

Adaptive Filters with Error Nonlinearities: Mean-Square Analysis and Optimum Design

Tareq Y. Al-Naffouri

Electrical Engineering Department, Stanford University, CA 94305, USA
 Email: naffouri@stanford.edu

Ali H. Sayed

Electrical Engineering Department, University of California, Los Angeles, CA 90095, USA
 Email: sayed@ee.ucla.edu

Received 7 August 2001 and in revised form 10 October 2001

This paper develops a unified approach to the analysis and design of adaptive filters with error nonlinearities. In particular, the paper performs stability and steady-state analysis of this class of filters under weaker conditions than what is usually encountered in the literature, and without imposing any restriction on the color or statistics of the input. The analysis results are subsequently used to derive an expression for the optimum nonlinearity, which turns out to be a function of the probability density function of the estimation error. Some common nonlinearities are shown to be approximations to the optimum nonlinearity. The framework pursued here is based on energy conservation arguments.

Keywords and phrases: adaptive filter, mean-square error, energy conservation, transient analysis, steady-state analysis, stability, error nonlinearity.

1. INTRODUCTION

The least-mean-squares (LMS) algorithm is a popular adaptive algorithm because of its simplicity and robustness [1, 2]. Many LMS-like algorithms have been suggested and analyzed in the literature with the aim of retaining the desirable properties of LMS and simultaneously offsetting some of its limitations. Of particular importance is the class of least-mean-squares algorithms with error nonlinearities. Table 1 lists examples from this class of algorithms for real-valued data.

TABLE 1: Examples of $f[e(i)]$.

ALGORITHM	ERROR NONLINEARITY $f[e(i)]$
LMS	$e(i)$
LMF	$e^3(i)$
LMF family	$e^{2k+1}(i), k \geq 0$
LMMN	$ae(i) + be^3(i)$
Sign error	$\text{sign}[e(i)]$
Saturation nonlinearity	$\int_0^{e(i)} \exp\left(-\frac{z^2}{2\sigma_z^2}\right) dz$

Despite the favorable behavior of many of these LMS variants, their choice and design are mostly justified by intuition rather than by rigorous theoretical arguments. Even the LMS algorithm, which has long been considered as an approximate solution to a least-mean-squares problem, has only recently been justified by a rigorous theory [3].

In this paper, we provide a unifying framework for the mean-square performance of adaptive filters that involve error nonlinearities in their update equations. We will use this analysis to design adaptive algorithms with optimized performance. Before discussing the features of the approach proposed herein and its contributions, we provide, as a motivation, a summary of selected studies dealing with adaptive filters with error nonlinearities. These studies can be classified into two categories.

1. Analysis using simplifying assumptions

Since adaptive algorithms with error nonlinearities are among the most difficult to analyze, it is not uncommon to resort to different methods and assumptions with the intent of performing tractable analysis. This includes:

Linearization: here the error nonlinearity is linearized around an operating point and higher-order terms are discarded as in [4, 5, 6, 7, 8]. Analyses that are based on this technique fail to accurately describe the adaptive filter

performance for large values of the error, for example, at early stages of adaptation. Linearization can be avoided by focusing on a specific nonlinearity (e.g., as in the sign algorithm [9]) or a subclass of nonlinearities (e.g., as in the case of the error saturation algorithm [10, 11]).

Restricting the class of input signals: such as assuming the input to be white and/or Gaussian (e.g., [5, 6, 9, 10, 11, 12, 13, 14, 15, 16]).

Independence assumption: it is very common to assume that successive regression vectors are independent.

Assumptions on the statistics of the error signal: while statistical assumptions are usually imposed on the regression and noise sequences, it is also common to impose statistical conditions on error quantities. For example, in studying the sign-LMS algorithm, it was assumed in [17] that the elements of the weight-error vector are jointly Gaussian. More accurate is the assumption that the residual error is Gaussian, which was adopted in [6, 9, 10, 11]. By central limit arguments, this assumption is justified for long adaptive filters. More importantly, the assumption is as valid in the early stages as in the final stages of adaptation.

Assuming Gaussian noise: noise is sometimes restricted to be i.i.d Gaussian as in [6, 9, 17, 18], although Gaussianity is not as common as the previous assumptions.

Most studies of adaptive algorithms with error nonlinearities rely on a selection from the above array of assumptions/techniques.

II. Optimal designs [8, 19, 20, 21]

Here one attempts to construct adaptive algorithms with optimum nonlinearities. A natural prerequisite is to evaluate some measures of performance (e.g., [6, 7, 16, 22]) and then minimize them to arrive at optimum choices for the nonlinearity [8, 19, 20, 21]. The difficulty, of course, is that the analysis is often plagued by the aforementioned assumptions and techniques. The result is that the optimum nonlinearities obtained are not any more valid than the restrictions imposed by the analysis.

1.1. The approach of this paper

In this paper, we address some of the above concerns. In particular, we present a unified approach to the mean-square analysis of adaptive algorithms with general error nonlinearities. The approach relies on energy conservation arguments and applies under weaker assumptions than what is available in the literature. Our performance results are subsequently optimized to obtain an expression for the optimum nonlinearity. In what follows, we list the contributions of the paper. This also serves as a layout for its organization.

(1) After introducing our notation, we set up the stage in the next section by defining the adaptive filtering problem. We also derive an energy relation that will be the starting point for much of the subsequent analysis.

(2) The energy relation is used in Section 3 to study mean-square stability. In particular, without relying on any independence-like assumptions, we derive bounds on the step-size for stability.

(3) Section 4 is devoted to studying the steady-state behavior, where we show that the mean-square error can be obtained as the fixed point of a nonlinear equation. The stability and steady-state analysis apply under weaker conditions than usual, and these conditions become reasonably accurate for long adaptive filters.

(4) The steady-state results are used in Section 5 to obtain an expression for the optimum nonlinearity, which is valid for all stages of adaptation. The nonlinearity turns out to be a function of the noise probability density function (pdf). We show how the nonlinearity manifests itself for different noise distributions and how it relates to more common nonlinearities.

1.2. Notation

We focus on real-valued data, although the extension to complex-valued data is immediate. Small boldface letters are used to denote vectors, for example, \mathbf{w} . Also, the symbol T denotes transposition. The notation $\|\mathbf{w}\|^2$ stands for the squared Euclidean norm of a vector, $\|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{w}$. All vectors are column vectors except for a single vector, namely the input data vector \mathbf{u}_i , which is taken to be a row vector. The time instant is placed as a subscript for vectors (e.g., \mathbf{w}_i) and between parentheses for scalars (e.g., $e(i)$).

2. ADAPTIVE ALGORITHMS WITH ERROR NONLINEARITIES

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

$$d(i) = \mathbf{u}_i \mathbf{w}^o + v(i). \quad (1)$$

Here $v(i)$ accounts for measurement noise and modeling errors. Many adaptive schemes have been proposed in the literature for this purpose (cf. [1, 2]). In this paper, we focus on the class of algorithms

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

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

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

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

Mean-square analysis of adaptive filters is best carried out in terms of the weight-error vector

$$\tilde{\mathbf{w}}_i = \mathbf{w}^o - \mathbf{w}_i \quad (4)$$

and the a priori and a posteriori errors defined by

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

We can use these quantities to rewrite the adaptation equation

(2) as

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - \mu f[e(i)] \mathbf{u}_i^T. \quad (6)$$

Moreover, by combining the defining expressions (3) and (5), we obtain

$$e(i) = e_a(i) + v(i). \quad (7)$$

A relation between the estimation errors $e_a(i)$, $e_p(i)$, and $e(i)$ can be obtained by pre-multiplying both sides of the adaptation equation (6) by \mathbf{u}_i ,

$$\mathbf{u}_i \tilde{\mathbf{w}}_{i+1} = \mathbf{u}_i \tilde{\mathbf{w}}_i - \mu f[e(i)] \|\mathbf{u}_i\|^2 \quad (8)$$

and incorporating the defining expressions (5), which yields

$$e_p(i) = e_a(i) - \mu \|\mathbf{u}_i\|^2 f[e(i)]. \quad (9)$$

2.1. An energy conservation relation

To motivate the subsequent analysis, it is worth listing first the questions that are usually of interest in an adaptive filtering setting. We are often interested in questions related to

Steady-state behavior: which relates to determining the steady-state values of $E[\|\tilde{\mathbf{w}}_i\|^2]$, $E[e_a^2(i)]$, and/or $E[e_p^2(i)]$.

Stability: which relates to determining the range of values of the step-size for which the variances $E[e_a^2(i)]$ and $E[\|\tilde{\mathbf{w}}_i\|^2]$ remain bounded.

Learning curves: which relates to determining the time evolution of the curves $E[e_a^2(i)]$ and $E[\|\tilde{\mathbf{w}}_i\|^2]$.

Observe that the above questions are all conveniently phrased in terms of the error quantities $\{e_a(i), e_p(i), \tilde{\mathbf{w}}_i\}$ or, more accurately, in terms of their energies. This fact motivates us to pursue an energy-based approach.

