TRANSIENT ANALYSIS OF ADAPTIVE FILTERS – PART II: THE ERROR NONLINEARITY CASE

TAREQ Y. AL-NAFFOURI[1] AND ALI H. SAYED[2]

[1]Electrical Engineering Department
Stanford University, CA 94305

[2]Electrical Engineering Department
University of California, Los Angeles, CA 90095

ABSTRACT

This paper is the second part of a framework for the mean-square analysis of adaptive filtering algorithms. In contrast to the companion article [1], the focus in this paper is on the class of long adaptive filters with general error nonlinearities. Among other results, the paper characterizes the learning behavior of this class of adaptive filters and derives expressions for steady-state performance. In addition, sufficient conditions for stability, expressed as bounds on the step-size parameter, are provided; these bounds are in terms of the Cramer-Rao bound of the underlying estimation process. The approach of this paper relies on energy conservation arguments and is carried out without restrictions on the input color or statistics.

1. ADAPTIVE FILTERS WITH ERROR NONLINEARITIES

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

$$d(i) = \mathbf{u}_i \boldsymbol{w}^o + v(i), \tag{1}$$

where $\boldsymbol{w}^o$ is an unknown *column* vector of size M that we wish to estimate, $v(i)$ accounts for measurement noise and modeling errors, and $\mathbf{u}_i$ denotes a *row* input (regressor) vector. In this paper, we consider the following class of adaptive algorithms

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

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	$\text{sign}[e(i)]$
Sat. nonlin.	$\int_0^{e(i)} \exp\left(-\frac{z^2}{2\sigma_z^2}\right) dz$

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

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

is the estimation error, and $f[e(i)]$ denotes a scalar function of the error $e(i)$, examples of which can be found in Table 1. The case when the update in (2) is linear in the error and nonlinear in that data is studied in the companion paper [1].

Our aim in this paper is to carry out a parallel study to that in [1] for the class of adaptive algorithms (2)-(3). There are many studies of such algorithms in the literature under varied assumptions (e.g., see [2]-[5]). In this paper, we pursue a unifying framework for adaptive filter analysis that does away with many of these assumptions. Our approach relies on energy conservation arguments developed in [6]–[9] and in the companion article [1].

2. PRELIMINARIES: DEFINITIONS AND NOTATION

The techniques that are used here are similar to those in the companion paper [1]. We shall therefore be brief. Analysis of (2)-(3) is best carried out in terms of the

Published in Proc. 5th IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, Baltimore, Maryland, June 2001. This work was partially supported by the National Science Foundation under award ECS-9820765. The work of T. Y. Al-Naffouri was also partially supported by a fellowship from King Fahd University of Petroleum and Minerals, Saudi Arabia.

following error quantities:

$\tilde{\boldsymbol{w}}_i \triangleq \boldsymbol{w}^o - \boldsymbol{w}_i$	weight-error vector
$e_a^{\Sigma}(i) \triangleq \mathbf{u}_i \boldsymbol{\Sigma} \tilde{\boldsymbol{w}}_i$	weighted a priori error
$e_p^{\Sigma}(i) \triangleq \mathbf{u}_i \boldsymbol{\Sigma} \tilde{\boldsymbol{w}}_{i+1}$	weighted a posteriori error.

where $\boldsymbol{\Sigma}$ is a weighting matrix. We reserve special notation for the case $\boldsymbol{\Sigma} = \boldsymbol{I}$: $e_a(i) = e_a^I(i)$ and $e_p(i) = e_p^I(i)$. We will also find it convenient to introduce the following notation for the weighted sum of squares:

$$\|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 \triangleq \tilde{\boldsymbol{w}}_i^T \boldsymbol{\Sigma} \tilde{\boldsymbol{w}}_i$$

For one reason, this notation is convenient because it enables us to transform many operations on $\tilde{\boldsymbol{w}}_i$ into operations on the norm subscript (see [1] for a list of such operations).

