J. Bucklew, "Asymptotic analysis of stochastic gradient-based adaptive filtering algorithms with general cost functions," *IEEE Trans. Signal Processing*, vol. 44, pp. 2186–2194, Sept. 1996.
[17] S. Florian and A. Feuer, "Performance analysis of the LMS algorithm with a tapped delay line (two-dimensional case)," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1542–1549, Dec 1986.
[18] S. C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Trans. Signal Processing*, vol. 43, pp. 2863–2871, Dec. 1995.
[19] V. H. Nascimento and A. H. Sayed, "Stability of the lms adaptive filter by means of a state equation," in *Proc. 36th Annu. Allerton Conf. Commun., Contr., Comput.*, 1998, pp. 242–251.
[20] W. A. Gardner, "Learning characteristic of stochastic-descent algorithms: A general study, analysis, and critique," *Signal Process.*, vol. 6, no. 2, pp. 113–133, Apr. 1984.
[21] 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.
[22] 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.
[23] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963–993, May/June 1979.
[24] J. Bermudez and N. J. Bershad, "A nonlinear analytical model for the quantized LMS algorithm- The arbitrary step size case," *IEEE Trans. Signal Processing*, vol. 44, pp. 1175–1183, May 1996.
[25] 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.
[26] 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.

[27] ——, "Ability of adaptive filters to track carrier offsets and random channel nonstationarities," *IEEE Trans. Signal Processing*, vol. 50, pp. 1533–1544, July 2002.
[28] ——, "A feedback analysis of the tracking performance of blind adaptive equalization algorithms," in *Proc. CD Conf.*, vol. 1, 1999, pp. 174–179.
[29] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive algorithms via the small gain theorem," in *Proc. SPIE*, vol. 2563, 1995, pp. 458–469.
[30] 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.
[31] A. H. Sayed and M. Rupp, "Robustness issues in adaptive filtering," in *DSP Handbook*. Boca Raton, FL: CRC, 1998, ch. 20.
[32] M. Rupp and A. H. Sayed, "On the convergence of blind adaptive equalizers for constant-modulus signals," *IEEE Trans. Commun.*, vol. 48, pp. 795–803, May 2000.
[33] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Processing*, vol. 51, pp. 639–652, Mar. 2003.
[34] R. Price, "A useful theorem for nonlinear devices having gaussian inputs," *IEEE Trans. Inform. Theory*, vol. IT-4, pp. 69–72, June 1958.
[35] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.
[36] S. C. Douglas and T. H. -Y. Meng, "Stochastic gradient adaptation under general error criterion," *IEEE Trans. Signal Processing*, vol. 42, pp. 1335–1351, June 1994.
[37] T. Birkel, "Laws of large numbers under dependence assumptions," *Statist. Probab. Lett.*, vol. 14, no. 4, pp. 355–362.
[38] T. Y. Al-Naffouri, A. Zerguine, and M. Bettayeb, "Convergence analysis of the LMS algorithm with a general error nonlinearity and an iid input," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 1, 1998, pp. 556–559.
[39] N. R. Yousef and A. H. Sayed, "Tracking analysis of the LMF and LMMN adaptive algorithms," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 1, 1999, pp. 786–790.
[40] M. K. Simon and M. -S. Alouini, "A unified approach to the probability of error for noncoherent and differentially coherent modulations over generalized fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1625–1638, Dec. 1998.
[41] E. Masry and F. Bullo, "Convergenece analysis of the sign algorithm for adaptive filtering," *IEEE Trans. Inform. Theory*, vol. 41, pp. 489–495, Mar. 1995.

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.