More specifically, in order to address questions of this kind, we will rely on an energy equality that relates the squared norms of the error quantities $\{e_a(i), e_p(i), \tilde{\mathbf{w}}_i, \tilde{\mathbf{w}}_{i+1}\}$. To derive the energy relation, we combine (6) and (9) so as to eliminate the nonlinearity $f[e(i)]$:

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - (e_a(i) - e_p(i)) \frac{\mathbf{u}_i^T}{\|\mathbf{u}_i\|^2}. \quad (10)$$

We then square both sides to get

$$\begin{aligned} \|\tilde{\mathbf{w}}_{i+1}\|^2 &= \left(\tilde{\mathbf{w}}_i - (e_a(i) - e_p(i)) \frac{\mathbf{u}_i^T}{\|\mathbf{u}_i\|^2} \right)^T \\ &\quad \times \left(\tilde{\mathbf{w}}_i - (e_a(i) - e_p(i)) \frac{\mathbf{u}_i^T}{\|\mathbf{u}_i\|^2} \right). \end{aligned} \quad (11)$$

This yields, after some straightforward manipulations, the energy relation

$$\|\tilde{\mathbf{w}}_{i+1}\|^2 + \frac{|e_a(i)|^2}{\|\mathbf{u}_i\|^2} = \|\tilde{\mathbf{w}}_i\|^2 + \frac{|e_p(i)|^2}{\|\mathbf{u}_i\|^2}. \quad (12)$$

This result is *exact* for any adaptive algorithm described by (2); no approximations whatsoever were used. It has proven

very useful in the study of the performance of adaptive filters, in both deterministic and stochastic analysis. The relation was originally derived in [23, 24, 25] and used in the context of robustness analysis of adaptive filters; it was later used in [26, 27, 28, 29, 30, 31] in the context of steady-state and transient analysis of adaptive filters.

Since we are interested in the mean-square behavior of adaptive filters, we take expectations of both sides of (12) and write

$$\begin{aligned} E[\|\tilde{\mathbf{w}}_{i+1}\|^2] + E\left[\frac{|e_a(i)|^2}{\|\mathbf{u}_i\|^2}\right] \\ = E[\|\tilde{\mathbf{w}}_i\|^2] + E\left[\frac{|e_p(i)|^2}{\|\mathbf{u}_i\|^2}\right] \end{aligned} \quad (13)$$

or, upon replacing the a posteriori error $e_p(i)$ by the equivalent expression (9),

$$\begin{aligned} E[\|\tilde{\mathbf{w}}_{i+1}\|^2] &= E[\|\tilde{\mathbf{w}}_i\|^2] - 2\mu E[e_a(i) f[e(i)]] \\ &\quad + \mu^2 E[\|\mathbf{u}_i\|^2 f^2[e(i)]]. \end{aligned} \quad (14)$$

This averaged form of the energy relation is the starting point of our analysis. In the course of answering the adaptive filtering questions, we will not attempt to develop (14) into a self-contained recursion as is usually done in literature. Rather, our efforts will be centered around manipulating the two expectations that appear on the right-hand side of (14) by imposing as little assumptions as necessary to answer the adaptive filtering question under consideration. In particular, the following two assumptions will be used throughout our analysis:

(AN) The noise sequence $\{v(i)\}$ is independent, identically distributed, and independent of the input sequence $\{\mathbf{u}_i\}$.

(AG) The filter is long enough such that $e_a(i)$ is Gaussian.

The independence assumption on the noise is valid in many practical situations. Notice, however, that we make no assumption on the noise statistics which is contrary to the Gaussian restriction that is sometimes imposed in literature (e.g., [6, 9, 17, 18]).

Assumption (AG) is justified for long filters by the central limit theorem. As such, the validity of the assumption is dependent on the filter order M . Nevertheless, the assumption remains as valid in the initial stages of adaptation as it is in the final stages. This comes contrary to the linearization arguments that are usually employed when dealing with error nonlinearities and which are only valid in the final stages of adaptation (see [6, 7, 8]). By expressing the two expectations in (14) in terms of the second-order moment $E[e_a^2(i)]$, we basically bypass the need for linearization.

3. MEAN-SQUARE STABILITY

Stability is usually studied in literature by first developing (14) into a *self-contained* recursion and subsequently determining conditions on the step-size in order to guarantee the stability

of the recursion. As we will now see, we can study stability *directly* from (14) thus doing away with the self-contained recursion and with any auxiliary assumptions that are invoked to develop this recursion. In particular, starting from (14), we pursue a Lyapunov approach to stability where we provide a nontrivial upper bound on μ for which $E[\|\tilde{\mathbf{w}}_i\|^2]$ remains uniformly bounded for all i . More specifically, we will show how to calculate a bound μ_0 for which

$$\mu \leq \mu_0 \implies E[\|\mathbf{w}_i\|^2] \leq C < \infty \quad (15)$$

for some constant C .

3.1. A monotone sequence of weight energies

Starting from (14), it is easy to see that

$$\begin{aligned} E[\|\tilde{\mathbf{w}}_{i+1}\|^2] &\leq E[\|\tilde{\mathbf{w}}_i\|^2] \\ &\iff -2\mu E[e_a(i)f[e(i)]] \\ &\quad + \mu^2 E[\|\mathbf{u}_i\|^2 f^2[e(i)]] \leq 0. \end{aligned} \quad (16)$$

Thus, if we choose μ such that for all i

$$\mu \leq 2 \frac{E[e_a(i)f[e(i)]]}{E[\|\mathbf{u}_i\|^2 f^2[e(i)]]}, \quad (17)$$

then the sequence $\{E[\|\tilde{\mathbf{w}}_i\|^2]\}$ will be decreasing and (being bounded from below) also convergent. Alternatively, a sufficient condition for stability would be

$$\mu \leq 2 \inf_{i \geq 0} \frac{E[e_a(i)f[e(i)]]}{E[\|\mathbf{u}_i\|^4]^{1/2} E[f^4[e(i)]]^{1/2}}, \quad (18)$$

where we appealed to the Cauchy-Schwartz inequality to bound the denominator by

$$E[\|\mathbf{u}_i\|^2 f^2[e(i)]] \leq E[\|\mathbf{u}_i\|^4]^{1/2} E[f^4[e(i)]]^{1/2}. \quad (19)$$

To proceed further, the convenience of the Gaussian assumption on $e_a(i)$ can be brought into fruit to evaluate the expectations in (18). In particular, the expectations can be written as functions of the second moment¹ $E[e_a^2(i)]$. This prompts

¹Since $e_a(i)$ is assumed Gaussian and independent of $v(i)$, we can, for example, write

$$\begin{aligned} E[f^4[e_a(i) + v(i)]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^4[e_a(i) + v(i)] \frac{1}{\sqrt{2\pi E[e_a^2(i)]}} \\ &\quad \times e^{-e_a^2(i)/2E[e_a^2(i)]} p_v(v) de_a(i) dv \\ &= \int_{-\infty}^{\infty} p_v(v) dv \int_{-\infty}^{\infty} f^4[e_a(i) + v(i)] \frac{1}{\sqrt{2\pi E[e_a^2(i)]}} \\ &\quad \times e^{-e_a^2(i)/2E[e_a^2(i)]} de_a(i). \end{aligned}$$

The inner integral depends on $e_a^2(i)$ through $E[e_a^2(i)]$ only and, hence, so does $E[f^4[e_a(i) + v(i)]]$.

us to define²

$$h_G[E[e_a^2(i)]] \triangleq \frac{E[e_a(i)f[e(i)]]}{E[e_a^2(i)]}, \quad (20)$$

$$h_C[E[e_a^2(i)]] \triangleq E[f^4[e(i)]]. \quad (21)$$

For future reference, h_G is tabulated in Table 2 for the error nonlinearities of Table 1.

Upon substituting (20) and (21) into (18), we see that a sufficient condition for convergence is that

$$\mu \leq \frac{2}{[E[\|\mathbf{u}_i^4\|]^{1/2}} \left(\inf_{i \geq 0} \frac{E[e_a^2(i)] \cdot h_G[E[e_a^2(i)]]}{\sqrt{h_C[E[e_a^2(i)]]}} \right), \quad (22)$$

where we moved the expectation $E[\|\mathbf{u}_i\|^4]$ outside the minimization since the input data is assumed stationary. Observe now that all terms in the above minimization are functions of $E[e_a^2(i)]$. We can therefore rewrite (22) as

$$\mu \leq \frac{2}{E[\|\mathbf{u}_i^4\|]^{1/2}} \left(\inf_{E[e_a^2(i)]} \frac{E[e_a^2(i)] \cdot h_G[E[e_a^2(i)]]}{\sqrt{h_C[E[e_a^2(i)]]}} \right). \quad (23)$$

We emphasize that the minimization takes place over the possible values of $E[e_a^2(i)]$ *only*; these values are not arbitrary but correspond to those assumed by the learning curve of the adaptive filter. As it stands, the bound (23) is still not useful, and we need to replace it by a *time-independent* bound.

3.1.1 Removing the time index (dependence)

We can replace the range of feasible values of $E[e_a^2(i)]$ by the larger set

$$\Omega = \{E[e_a^2] : 0 \leq E[e_a^2] < \infty\}. \quad (24)$$

Minimization over Ω is easier to carry out and we additionally have

$$\begin{aligned} &\inf_{E[e_a^2] \in \Omega} \left(\frac{E[e_a^2] \cdot h_G[E[e_a^2]]}{\sqrt{h_C[E[e_a^2]]}} \right) \\ &\leq \inf_{E[e_a^2(i)]} \left(\frac{E[e_a^2(i)] \cdot h_G[E[e_a^2(i)]]}{\sqrt{h_C[E[e_a^2(i)]]}} \right). \end{aligned} \quad (25)$$

Almost always, however, minimization over Ω yields a null value and hence is useless. A more intelligent choice of the feasible set is thus called for.