The estimation errors $e_a^{\Sigma}(i)$, $e_p^{\Sigma}(i)$, and $e(i)$ can be related by premultiplying both sides of the adaptation equation (2) by $\mathbf{u}_i \boldsymbol{\Sigma}$

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

and incorporating the defining expressions for $e_p^{\Sigma}(i)$ and $e_a^{\Sigma}(i)$. This process leads to

$$\boxed{e_p^{\Sigma}(i) = e_a^{\Sigma}(i) - \mu \|\mathbf{u}_i\|_{\Sigma}^2 f[e(i)]} \tag{4}$$

If we now solve for $f[(e(i)]$ from the above equation, substitute into the recursion for $\tilde{\boldsymbol{w}}_{i+1}$, and compare energies, we arrive at the following relation (see [1, 8, 9]):

$$\boxed{\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2 + \frac{\left|e_a^{\Sigma}(i)\right|^2}{\|\mathbf{u}_i\|_{\Sigma}^2} = \|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 + \frac{\left|e_p^{\Sigma}(i)\right|^2}{\|\mathbf{u}_i\|_{\Sigma}^2}} \tag{5}$$

No assumptions or approximations are used to derive this energy relation, which applies to any adaptation algorithm of the form (2)-(3). The unweighted version of this recursion was originally developed in [6, 7] in the context of deterministic analysis of adaptive filters, and subsequently used in [8, 9] to study the steady-state performance of various adaptive filtering and equalization algorithms.

3. DYNAMICAL BEHAVIOR OF THE WEIGHT-ERROR VECTOR

Starting from the energy conservation relation (5) and replacing the a- posteriori error $e_p^{\Sigma}(i)$ by its equivalent expression (4) leads to

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

or, upon taking the expectation of both sides,

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2\right] - 2\mu \overbrace{E\left[e_a^{\Sigma}(i) f[e(i)]\right]}^{①}$$
$$+\mu^2 \overbrace{E\left[\|\mathbf{u}_i\|_{\Sigma}^2 f^2[e(i)]\right]}^{②} \tag{6}$$

Now, two expectations call for evaluation. As in [1], this is facilitated by the following assumption on the noise sequence:

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

3.1. Evaluating ①

To evaluate the first expectation, we further assume that the adaptive filter is long enough such that

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

This assumption is justified for long filters by central limit arguments. Assumption AG can now be used together with the independence assumption AN and Price Theorem to write

$$① \triangleq E\left[e_a^{\Sigma} f[e_a(i) + v(i)]\right] \tag{7}$$
$$= E\left[e_a^{\Sigma}(i) e_a(i)\right] \frac{E\left[e_a(i) f[e_a(i) + v(i)]\right]}{E[e_a^2(i)]} \tag{8}$$

Since $e_a(i)$ is Gaussian, the expectation $E[e_a(i) f(e(i))]$ in (8) depends on $e_a(i)$ through *the second moment* $E[e_a^2(i)]$ *only*. This fact motivates the following definition:[1]

$$h_G\left[E[e_a^2(i)]\right] \triangleq \frac{E\left[e_a(i) f[e(i)]\right]}{E[e_a^2(i)]} \tag{9}$$

For future reference, h_G is evaluated for the algorithms of Table 1 and the results are shown in Table 2.
Combining (8) and (9) yields

$$① = E\left[e_a^{\Sigma}(i) e_a(i)\right] h_G\left[E[e_a^2(i)]\right]$$
$$= E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma u_i^T u_i}^2\right] h_G\left[E[e_a^2(i)]\right] \tag{10}$$

[1]The Gaussianity assumption AG is the major assumption leading to the defining expression (9) for h_G, hence the subscript G. The subscript U of h_U, which is defined in (12), is similarly motivated.

Table 1: $h_G[\cdot]$ *for the error nonlinearities of Table 1* ($\sigma_e^2 \triangleq E[e_a^2(i)]$.)