3.1.2 A lower bound on $E[e_a^2(i)]$

A nonzero lower bound on $E[e_a^2(i)]$ can be obtained by noting that $E[e_a^2(i)]$ cannot be lower than the Cramer-Rao bound λ associated with the underlying estimation process, that is, the problem of estimating the random quantity $\mathbf{u}_i \mathbf{w}^o$

²The subscript in h_G points to the fact that the Gaussian assumption (AG) is the major assumption in evaluating $E[e_a(i)f[e(i)]]$. The subscript in h_C is a reminder that the Cauchy-Schwartz inequality is the key step in approximating $E[\|\mathbf{u}_i\|^2 f^2[e_a(i) + v(i)]]$.

TABLE 2: $h_G[\cdot]$ for the error nonlinearities of Table 1 ($\sigma_{e_a}^2 \triangleq E[e_a^2(i)]$).

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

by using $\mathbf{u}_i \mathbf{w}_i$ (see [32, page 72]). Thus, we can write

$$E[e_a^2(i)] \geq \lambda \quad (26)$$

and subsequently replace Ω with the smaller set

$$\Omega' = \{E[e_a^2] : \lambda \leq E[e_a^2]\}. \quad (27)$$

This results in the tighter bound

$$\mu \leq \frac{2}{[E\|\mathbf{u}_i\|^4]^{1/2}} \left(\inf_{E[e_a^2] \in \Omega'} \frac{E[e_a^2] \cdot h_G[E[e_a^2]]}{\sqrt{h_G[E[e_a^2]]}} \right). \quad (28)$$

3.1.3 An upper bound on $E[e_a^2(i)]$

It turns out that for adaptive filters that employ a linear or sublinear error nonlinearities (e.g., the LMS and the sign algorithms), the feasible set Ω' defined by (27) is good enough (as we will show in the examples further ahead). In other words, the set is small enough to produce a positive upper bound on the step size for stability. In other cases, for example, the LMF algorithm and its family, we need a more compact set and an upper bound on $E[e_a^2(i)]$ is also necessary. Not surprisingly perhaps, the upper bound depends on the initial condition of the adaptive filter.

To this end, observe that since $e_a(i)$ is Gaussian, by assumption, it satisfies

$$\begin{aligned} E[e_a^2(i)] &= \frac{1}{4} (E[|e_a(i)|])^2 \\ &= \frac{1}{4} (E[|\mathbf{u}_i \tilde{\mathbf{w}}_i|])^2 \\ &\leq \frac{1}{4} (E[\|\mathbf{u}_i\| \|\tilde{\mathbf{w}}_i\|])^2, \end{aligned} \quad (29)$$

where the last line follows from the Cauchy-Schwartz inequality. To proceed further, we apply the same inequality to the expectation operator this time, splitting it as

$$\begin{aligned} E[e_a^2(i)] &\leq \frac{1}{4} \left([E\|\mathbf{u}_i\|^2]^{1/2} [E\|\tilde{\mathbf{w}}_i\|^2]^{1/2} \right)^2 \\ &= \frac{1}{4} E[\|\mathbf{u}_i\|^2] E[\|\tilde{\mathbf{w}}_i\|^2] \\ &= \frac{1}{4} \text{Tr}(\mathbf{R}) E[\|\tilde{\mathbf{w}}_i\|^2]. \end{aligned} \quad (30)$$

Here \mathbf{R} denotes the covariance matrix of the regression vector \mathbf{u}_i . Now since the bounds (18) or (23) on μ ensure that $E[\|\tilde{\mathbf{w}}_i\|^2]$ is decreasing, we have

$$E[\|\tilde{\mathbf{w}}_i\|^2] \leq E[\|\tilde{\mathbf{w}}_0\|^2]. \quad (31)$$

This gives us the desired upper bound

$$E[e_a^2(i)] \leq \frac{1}{4} \text{Tr}(\mathbf{R}) E[\|\tilde{\mathbf{w}}_0\|^2]. \quad (32)$$

The two bounds (26) and (32) produce the alternative feasibility set

$$\Omega'' = \left\{ E[e_a^2] : \lambda \leq E[e_a^2] \leq \frac{1}{4} \text{Tr}(\mathbf{R}) E[\|\tilde{\mathbf{w}}_0\|^2] \right\} \quad (33)$$

which leads to the following conclusion.

Theorem 1 (stability). Consider an adaptive filter of the form

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu \mathbf{u}_i^T f[e(i)], \quad i \geq 0, \quad (34)$$

where $e(i) = d(i) - \mathbf{u}_i \mathbf{w}_i$ and $d(i) = \mathbf{u}_i \mathbf{w}^0 + v(i)$. Assume the noise sequence $v(i)$ is i.i.d and independent of \mathbf{u}_i , and that the filter is long enough so that $e_a(i) = \mathbf{u}_i (\mathbf{w}^0 - \mathbf{w}_i)$ is Gaussian. Then sufficient conditions for stability are

$$\mu \leq \frac{2}{[E\|\mathbf{u}_i\|^4]^{1/2}} \left(\inf_{E[e_a^2] \in \Omega'} \frac{E[e_a^2] \cdot h_G[E[e_a^2]]}{\sqrt{h_G[E[e_a^2]]}} \right) \quad (35)$$

or

$$\mu \leq \frac{2}{[E\|\mathbf{u}_i\|^4]^{1/2}} \left(\inf_{E[e_a^2] \in \Omega''} \frac{E[e_a^2] \cdot h_G[E[e_a^2]]}{\sqrt{h_C[E[e_a^2]]}} \right), \quad (36)$$

where

$$\begin{aligned} \Omega' &= \{E[e_a^2] : \lambda \leq E[e_a^2]\}, \\ \Omega'' &= \left\{ E[e_a^2] : \lambda \leq E[e_a^2] \leq \frac{1}{4} \text{Tr}(\mathbf{R}) E[\|\tilde{\mathbf{w}}_0\|^2] \right\} \end{aligned} \quad (37)$$

and λ is the Cramer-Rao bound associated with the problem of estimating the random quantity $\mathbf{u}_i \mathbf{w}^o$ by using $\mathbf{u}_i \mathbf{w}_i$.

As indicated earlier, the bound (35) will be zero for superlinear functions $f[\cdot]$ (e.g., LMF and LMMN), in which case the tighter bound (36) will be more useful. Notice also that the above bounds are derived without relying on any independence-like assumptions.

3.2. Examples: explicit bounds on μ for stability

3.2.1 The LMS algorithm

Instead of applying Theorem 1, we can in the LMS case be more specific. Thus starting from (17), we obtain

$$\begin{aligned} \mu &\leq 2 \inf_{E[e_a^2(i)]} \frac{E[e_a(i)e(i)]}{E[\|\mathbf{u}_i\|^2 e^2(i)]} \\ &= 2 \inf_{E[e_a^2(i)]} \frac{E[e_a^2(i)]}{E[\|\mathbf{u}_i\|^2 e_a^2(i) + \sigma_v^2 E[\|\mathbf{u}_i\|^2]]} \\ &\leq 2 \inf_{E[e_a^2(i)]} \frac{E[e_a^2(i)]}{[E\|\mathbf{u}_i\|^4]^{1/2} E[e_a^4(i)]^{1/2} + \sigma_v^2 E[\|\mathbf{u}_i\|^2]} \\ &= 2 \inf_{E[e_a^2(i)]} \frac{E[e_a^2(i)]}{\sqrt{3} [E\|\mathbf{u}_i\|^4]^{1/2} E[e_a^2(i)] + \sigma_v^2 E[\|\mathbf{u}_i\|^2]}, \end{aligned} \quad (38)$$

where the last line follows from the Gaussian assumption (AG). By performing the minimization over Ω' , we obtain the tighter bound

$$\begin{aligned} \mu &\leq 2 \inf_{E[e_a^2(i)] \geq \lambda} \frac{E[e_a^2(i)]}{\sqrt{3} [E\|\mathbf{u}_i\|^4]^{1/2} E[e_a^2(i)] + \sigma_v^2 E[\|\mathbf{u}_i\|^2]} \\ &= \frac{2\lambda}{\sqrt{3} [E\|\mathbf{u}_i\|^4]^{1/2} \lambda + \sigma_v^2 E[\|\mathbf{u}_i\|^2]}. \end{aligned} \quad (39)$$

For binary inputs, stability of the LMS algorithm can be established even without the Gaussian assumption (AG). For then, $\|\mathbf{u}_i\|^2 = M$ and stability is guaranteed if

$$\begin{aligned} \mu &\leq 2 \inf_{E[e_a^2(i)]} \frac{E[e_a(i)e(i)]}{E[\|\mathbf{u}_i\|^2 e^2(i)]} = \frac{2}{M} \inf_{E[e_a^2(i)]} \frac{E[e_a^2(i)]}{E[e_a^2(i)] + \sigma_v^2} \\ &\leq \frac{2}{M} \inf_{E[e_a^2] \geq \lambda} \frac{E[e_a^2]}{E[e_a^2] + \sigma_v^2} = \frac{2}{M} \frac{\lambda}{\lambda + \sigma_v^2}. \end{aligned} \quad (40)$$

3.2.2 The sign algorithm

For the sign algorithm, the bound (35) reads

$$\begin{aligned} \mu &\leq \frac{2}{[E\|\mathbf{u}_i\|^4]^{1/2}} \inf_{E[e_a^2] \geq \lambda} \sqrt{\frac{2}{\pi}} (E[e_a^2])^{3/2} E[e^{-v^2(i)/2E[e_a^2]}] \\ &= \sqrt{\frac{8}{\pi}} \frac{\lambda^{3/2}}{[E\|\mathbf{u}_i\|^4]^{1/2}} E[e^{-v^2(i)/2\lambda}]. \end{aligned} \quad (41)$$