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

3.2. Evaluating ②

To handle the second expectation in (6), we use the long filter assumption:

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

The assumption becomes more realistic as the filter gets longer. It enables us to split the expectation② as

$$E\left[\|\mathbf{u}_i\|_\Sigma^2 f^2[e(i)]\right] = E\left[\|\mathbf{u}_i\|_\Sigma^2\right] E\left[f^2[e(i)]\right] \tag{11}$$

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

$$h_U\left[E[e_a^2(i)]\right] \triangleq E\left[f^2[e(i)]\right] \tag{12}$$

which together with (11) yields

$$\boxed{\textcircled{2} = E\left[\|\mathbf{u}_i\|_\Sigma^2\right] h_U\left[E\left[e_a^2(i)\right]\right]} \tag{13}$$

The function h_U is evaluated for the algorithms of Table 1 and the results appear in Table 3 in the same order.

By substituting (10) and (13) into (6), we obtain

$$\begin{aligned} E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_\Sigma^2\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2\right] - 2\mu h_G\left[E[e_a^2(i)]\right] E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma u_i^T u_i}\right] \\ &\quad + \mu^2 E\left[\|\mathbf{u}_i\|_\Sigma^2\right] h_U\left[E[e_a^2(i)]\right] \end{aligned} \tag{14}$$

3.3. Learning Curves

To construct the learning curves, we impose the independence assumption on the data:

AI The sequence $\mathbf{u}_i$ is independent with zero mean and autocorrelation matrix $\boldsymbol{R}$.

Table 2: $h_U[\cdot]$ *for the error nonlinearities of Table 1* ($\sigma_e^2 \triangleq E[e_a^2(i)]$.)

$h_U[\sigma_e^2]$
$15\sigma_e^6 + 45\sigma_e^4\sigma_v^2 + 15\sigma_e^2 E\left[v^4(i)\right] + E\left[v^6(i)\right]$
$\sum_{j=0}^{2k+1} \binom{4k+2}{2j} \frac{(2j)!}{2^j j!}\sigma_e^{2j} E\left[v^{2(2k-j+1)}(i)\right]$
$15b^2\sigma_e^6 + (45b^2\sigma_v^2 + 6ab)\sigma_e^4$ $+(15b^2 E\left[v^4(i)\right] + 12ab\sigma_v^2 + a^2)\sigma_e^2$ $+E\left[(bv^2(i)+a)^2 v^2(i)\right]$
1
$\frac{\pi}{2}\sigma_z^2 - 2\sigma_z^3 \int_0^{\frac{1}{\sqrt{2}}} \frac{1}{\sqrt{\sigma_e^2+\sigma_z^2(1-x^2)}} E\left[e^{-\frac{v^2(i)}{2(\sigma_e^2+\sigma_z^2(1-x^2))}}\right] dx$

Under this assumption, recursion (14) takes the more homogeneous form

$$\begin{aligned} E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_\Sigma^2\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2\right] - 2\mu h_G\left[E\left[\|\tilde{\boldsymbol{w}}_i\|_R^2\right]\right] E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma R}^2\right] \\ &\quad + \mu^2 E\left[\|\mathbf{u}_i\|_\Sigma^2\right] h_U\left[E\left[\|\tilde{\boldsymbol{w}}_i\|_R^2\right]\right] \end{aligned} \tag{15}$$

Notice that (15) is not self-contained. However, we can take advantage of the free parameter $\boldsymbol{\Sigma}$ to go around this problem. Let us in particular write (15) for the choices $\boldsymbol{\Sigma} = \boldsymbol{I}, \boldsymbol{R}, \cdots, \boldsymbol{R}^{M-1}$ (the arguments of the functions h_G and h_U remain the same (i.e., $E\left[\|\tilde{\boldsymbol{w}}_i\|_R^2\right]$) regardless of the choice of $\boldsymbol{\Sigma}$ and are therefore suppressed for convenience of notation):

$$\left\{ \begin{aligned} E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_I^2\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|_I^2\right] - 2\mu h_G E\left[\|\tilde{\boldsymbol{w}}_i\|_R^2\right] \\ &\quad + \mu^2 E\left[\|\mathbf{u}_i\|_I^2\right] h_U \\ E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_R^2\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|_R^2\right] - 2\mu h_G E\left[\|\tilde{\boldsymbol{w}}_i\|_{R^2}^2\right] \\ &\quad + \mu^2 E\left[\|\mathbf{u}_i\|_R^2\right] h_U \\ &\;\;\vdots \\ E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_{R^{M-1}}^2\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|_{R^{M-1}}^2\right] - 2\mu h_G E\left[\|\tilde{\boldsymbol{w}}_i\|_{R^M}^2\right] \\ &\quad + \mu^2 E\left[\|\mathbf{u}_i\|_{R^{M-1}}^2\right] h_U \end{aligned} \right. \tag{16}$$

Now the additional variable $E\left[\|\tilde{\boldsymbol{w}}_i\|_{R^M}^2\right]$ in the last equation of (16) can be expressed in terms of the lower-order variables. In particular, by the Cayley-Hamilton theorem, we can write

$$\boldsymbol{R}^M = -p_0\boldsymbol{I} - p_1\boldsymbol{R} - \cdots - p_{M-1}\boldsymbol{R}^{M-1} \tag{17}$$

where

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

is the characteristic polynomial of $\boldsymbol{R}$. This induces the desired "order-reducing" relation

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

and enables us to rewrite the last equation in (16) as

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

The system (16) now becomes truly self-contained as can be seen from the equivalent state-space form

$$\boxed{\mathcal{W}_{i+1} = \mathcal{A}\mathcal{W}_i + \mu^2\mathcal{Y}} \tag{18}$$

where

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

and

$$\mathcal{A} = \begin{bmatrix} 1 & -2\mu h_G & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -2\mu h_G \\ 2\mu p_0 h_G & 2\mu p_1 h_G & \cdots & 1 + 2\mu p_{M-1} h_G \end{bmatrix} \tag{19}$$

This state-space model is nonlinear but time-invariant. It describes the evolution of the weight-error energy.

4. STEADY-STATE ANALYSIS

Our starting point for steady-state analysis is the averaged energy relation (14) reproduced in an equivalent form here:

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma}\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma}\right] - 2\mu E\left[e_a^{\Sigma}(i)e_a(i)\right] h_G\left[E[e_a^2(i)]\right]$$
$$+\mu^2 E\left[\|\mathbf{u}_i\|^2_{\Sigma}\right] h_U\left[E[e_a^2(i)]\right] \tag{20}$$

Assuming the filter is stable, the energy of the weight-error vector eventually reaches a steady-state value, i.e.

$$\lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\Sigma}\right] = \lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma}\right] \tag{21}$$

Then, in the limit, (20) becomes

$$\lim_{i\to\infty} E[e_a^{\Sigma}(i)e_a(i)] = \frac{\mu E\left[\|\mathbf{u}_i\|^2_{\Sigma}\right]}{2} \frac{\lim_{i\to\infty} h_U\left[E\left[e_a^2(i)\right]\right]}{\lim_{i\to\infty} h_G[E[e_a^2(i)]]} \tag{22}$$

Now, let S denote the (asymptotic) value of the mean-square error (MSE), i.e.

$$S = \lim_{i\to\infty} E[e_a^2(i)] \tag{23}$$

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

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

and, accordingly, (22) can be written more compactly as

$$\boxed{\lim_{i\to\infty} E\left[e_a^{\Sigma}(i)e_a(i)\right] = \frac{\mu}{2} E\left[\|\mathbf{u}_i\|^2_{\Sigma}\right] \frac{h_U[S]}{h_G[S]}} \tag{24}$$