3.2.3 The LMF algorithm

For the LMF algorithm, we employ the tighter bound (36)

$$\begin{aligned} \mu &\leq \frac{2}{[E\|\mathbf{u}_i\|^4]^{1/2}} \inf_{E[e_a^2(i)] \in \Omega''} \frac{E[e_a^2(i)] \cdot h_G[E[e_a^2(i)]]}{\sqrt{h_C[E[e_a^2(i)]]}} \\ &= \frac{2}{[E\|\mathbf{u}_i\|^4]^{1/2}} \inf_{E[e_a^2(i)] \in \Omega''} \frac{3E[e_a^2(i)](E[e_a^2(i)] + \sigma_v^2)}{E[e^{12}(i)]^{1/2}}. \end{aligned} \quad (42)$$

(43)

Using the Gaussian assumption, it is easy to evaluate the expectation $E[e^{12}(i)]$, and subsequently carry out the minimization in (43). However, this is not necessary for all what we need is to make sure that the bound in (43) is positive. Since $E[e_a^2] h_G/\sqrt{h_C}$ is strictly positive over Ω'' and since Ω'' is compact, we conclude that

$$\inf_{E[e_a^2] \in \Omega''} \frac{E[e_a^2] \cdot h_G(E[e_a^2])}{\sqrt{h_C(E[e_a^2])}} > 0. \quad (44)$$

For design purposes, we could determine the infimum of $E[e_a^2] h_G/\sqrt{h_C}$ over Ω'' . Here we are interested in the simpler task of establishing stability.

Remark. Linearization was employed in [4] to prove the stability of the LMF algorithm. While this might be reasonable for steady-state analysis, it need not be valid when stability is concerned. Notice that no linearization arguments were used here. Observe also that for the LMF algorithm, the infimum of $E[e_a^2] h_G/\sqrt{h_C}$ over Ω' is zero. That explains why we had to perform minimization over the smaller set Ω'' , which is initial-condition dependent. This suggests that the performance of the LMF is initial-condition dependent too, as is often confirmed by simulation. In general, this is expected to be the case for algorithms employing super-nonlinearity.

4. STEADY-STATE BEHAVIOR

To investigate the steady-state behavior, we again start from the averaged energy relation (14). Assuming that the filter is stable, it should attain a steady-state where it holds that

$$\lim_{i \rightarrow \infty} E[\|\tilde{\mathbf{w}}_{i+1}\|^2] = \lim_{i \rightarrow \infty} E[\|\tilde{\mathbf{w}}_i\|^2]. \quad (45)$$

Therefore, (14) becomes in steady-state

$$\lim_{i \rightarrow \infty} E[e_a(i) f[e(i)]] = \frac{\mu}{2} \lim_{i \rightarrow \infty} E[\|\mathbf{u}_i\|^2 f^2[e(i)]]. \quad (46)$$

Both of the expectations in (46) were dealt with as part of the stability analysis. Using the Gaussian assumption, we have already argued that $E[e_a(i)f[e(i)]]$ can be written as a function of $E[e_a^2(i)]$. In particular, from (20) we have

$$E[e_a(i)f[e(i)]] = E[e_a^2(i)]h_G[E[e_a^2(i)]], \quad (47)$$

where the function h_G was tabulated in Table 2 for the nonlinearities in Table 1.

In a similar fashion, we now proceed to evaluate $E[\|\mathbf{u}_i\|^2 f^2[e(i)]]$ in terms of $E[e_a^2(i)]$ (rather than bound it as in the stability analysis). This prompts us to introduce the following ‘‘asymptotic’’ assumption:

(AU) The random variables $\|\mathbf{u}_i\|^2$ and $f^2[e(i)]$ are asymptotically uncorrelated, that is,

$$\lim_{i \rightarrow \infty} E[\|\mathbf{u}_i\|^2 f^2[e(i)]] = E[\|\mathbf{u}_i\|^2] \lim_{i \rightarrow \infty} E[f^2[e(i)]]. \quad (48)$$

Assumption (AU) has the same spirit as the independence assumption³ but it is weaker. For one thing, relation (48) is exact for constant modulus inputs while the independence assumption is not. Moreover, the separation property (48) need only be satisfied asymptotically. Fortunately, assumption (AU) acts in harmony with the Gaussianity assumption on $e_a(i)$ in that it also becomes more realistic as the filter gets longer. For then, by an ergodic argument, $\|\mathbf{u}_i\|^2$ behaves like the second moment of the input (scaled by the filter length M).

To proceed, we use the Gaussian assumption (AG) on $e_a(i)$ to express the expectation $E[f^2[e(i)]]$ as a function of the second-order moment $E[e_a^2(i)]$, which motivates the definition

$$h_U[E[e_a^2(i)]] \triangleq E[f^2[e(i)]]. \quad (49)$$

The function h_U is tabulated in Table 3 for the nonlinearities of Table 1.

This definition, together with (48), yields

$$\begin{aligned} \lim_{i \rightarrow \infty} E[\|\mathbf{u}_i\|^2 f^2[e(i)]] &= E[\|\mathbf{u}_i\|^2] \lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]] \\ &= \text{Tr}(\mathbf{R}) \lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]]. \end{aligned} \quad (50)$$

Upon substituting (47) and (50) into (46), we obtain

$$\lim_{i \rightarrow \infty} E[e_a^2(i)] \triangleq \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{\lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]]}{\lim_{i \rightarrow \infty} h_G[E[e_a^2(i)]]}. \quad (51)$$

Now denote the steady-state mean-square error by

$$S \triangleq \lim_{i \rightarrow \infty} E[e_a^2(i)]. \quad (52)$$

Then since both h_U and h_G are analytic in their arguments,

we have

$$\begin{aligned} \lim_{i \rightarrow \infty} h_U[E[e_a^2(i)]] &= h_U[S], \\ \lim_{i \rightarrow \infty} h_G[E[e_a^2(i)]] &= h_G[S], \end{aligned} \quad (53)$$

so that the MSE is the positive root of the nonlinear equation

$$S = \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{h_U[S]}{h_G[S]}. \quad (54)$$

In other words, the MSE is the fixed point of the function $(\mu/2) \text{Tr}(\mathbf{R}) (h_U[S]/h_G[S])$. For a given error nonlinearity f , we can evaluate h_U and h_G (as done in Tables 2 and 3) and subsequently solve for the MSE. Our findings are summarized in the following theorem.

Theorem 2 (steady-state behavior). *Consider the setting of Theorem 1. Assume further that $\|\mathbf{u}_i\|^2$ and $f^2[e(i)]$ are asymptotically uncorrelated, and that the filter is mean-square stable with MSE denoted by S . Then the following equality holds:*

$$S = \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{h_U[S]}{h_G[S]}. \quad (55)$$

To demonstrate the use of this theorem, we provide in what follows expressions for the mean-square error of some of the nonlinearities in Table 1.

4.1. Examples: MSE expressions

4.1.1 The LMS algorithm

In the LMS case, (54) reads

$$S = \frac{\mu}{2} \text{Tr}(\mathbf{R}) (S + \sigma_v^2) \quad (56)$$

or, equivalently,

$$S = \sigma_v^2 \frac{\mu \text{Tr}(\mathbf{R})}{2 - \text{Tr}(\mathbf{R})}. \quad (57)$$

This is a well-known result that was derived in [13] by relying on the independence assumption. Using the energy-based approach of this paper, we only need the weaker assumption (AU), as indicated above and also in [27, 29].⁴

4.1.2 The sign algorithm

In the case of the sign algorithm, we can show that the MSE satisfies

$$S = \mu \sqrt{\frac{\pi}{8}} \text{Tr}(\mathbf{R}) \frac{\sqrt{S}}{E[e^{-v^2(i)/2S}]}. \quad (58)$$

This relation applies irrespective of assumption (AU). We have only appealed to the Gaussian assumption (AG) in arriving at (58)—see also [28]. The expectation that appears

³The independence assumption requires that the input regressors $\{\mathbf{u}_i\}$ form an independent and identically distributed sequence. This assumption is heavily used in the adaptive filtering literature.

⁴This result remains valid irrespective of the Gaussian assumption on $e_a(i)$. The reason is that in the LMS case, the two expectations that appear in (46) are already quadratic in $e_a(i)$ —see [27, 29].

TABLE 3: $h_U[\cdot]$ for the error nonlinearities of Table 1 ($\sigma_{e_a}^2 \triangleq E[e_a^2(i)]$).

ALGORITHM	$h_U[\sigma_{e_a}^2]$ ($v(i)$ Gaussian)	$h_U[\sigma_{e_a}^2]$
LMS	$\sigma_{e_a}^2 + \sigma_v^2$	$\sigma_{e_a}^2 + \sigma_v^2$
LMF	$15(\sigma_{e_a}^2 + \sigma_v^2)^3$	$15\sigma_{e_a}^6 + 45\sigma_{e_a}^4\sigma_v^2 + 15\sigma_{e_a}^2E[v^4(i)] + E[v^6(i)]$
LMF family	$\frac{(4k+2)!}{2^{2k+1}(2k+1)!}(\sigma_{e_a}^2 + \sigma_v^2)^{2k+1}$	$\sum_{j=0}^{2k+1} \binom{4k+2}{2j} \frac{(2j)!}{2^j j!} \sigma_{e_a}^{2j} E[v^{2(2k-j+1)}(i)]$
LMMN	$a^2(\sigma_{e_a}^2 + \sigma_v^2) + 6ab(\sigma_{e_a}^2 + \sigma_v^2)^2 + 15b^2(\sigma_{e_a}^2 + \sigma_v^2)^3$	$15b^2\sigma_{e_a}^6 + (45b^2\sigma_v^2 + 6ab)\sigma_{e_a}^4 + (15b^2E[v^4(i)] + 12ab\sigma_v^2 + a^2)\sigma_{e_a}^2 + E[(bv^2(i) + a)^2v^2(i)]$
Sign error	1	1
Sat. nonlin.	$\sigma_z^2 \sin^{-1}\left(\frac{\sigma_{e_a}^2 + \sigma_v^2}{\sigma_{e_a}^2 + \sigma_v^2 + \sigma_z^2}\right)$	$\frac{\pi}{2}\sigma_z^2 - 2\sigma_z^3 \int_0^{1/\sqrt{2}} \frac{1}{\sqrt{\sigma_{e_a}^2 + \sigma_z^2(1-x^2)}} E[e^{-v^2(i)/2(\sigma_{e_a}^2 + \sigma_z^2(1-x^2))}] dx$

in (58) is carried over the noise pdf. By specifying the noise statistics, we obtain the following special cases:

$$S = \begin{cases} \frac{\alpha + \sqrt{\alpha + 4\sigma_v^2}}{2} & \text{Gaussian noise,} \\ \alpha \sqrt{\frac{6}{\pi}} \frac{\sigma_v}{\text{erf}(\sqrt{3\sigma_v^2/2S})} & \text{Uniform noise,} \\ \alpha \sqrt{S} e^{\sigma_v^2/2S} & \text{Binary noise,} \end{cases} \quad (59)$$

where $\alpha = \mu\sqrt{\pi/8}\text{Tr}(\mathbf{R})$. Each of these equations can be uniquely solved for the MSE. They were derived in [12] under the independence assumption and under an i.i.d restriction on the input, but are rederived here without relying on these restrictions.

4.1.3 The LMF algorithm

For the LMF algorithm, and with the aid of Tables 1 and 2, equation (54) takes the form

$$S = \frac{\mu}{6} \frac{15S^3 + 45\sigma_v^2 S^2 + 15m_{v,4}S + m_{v,6}}{S + \sigma_v^2} \text{Tr}(\mathbf{R}), \quad (60)$$

where $m_{v,4}$ and $m_{v,6}$ denote the fourth and sixth moments of the noise $v(i)$. Finding the MSE 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 MSE.

Gaussian noise

In the Gaussian noise case, (60) simplifies to

$$S = \frac{5\mu}{2} \frac{(S + \sigma_v^2)^3}{S + \sigma_v^2} \text{Tr}(\mathbf{R}) = \frac{\alpha}{2} (S + \sigma_v^2)^2, \quad (61)$$

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

$$S = \frac{(1 - \alpha\sigma_v^2) \pm \sqrt{1 - 2\alpha\sigma_v^2}}{\alpha}. \quad (62)$$

Only the larger root is meaningful.⁵

Remarks. Although several of the MSE expressions above appeared previously in literature, the advantage of the energy-based approach of this article is threefold:

(1) All expressions are obtained as a fall out of the same approach (i.e., by using the same expression (54))—an approach that avoids the need for a self-contained recursion.

(2) The expressions are either new (e.g., (58), (60), and (62)) or are otherwise rederived under a more relaxed set of assumptions (e.g., (57) and (59)).

(3) Here we avoid the need for linearization as in the LMF example. It seems that calculating the steady-state error for super nonlinearities (e.g., the LMF and its family) has always involved some form of linearization arguments (e.g., [4, 6, 8, 16]).

5. OPTIMUM CHOICE OF THE NONLINEARITY

In this section, we build upon the second-order analysis performed above to optimize the choice of the error nonlinearity f . To this end, consider expression (54) for the mean-square error written in a more explicit form

$$S = \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{E[f^2[e(i)]]}{E[e_a(i)f[e(i)]]/E[e_a^2(i)]}. \quad (63)$$

⁵We can show that the smaller root is $O(\mu^3)$. It is well known that the MSE is linearly proportional to the step size, and hence the smaller root can be ignored.

We would like to choose a nonlinearity f that minimizes the mean-square error. If we confine our attention to the class of smooth nonlinearities, we can write, using the Gaussian assumption (AG) and Price theorem [11, 33],

$$\begin{aligned} E[e_a(i)f[e(i)]] &= E[e_a(i)f[e_a(i) + v(i)]] \\ &= E[e_a(i)e(i)]E[f'(e(i))] \quad (64) \\ &= E[e_a^2(i)]E[f'(e(i))]. \end{aligned}$$

Thus, for a smooth error nonlinearity f , the MSE takes the alternative form

$$S = \frac{\mu}{2} \text{Tr}(\mathbf{R}) \frac{E[f^2[e(i)]]}{E[f'[e(i)]]}. \quad (65)$$

The mean-square error cannot be minimized beyond the limit λ , which corresponds to the Cramer-Rao bound of the underlying estimation process. We can thus write

$$\frac{E[f^2[e(i)]]}{E[f'[e(i)]]} \geq \frac{2}{\mu \text{Tr}(\mathbf{R})} \lambda \triangleq \alpha. \quad (66)$$

Now let p_e denote the pdf of $e(i)$. We claim that the nonlinearity

$$f[e(i)] = -\alpha \frac{p'_e[e(i)]}{p_e[e(i)]} \quad (67)$$

attains the lower bound on the MSE and hence is optimum. To see this, we evaluate the numerator and denominator of (66) for this choice of f . Using integration by parts, we have

$$\begin{aligned} E[f'[e(i)]] &= \int_{-\infty}^{\infty} f'[e(i)]p_e[e(i)]de(i) \\ &= f[e(i)]p_e[e(i)] \Big|_{-\infty}^{\infty} \\ &\quad - \int_{-\infty}^{\infty} f[e(i)]p'_e[e(i)]de(i). \end{aligned} \quad (68)$$

For the choice (67) of f , this yields

$$\begin{aligned} E[f'[e(i)]] &= -\alpha p'_e[e(i)] \Big|_{-\infty}^{\infty} \\ &\quad + \alpha \int_{-\infty}^{\infty} \frac{(p'_e[e(i)])^2}{p_e[e(i)]} de(i) \end{aligned} \quad (69)$$

or, assuming that p'_e decays to zero as $e(i)$ approaches $\pm\infty$,

$$E[f'[e(i)]] = \alpha \int_{-\infty}^{\infty} \frac{(p'_e[e(i)])^2}{p_e[e(i)]} de(i). \quad (70)$$

Now for the same choice of f , we have

$$E[f^2[e(i)]] = \alpha^2 \int_{-\infty}^{\infty} \left(\frac{p'_e[e(i)]}{p_e[e(i)]} \right)^2 p_e[e(i)] de(i) \quad (71)$$

$$= \alpha^2 \int_{-\infty}^{\infty} \frac{(p'_e[e(i)])^2}{p_e[e(i)]} de(i). \quad (72)$$

We thus conclude that

$$\frac{E[f^2[e(i)]]}{E[f'[e(i)]]} = \alpha. \quad (73)$$

In other words, the nonlinearity (67) ensures that the MSE approaches the minimum limit determined by the Cramer-Rao bound.

5.1. Removing the constant α

The optimum nonlinearity is specified up to the constant $\alpha = 2\lambda/\mu \text{Tr}(\mathbf{R})$. It turns out that we can use this nonlinearity without calculating α . To see this, we examine the adaptation equation (6) with the optimum choice (67) of f :

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - \mu \left(-\alpha \frac{p'_e[e(i)]}{p_e[e(i)]} \right) \mathbf{u}_i. \quad (74)$$

Since the step size μ is a design parameter that is usually varied, we can absorb the constant α into μ so that

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - \mu \left(-\frac{p'_e[e(i)]}{p_e[e(i)]} \right) \mathbf{u}_i \quad (75)$$

in which case the optimum nonlinearity effectively reads

$$f_{\text{opt}}[e(i)] = -\frac{p'_e[e(i)]}{p_e[e(i)]}. \quad (76)$$

Theorem 3 (optimum nonlinearity). *Consider the setting of Theorem 2. Let p_e denote the pdf of the estimation error $e(i)$. The optimum nonlinearity that minimizes the steady-state mean-square error is given by*

$$f_{\text{opt}}[e(i)] = -\frac{p'_e[e(i)]}{p_e[e(i)]}. \quad (77)$$

5.2. Incorporating the Gaussian assumption on $e_a(i)$

Implementing the optimum nonlinearity (67) or (76) requires that we evaluate the pdf of $e(i)$ at each time instant. It turns out, however, that we can replace this task with the simpler task of evaluating the variance of $e_a(i)$ in addition to specifying the (time-invariant) noise pdf. To this end, recall that our derivation of the MSE expression relied on a Gaussian assumption on the error $e_a(i)$, and, hence, so does our subsequent derivation of the optimum error nonlinearity. Fortunately, this assumption helps us obtain a more explicit expression for the optimum nonlinearity (76). To see this, notice that the estimation error is the sum of two independent random variables,

$$e(i) = e_a(i) + v(i). \quad (78)$$

Therefore, its pdf, $p_e[e(i)]$, is the convolution of the pdf's of $e_a(i)$ and $v(i)$, that is,

$$p_e[e(i)] = p_{e_a}[e(i)] \star p_v[e(i)] \quad (79)$$

$$= \frac{1}{\sqrt{2\pi\sigma_{e_a}^2}} e^{-e^2(i)/2\sigma_{e_a}^2} \star p_v[e(i)]. \quad (80)$$

Here $\sigma_{e_a}^2$ denotes the variance of $e_a(i)$.⁶ The above calculation reduces the determination of the optimum nonlinearity to the task of modeling the noise pdf.