This relation can now be used to calculate various steady-state measures. In particular, to calculate the MSE we employ (24) with $\mathbf{\Sigma}$ set to the identity matrix to get

$$\lim_{i\to\infty} E[e_a^2] = \frac{\mu}{2}\mathrm{Tr}\left(\boldsymbol{R}\right)\frac{h_U[S]}{h_G[S]}.$$

Or, since $S = \lim_{i\to\infty} E[e_a^2]$, the MSE is the positive solution of the equation

$$\boxed{S = \frac{\mu}{2}\mathrm{Tr}\left(\boldsymbol{R}\right)\frac{h_U[S]}{h_G[S]}} \tag{25}$$

i.e., the MSE is a *fixed point* of the function $\frac{\mu}{2}\mathrm{Tr}\left(\boldsymbol{R}\right)\frac{h_U[S]}{h_G[S]}$. For a given error nonlinearity, we can evaluate h_U and h_G, as done in Tables 1 and 2, and proceed to calculate the MSE. The mean-square deviation (MSD)

$$\mathsf{MSD} \triangleq \lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|^2\right]$$

is calculated by setting $\mathbf{\Sigma} = \boldsymbol{R}^{-1}$ in (24) and invoking the independence assumption AI to write

$$\mathsf{MSD} \triangleq \lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|^2\right] = \lim_{i\to\infty} E\left[e_a^{R^{-1}}(i)e_a(i)\right] = \frac{\mu M}{2}\frac{h_U[S]}{h_G[S]}$$

Upon comparing this relation with the corresponding one for the MSE (25), we see that the MSD can also be calculated from

$$\boxed{\mathsf{MSD} = \frac{M}{\mathrm{Tr}(\boldsymbol{R})}\mathsf{MSE}} \tag{26}$$

5. STABILITY CONDITIONS

Consider the energy relation (6) for $\mathbf{\Sigma} = \boldsymbol{I}$,

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

This shows that if we choose μ such that

$$\mu \le \mu_{\max} = 2\inf_i \frac{E\left[e_a(i)f[e(i)]\right]}{E\left[\|\mathbf{u}_i\|^2 f^2[e(i)]\right]} \tag{27}$$

then the sequence $\{E\left[\|\tilde{\boldsymbol{w}}_i\|^2\right]\}$ would be decreasing and (being bounded from below) also convergent. The difficulty is to calculate $\mu_{\max}$, or any nonnegative lower bound for that matter. Using the Cauchy-Schwartz inequality, we can write

$$\begin{aligned}\frac{E\left[e_a(i)f[e(i)]\right]}{E\left[\|\mathbf{u}_i\|^2 f^2[e(i)]\right]} &\geq \frac{1}{\left[E\|\mathbf{u}_i\|^4\right]^{\frac{1}{2}}}\frac{E\left[e_a(i)f[e(i)]\right]}{E\left[f^4[e(i)]\right]^{\frac{1}{2}}} \\ &= \frac{1}{\left[E\|\mathbf{u}_i\|^4\right]^{\frac{1}{2}}}\frac{h_G\left[E[e_a^2(i)]\right]}{h_C\left[E[e_a^2(i)]\right]} \qquad (28)\end{aligned}$$

where the second line is obtained by invoking the Gaussian assumption on $e_a(i)$ to write $h_G\left[E[e_a^2(i)]\right] = E\left[e_a(i)f[e(i)]\right]$ (as defined in (9)) and $h_C\left[E[e_a^2(i)]\right] = E\left[f^4[e(i)]\right]$. Thus, a more conservative bound on μ is

$$\mu \leq \frac{1}{\left[E\|\mathbf{u}_i\|^4\right]^{\frac{1}{2}}}\inf_{E[e_a^2(i)]}\frac{h_G\left[E[e_a^2(i)]\right]}{h_C\left[E[e_a^2(i)]\right]} \qquad (29)$$