5.3. Estimating the variance $\sigma_{e_a}^2$

Perhaps the most challenging task of the algorithm is estimating the variance of the a priori error $e_a(i)$ —a nonstationary quantity. The easiest way out is to set $\sigma_{e_a}^2$ to some constant value. Alternatively, as done in the simulation, we first estimate the variance of $e(i)$ using a window of samples of $e(i)$, and subsequently estimate $\sigma_{e_a}^2$ from

$$\hat{\sigma}_{e_a}^2 = \hat{\sigma}_e^2 - \sigma_v^2. \quad (81)$$

Furthermore, to avoid malfunctioning of the algorithm, we confine the estimate $\sigma_{e_a}^2$ to a bounded interval $[a, b]$ that is determined by the designer.

Remarks. (1) From above, we see that the derivation of the optimum nonlinearity blends smoothly with the stability and steady-state analysis in that it relies on the same set of assumptions, and is also obtained as a fallout of the same energy conservation approach.

(2) Our derivation of the optimum nonlinearity shares another feature with the (stability) analysis in that it makes use of the fundamental limit set by the Cramer-Rao bound of the underlying estimation process.

(3) Also note that no heavy machinery is appealed to in developing the optimum nonlinearity, maintaining the general themes of clarity and simplicity. In particular, we avoid the variational approaches that are usually employed in literature in designing optimum adaptation schemes (see [8, 20]).

(4) The nonlinearity (76) is derived under simpler assumptions compared to what is available in literature. For instance, we employ a weaker version of the independence assumption (compare with [8, 10, 19, 20, 21]) and make no restriction on the color or statistics of the input (compare with [8, 10, 20]). The nonlinearity (76) also applies irrespective of the noise statistics or whether its pdf is symmetric or not (contrary to what is assumed in [8, 19]). We only require the noise to have zero mean.

(5) More importantly, perhaps, we avoid the need for any linearization arguments making the nonlinearity (76) accurate over all stages of adaptation. In contrast, the optimum nonlinearity

$$f[e(i)] = -\frac{p'_v[e(i)]}{p_v[e(i)]} \quad (82)$$

that was derived in [8] using linearization arguments is only accurate in the final stages of adaptation. In fact, the more accurate expression (76) for the optimum nonlinearity collapses to (82) as the filter reaches its steady-state.

(6) Notice further that expression (76) for the optimum nonlinearity applies irrespective of whether the noise pdf is smooth enough (differentiable) or not. Thanks to the

smoothing convolution operator (see (80)), we can, for example, directly calculate the optimum nonlinearity for binary and uniform noise (see examples below). This comes contrary to the nonlinearity (82) where an artificial smoothing kernel needs to be employed for such singular cases [8].

5.4. Examples

In what follows, we show how the error nonlinearity manifests itself for different noise statistics.

Gaussian noise

When v is Gaussian, so is the estimation error $e(i)$ (since $e(i)$ is then the sum of two Gaussian random variables, $v(i)$ and $e_a(i)$). In this case, the optimum nonlinearity (76) becomes

$$f_{\text{opt}}[e(i)] = -\frac{p'_e[e(i)]}{p_e[e(i)]} = \frac{1}{\sigma_e^2} e(i) \quad (83)$$

which, up to the scaling factor $1/\sigma_e^2$, is the error function of the LMS. Therefore, the LMS is the optimum adaptive algorithm in the presence of Gaussian noise.

Laplacian noise

When v follows a Laplacian distribution, its pdf takes the form

$$p_v[v] = \frac{1}{2} e^{-|v|}. \quad (84)$$

Upon substituting this expression into (80), we can show that the pdf of $e(i)$ takes the form

$$p_e[e(i)] = \frac{1}{4} e^{\sigma_{e_a}^2/2} \left\{ e^{e(i)} \left(1 - \operatorname{erf} \left[\frac{e(i) + \sigma_{e_a}^2}{\sqrt{2\sigma_{e_a}^2}} \right] \right) + e^{-e(i)} \left(1 + \operatorname{erf} \left[\frac{e(i) - \sigma_{e_a}^2}{\sqrt{2\sigma_{e_a}^2}} \right] \right) \right\}. \quad (85)$$

After some straight forward manipulations, we can show that this leads to the following form for the nonlinearity $-p'_e/p_e$

$$\begin{aligned} f_{\text{opt}} = & - \left(e^{e(i)} \left(1 - \operatorname{erf} \left[\frac{e(i) + \sigma_{e_a}^2}{\sqrt{2\sigma_{e_a}^2}} \right] \right) \right. \\ & \left. - \sqrt{2/\pi\sigma_{e_a}^2} e^{-(e(i) + \sigma_{e_a}^2)^2/2\sigma_{e_a}^2} \right) \\ & - e^{-e(i)} \left(1 + \operatorname{erf} \left[\frac{e(i) - \sigma_{e_a}^2}{\sqrt{2\sigma_{e_a}^2}} \right] \right) \\ & \left. - \sqrt{2/\pi\sigma_{e_a}^2} e^{-(e(i) - \sigma_{e_a}^2)^2/2\sigma_{e_a}^2} \right) \\ & \times \left(e^{e(i)} \left(1 - \operatorname{erf} \left[\frac{e(i) + \sigma_{e_a}^2}{\sqrt{2\sigma_{e_a}^2}} \right] \right) \right. \\ & \left. + e^{-e(i)} \left(1 + \operatorname{erf} \left[\frac{e(i) - \sigma_{e_a}^2}{\sqrt{2\sigma_{e_a}^2}} \right] \right) \right)^{-1}, \end{aligned} \quad (86)$$

⁶The time dependence of $\sigma_{e_a}^2$ is suppressed for notational convenience.

where erf is the error function defined by

$$\operatorname{erf}[x] \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (87)$$

Uniform noise

When v is uniformly distributed over $[-1, 1]$, we have

$$p_v[v] = \begin{cases} \frac{1}{2}, & -1 \leq v \leq 1, \\ 0, & \text{otherwise.} \end{cases} \quad (88)$$

Upon substituting (88) into (80), we obtain

$$p_e[e(i)] = \frac{\sigma_{e_a}}{2} \sqrt{\frac{\pi}{2}} \left(\operatorname{erf} \left[\frac{e(i)+1}{\sqrt{2\sigma_{e_a}^2}} \right] - \operatorname{erf} \left[\frac{e(i)-1}{\sqrt{2\sigma_{e_a}^2}} \right] \right). \quad (89)$$

We can use this expression to show that

$$f_{\text{opt}}[e(i)] = \sqrt{\frac{8}{\pi}} \frac{1}{\sigma_{e_a}} e^{-((e(i)+1)/2\sigma_{e_a}^2)} \times \frac{\sinh[e(i)/\sigma_{e_a}^2]}{\operatorname{erf}[(e(i)+1)/\sqrt{2\sigma_{e_a}^2}] - \operatorname{erf}[(e(i)-1)/\sqrt{2\sigma_{e_a}^2}]}. \quad (90)$$

Binary noise

In the binary noise case, we have

$$p_v[v] = \begin{cases} 1 & \text{with probability 0.5,} \\ -1 & \text{with probability 0.5.} \end{cases} \quad (91)$$

In this case, the optimum nonlinearity reads

$$f_{\text{opt}}[e(i)] = \frac{1}{\sigma_{e_a}^2} \left(e(i) - e^{-((e(i)+1)/2\sigma_{e_a}^2)} \tanh \left[\frac{e(i)}{\sigma_{e_a}^2} \right] \right). \quad (92)$$

5.4.1 Simulations

Here we use simulations to illustrate the favorable behavior of the optimum algorithm in comparison to the LMS. The system to be identified is an FIR channel with 15 taps normalized so that the SNR relative to the input and output is the same (10 dB in our case). The input is taken to be Gaussian while the additive output noise is assumed to be binary or Laplacian. The variance σ_e^2 is estimated using the most recent four samples of $e(i)$, and the estimate is in turn used in (81) to estimate the variance of $e_a(i)$. Whenever the estimate $\hat{\sigma}_{e_a}^2$ falls outside the range $[\sigma_v^2/2, 5\sigma_v^2]$ ($[\sigma_v^2/10, 5\sigma_v^2]$) in the binary (Laplacian) noise case, we enforce the assignment $\hat{\sigma}_{e_a}^2 = \sigma_v^2$ instead. The experiment is averaged over one thousand runs.

The LMS and the optimum adaptive algorithms are compared (Figures 1 and 2) in terms of their learning curves; the evolution of $E[\|\tilde{\mathbf{w}}_i\|^2]$ with time (also known as the

mean-square deviation or MSD). We also plot the nonlinearities employed by both algorithms. Since the optimum nonlinearity is *time varying* (through its dependence on $\sigma_{e_a}^2$), it has a stochastic nature. The plots thus show the optimum nonlinearities in their averaged forms.

5.5. Relation between optimum and other nonlinearities

The optimum nonlinearities (76) and (82) are expressed in terms of some pdf ($p_e[e(i)]$ or $p_v[e(i)]$) and its derivative. This makes the nonlinearities difficult to implement since the pdf is usually unknown and/or time-varying. Even if the pdf is known, the corresponding nonlinearity would be expressed in terms of transcendental functions (e.g., as in (90) and (92)), which do not lend themselves to real-time implementations. This is compounded by the fact that different distributions (i.e., pdf's) call for different nonlinearities. Thus, the optimum nonlinearities defy an important feature of least-mean-square algorithms, namely computational simplicity. A more alarming issue though is that these nonlinearities do not seem to relate to the ubiquitous LMS algorithm or its common variants. In the following, we address both of these issues for the nonlinearity (76) by representing the pdf $p_e[e(i)]$ in an Edgeworth expansion, which we now digress to introduce. The nonlinearity (82) can be dealt with similarly.

5.5.1 Edgeworth expansion of p_e

Let y_j denote the j th cumulant of $e(i)$ and let σ_e^2 denote its variance. Assuming p_e to be even, its Edgeworth expansion is given by [34]

$$p_e[e(i)] = \frac{1}{\sigma_e} \phi \left[\frac{e(i)}{\sigma_e} \right] \sum_{j=0}^{\infty} a_{2j} \operatorname{He}_{2j} \left[\frac{e(i)}{\sigma_e} \right], \quad (93)$$

where ϕ is the standard (zero-mean, unit variance) Gaussian pdf and He_{2j} is the Hermite polynomial of degree $2j$ [35]. The coefficients a_j are defined recursively by

$$\begin{aligned} a_0 &= 0, & a_1 &= 1, & a_2 &= 0, \\ a_j &= \frac{1}{j} \left[\frac{y_1}{\sigma_e} a_{j-1} + \left(\frac{y_2}{\sigma_e^2} - 1 \right) a_{j-2} \right. \\ &\quad \left. + \sum_{m=3}^j \frac{y_m}{(m-1)! \sigma_e^m} a_{j-m} \right] \quad (j \geq 3). \end{aligned} \quad (94)$$

Since p_e is even, we can show that $a_{2j+1} = 0$ for all j , and only the even indexed coefficients appear in (93). Thus, $p_e[e(i)]/\phi[e(i)/\sigma_e]$ is an infinite linear combination of Hermite polynomials of even degree. Alternatively, since $\operatorname{He}_{2j}[e(i)/\sigma_e]$ is a linear combination of even powers of its argument, so is the expansion (93). This series expansion is therefore similar to the familiar Taylor expansion except for the fact that it is expressed in statistically relevant terms—the a_j 's, which are defined in terms of the cumulants rather than the derivatives of p_e . The expansion (93) is also different from a Taylor series in that we are not interested in its convergence as much as in representing p_e in as few terms

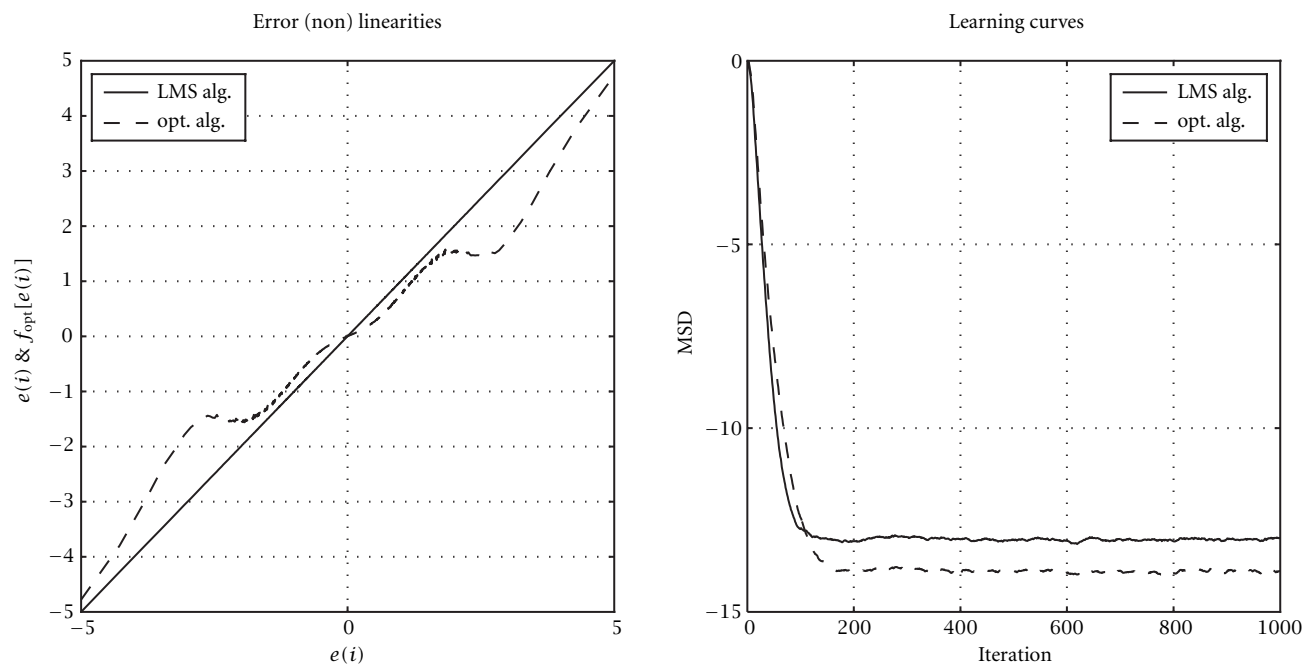


FIGURE 1: Error updates and learning curves for the LMS and optimum algorithm (binary noise case).

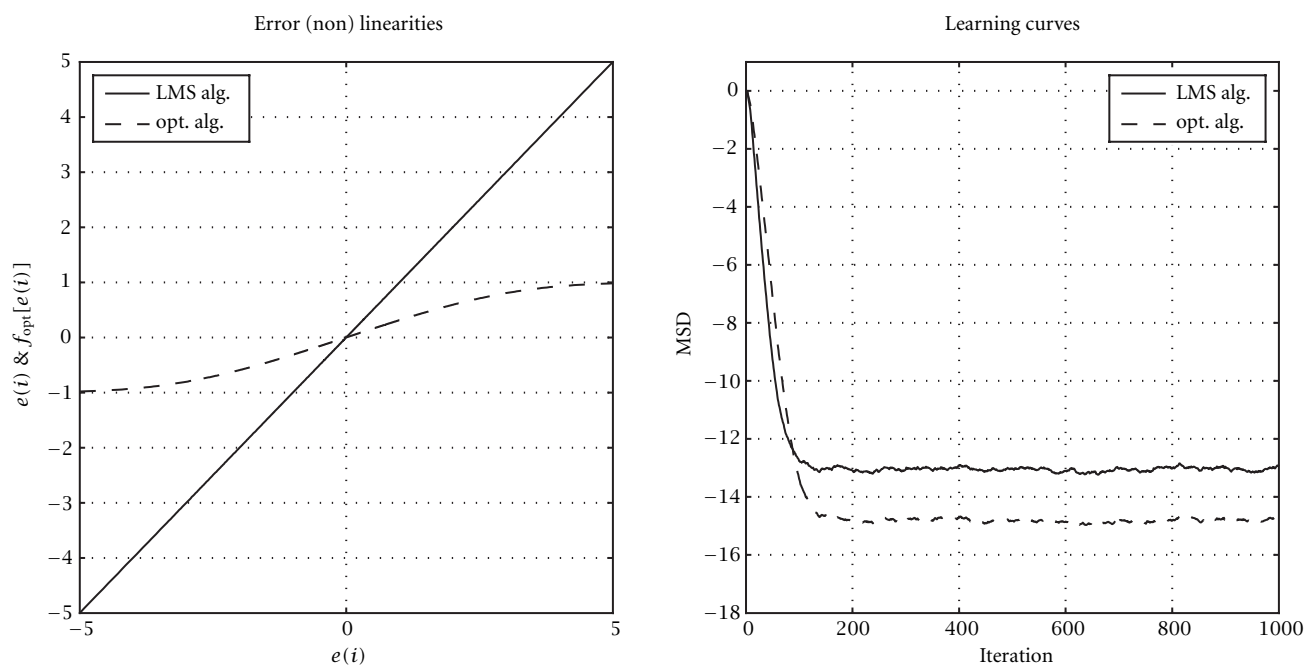


FIGURE 2: Error updates and learning curves for the LMS and optimum algorithm (Laplacian noise case).

of (93) as possible. For most practical purposes, the first few terms of (93) are sufficient for a good approximation [34].

5.5.2 Relating the nonlinearities

With the Edgeworth expansion (93) at hand, we now show how the optimum nonlinearity (76) relates to the (non)linearities of the LMS algorithm and its variants. Using (93), we can represent the optimum nonlinearity (76),

after some straightforward manipulations, as

$$\begin{aligned} f_{\text{opt}}[e(i)] &= -\frac{p'[e(i)]}{p[e(i)]} \\ &= \frac{e(i)}{\sigma_e^2} - \frac{1}{\sigma_e^2} \frac{\sum_{j=2}^{\infty} (2j) a_{2j} \text{He}_{2j-1}[e(i)/\sigma_e]}{\sum_{j=2}^{\infty} a_{2j} \text{He}_{2j}[e(i)/\sigma_e]}. \end{aligned} \quad (95)$$

We can finally put f_{opt} in a more familiar form by approximating the rational function in the last equation by its Taylor series. As the rational function is odd, the Taylor series will contain only odd powers of $e(i)$, and we can, therefore, write

$$f_{\text{opt}}[e(i)] = \sum_{j=0}^{\infty} c_{2j+1} e^{2j+1}(i). \quad (96)$$

The c_{2j+1} 's can be written in terms of the a_{2j} 's but the explicit dependence is not essential for the subsequent discussion. The following remarks are in order.