Minimizing over $E\left[e_a^2(i)\right]$ is difficult. Instead, we carry out the minimization over a larger feasibility set. Notice first that $E\left[e_a^2(i)\right]$ is lower bounded by the Cramer-Rao bound γ associated with the underlying estimation process (estimating $\mathbf{u}_i\boldsymbol{w}^o$ as $\mathbf{u}_i\boldsymbol{w}_i$):

$$\boxed{E\left[e_a^2\right] \geq \gamma} \qquad (30)$$

To obtain an upper bound on $E\left[e_a^2(i)\right]$, observe that if μ is chosen to satisfy (27) for all i, then the sequence $E\left[\|\tilde{\boldsymbol{w}}_i\|^2\right]$ will be monotonically decreasing. In other words, $E\left[\|\tilde{\boldsymbol{w}}_i\|^2\right] \leq E\left[\|\tilde{\boldsymbol{w}}_0\|^2\right]$. This fact together with the Gaussian assumption on $e_a(i)$ produces the upper bound

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

In other words,

$$\boxed{E\left[e_a^2(i)\right] \leq \tfrac{1}{4}\mathrm{Tr}\,(\boldsymbol{R})\,E[\|\tilde{\boldsymbol{w}}_o\|^2]} \qquad (31)$$

The bounds (30) and (31) suggest the feasibility set

$$\Omega = \left\{E[e_a^2]:\ \gamma \leq E[e_a^2] \leq \frac{1}{4}\mathrm{Tr}\,(\boldsymbol{R})\,E\left[\|\tilde{\boldsymbol{w}}_0\|^2\right]\right\} \qquad (32)$$

which leads to the following sufficient condition for stability

$$\boxed{\mu \leq \frac{2}{\left[E\|\mathbf{u}_i\|^4\right]^{\frac{1}{2}}}\left(\inf_{E[e_a^2]\epsilon\Omega}\frac{h_G\left[E[e_a^2]\right]}{h_C\left[E[e_a^2]\right]}\right)} \qquad (33)$$

Notice that no independence assumptions were used to derive this bound. For a given error nonlinearity, we can obtain explicit expressions for h_G and h_C and subsequently calculate the bound (33).

6. CONCLUSION

In this paper we employed energy-conservation arguments to perform mean-square analysis of adaptive filters with error nonlinearities. In particular, we have constructed a state-space model that characterizes the mean-square behavior of this class of algorithms. We have also shown how to calculate the steady-state error for an adaptive filter in this class by finding the fixed point of some function. We finally provided a bound on the step size for stability without invoking any independence arguments. Our results apply for any error nonlinearity and for any input color and statistics.

Our study is centered around a weighted energy relation. The relation was used in the first part of this work [1] to carry out a parallel study of the class of adaptive filters with data nonlinearities.

7. REFERENCES

[1] T. Y. Al-Naffouri and A. H. Sayed. Mean-square analysis of adaptive filters – Part I: The data nonlinearity case, *Proc. 5th IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing,* Baltimore, Maryland, June 2001.

[2] D. L. Duttweiler. Adaptive filter performance with nonlinearities in the correlation multiplier, *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 30, no. 4, pp. 578-586, Aug. 1982.

[3] 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, Aug. 1984.

[4] N. Bershad, and M. Bonnet, Saturation effects in LMS adaptive echo cancellation for binary data, *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 38, no. 10, pp. 1687-1696, Oct. 1990.

[5] V. Mathews and S. Cho. Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm. *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. 35, no. 4, pp. 450–454, Apr. 1987.

[6] A. H. Sayed and M. Rupp. A time-domain feedback analysis of adaptive algorithms via the small gain theorem. *Proc. SPIE,* vol. 2563, pp. 458–469, San Diego, CA, Jul. 1995.

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

[8] J. Mai and A. H. Sayed. A feedback approach to the steady-state performance of fractionally-spaced blind adaptive equalizers, *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 80–91, Jan. 2000

[9] N. R. Yousef and A. H. Sayed. A unified approach to the steady-state and tracking analyses of adaptive filters, *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314–324, Feb. 2001.