Remarks. (1) The lowest order term in (96) is that employed by the LMS algorithm. Thus, LMS is a first-order approximation of the optimum nonlinearity, which explains the robustness of the LMS in noisy environments (see also [3] for a deterministic account of the robustness of the LMS). Moreover, the cubic term of (96) corresponds to the error nonlinearity of the LMF algorithm, while the higher order terms are those of the LMF-family.

(2) The approximation (96) also suggests that a mixture of the LMS algorithm and the LMF family of algorithms will outperform any of the individual algorithms as such mixtures represent a better approximation of the optimum nonlinearity. The LMS-LMF mixture (also known as the LMMN algorithm) was actually simulated in [36] and shown to outperform both of the constituent algorithms. The approximation presented here not only justifies such mixtures but can actually be used to design optimal mixtures by calculating the coefficients c_{2j+1} (which, in turn, can be explicitly expressed in terms of the estimation error cumulants, γ_j).

(3) From the above, it follows that the optimum nonlinearity is nothing but an optimal mixture of familiar nonlinearities.

(4) The approximation (96) alleviates the difficulties associated with the implementation of the optimum nonlinearity. In particular, it applies irrespective of the distribution of the estimation error pdf p_e . The approximation also provides for a tradeoff between numerical simplicity and more accurate approximation of the nonlinearity. Notice, however, that (96) still calls for the estimation of the coefficients c_{2j+1} , or, equivalently, of the cumulants γ_i of the estimation error $e(i)$.

(5) By expanding the noise pdf, p_v , in an Edgeworth series similar to (93), we can extend the above remarks to the nonlinearity (82) (see [37]).

6. CONCLUSION

In this paper, we pursued a unified approach to mean-square analysis of adaptive filters with arbitrary error nonlinearities. In particular, starting from an energy conservation relation, we were able to arrive at sufficient conditions for stability without relying on any independence assumptions. Using the same relation we also showed that the MSE is the fixed point of a nonlinear function. This nonlinear expression for the MSE was subsequently used to derive an expression for the optimum nonlinearity.

We would like to emphasize that all our results apply for any error nonlinearity and for arbitrary input color and statistics. They are obtained as a fall out of the same energy conservation relation and rely on weak assumptions that are quite accurate for long enough filters.

ACKNOWLEDGEMENTS

The work of T. Y. Al-Naffouri was partially supported by a fellowship from King Fahd University of Petroleum and Minerals, Saudi Arabia. The work of A. H. Sayed was partially supported by the National Science Foundation under grant ECS-9820765.

REFERENCES

- [1] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmissions*, John Wiley & Sons, New York, 1995.
- [2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 3rd edition, 1996.
- [3] B. Hassibi, A. Sayed, and T. Kailath, " \mathcal{H}^∞ optimality of the LMS algorithm," *IEEE Trans. Signal Processing*, vol. 44, no. 2, pp. 267–280, 1996.
- [4] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Transactions on Information Theory*, vol. 30, no. 2, pp. 275–283, 1984.
- [5] J. D. Gibson and S. D. Gray, "MVSE adaptive filtering subject to a constraint on MSE," *IEEE Trans. Circuits and Systems*, vol. 35, no. 5, pp. 603–608, 1988.
- [6] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 30, no. 4, pp. 578–586, 1982.
- [7] W. A. Sethares, "Adaptive algorithms with nonlinear data and error functions," *IEEE Trans. Signal Processing*, vol. 40, no. 9, pp. 2199–2206, 1992.
- [8] S. C. Douglas and T. H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Processing*, vol. 42, no. 6, pp. 1335–1351, 1994.
- [9] V. J. Mathews and S. H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 4, pp. 450–454, 1987.
- [10] N. J. Bershad, "On error-saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 440–452, 1988.
- [11] N. Bershad and M. Bonnet, "Saturation effects in LMS adaptive echo cancellation for binary data," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1687–1696, 1990.
- [12] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Circuits and Systems*, vol. 28, no. 6, pp. 510–518, 1981.
- [13] W. A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: a general study, analysis and critique," *Signal Processing*, vol. 6, no. 2, pp. 113–133, 1984.
- [14] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 222–229, 1985.
- [15] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Trans. Signal Processing*, vol. 41, no. 3, pp. 1149–1160, 1993.
- [16] 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 *Proceedings of the 32nd Asilomar Conference on Signals, Systems, and Computers*, vol. I, pp. 556–559, Asilomar, CA, November 1998.

[17] S. Koike, "Convergence analysis of a data echo canceller with a stochastic gradient adaptive FIR filter using the sign algorithm," *IEEE Trans. Signal Processing*, vol. 43, no. 12, pp. 2852–2861, 1995.

[18] J. C. M. Bermudez and N. Bershad, "A nonlinear analytical model for the quantized LMS algorithm—the arbitrary step size case," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1175–1183, 1996.

[19] B. Polyak and Y. Tsytkin, "Adaptive estimation algorithms (convergence, optimality, stability)," *Avtomatika i Telemekhanika*, no. 3, pp. 71–84, 1979.

[20] N. J. Bershad, "On the optimum data nonlinearity in LMS adaptation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 69–76, 1986.

[21] T. Y. Al-Naffouri, A. H. Sayed, and T. Kailath, "On the selection of optimal nonlinearities for stochastic gradient adaptive algorithms," in *Proc. ICASSP*, vol. 1, pp. 464–467, Istanbul, Turkey, June 2000.

[22] R. Sharma, W. A. Sethares, and J. A. Bucklew, "Asymptotic analysis of stochastic gradient based adaptive filtering algorithms with general cost functions," *IEEE Trans. Signal Processing*, vol. 44, no. 9, pp. 2186–2194, 1996.

[23] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive gradient algorithms via the small gain theorem," in *Proc. SPIE Conference on Advanced Signal Processing: Algorithms, Architectures, and Implementations*, F. T. Luk, Ed., vol. 2563, pp. 458–469, San Diego, CA, July 1995.

[24] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Processing*, vol. 44, no. 6, pp. 1428–1439, 1996.

[25] A. H. Sayed and M. Rupp, "Robustness issues in adaptive filtering," in *Digital Signal Processing Handbook*, chapter 20, CRC Press, January 1998.

[26] 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, no. 1, pp. 80–91, 2000.

[27] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analysis of adaptive filters," *IEEE Trans. Signal Processing*, vol. 49, no. 2, pp. 314–324, 2001.

[28] N. R. Yousef and A. H. Sayed, "Steady-state and tracking analyses of the sign algorithm without the explicit use of the independence assumption," *IEEE Signal Processing Letters*, vol. 7, no. 11, pp. 307–309, 2000.

[29] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters," in *Proc. ICASSP*, Salt Lake City, Utah, May 2001.

[30] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters—Part I: the data nonlinearity case," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Baltimore, Maryland, June 2001.

[31] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters—Part II: the error nonlinearity case," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Baltimore, Maryland, June 2001.

[32] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*, John Wiley and Sons, New York, 1968.

[33] R. Price, "A useful theorem for nonlinear devices having gaussian inputs," *IEEE Transactions on Information Theory*, vol. 4, no. 6, pp. 69–72, 1958.

[34] A. H. Nuttall, "Evaluation of densities and distributions via Hermite and generalized Laguerre series employing higher-order expansion coefficients determined recursively via moments or cumulants," Tech. Rep. TR 7377, Naval Under-Water Systems Center, February 1985.

[35] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover Publications, Dover, New York, 1970.

[36] O. Tanrikulu and J. A. Chambers, "Convergence and steady-state properties of the least-mean mixed norm (LMMN) adaptive algorithm," in *IEE Proc.-Vision, Image and Signal Processing*, vol. 143, pp. 137–142, June 1996.

[37] T. Y. Al-Naffouri, A. Zerguine, and M. Bettayeb, "A unifying view of error nonlinearities in LMS adaptation," in *Proc. ICASSP*, vol. III, pp. 1697–1700, Seattle, May 1998.

Tareq Y. Al-Naffouri received a BS degree in Mathematics (with honors) and an MS degree in Electrical Engineering from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 1994 and 1997, respectively. He received an MS degree in Electrical Engineering from Georgia Institute of Technology in 1998, and subsequently joined Stanford University where he is a PhD candidate in the Electrical Engineering Department. Mr. Al-Naffouri's research interests include system identification, adaptive filtering analysis and design, echo cancellation, and design under uncertainty. He has recently been interested in channel identification and equalization in OFDM transmission. He is the recipient of a 2001 best student paper award at an international meeting for his work on adaptive filtering analysis.



Ali H. Sayed received the Ph.D. degree in electrical engineering in 1992 from Stanford University, Stanford, CA. He is Professor of Electrical Engineering at the University of California, Los Angeles. He has over 160 journal and conference publications, is coauthor of the research monograph *Indefinite Quadratic Estimation and Control* (SIAM, PA 1999) and of the graduate-level textbook *Linear Estimation* (Prentice Hall, NJ, 2000). He is also coeditor of the volume *Fast Reliable Algorithms for Matrices with Structure* (SIAM, PA 1999). He is a member of the editorial boards of the *SIAM Journal on Matrix Analysis and Its Applications*, of 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, equalization techniques for communications, interplays between signal processing and control methodologies, and fast algorithms for large-scale problems. To learn more about his work, visit the website of the UCLA Adaptive Systems Laboratory at <http://www.ee.ucla.edu/asl>. Dr Sayed is a recipient of the 1996 IEEE Donald G. Fink Award. He is Associate Editor of the *IEEE Transactions on Signal Processing*.

