TRANSIENT ANALYSIS OF ADAPTIVE FILTERS – PART I: THE DATA 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 develops a framework for the mean-square analysis of adaptive filters with general data nonlinearities. The approach relies on energy conservation arguments and is carried out without restrictions on the input color or statistics. Among other results, the paper characterizes the learning behavior of adaptive filters with diagonal matrix nonlinearities. It also provides closed form expressions for the steady-state performance, and necessary and sufficient conditions for mean-square stability. This study encompasses earlier results and addresses some open issues. A companion article studies the case of adaptive filters with error nonlinearities.

1. INTRODUCTION

Adaptive filters are, by design, time-variant and nonlinear systems that adapt to variations in signal statistics and that learn from their interactions with the environment. The success of their learning mechanism can be measured in terms of how fast they adapt to changes in the signal characteristics and how well they can learn given sufficient time.

There have been extensive works in the literature on the performance of adaptive filters with many ingenious results and approaches (e.g., [1]–[9]). However, it is generally observed that most of these works study individual algorithms separately. This is because different adaptive schemes have different nonlinear update equations, and the particularities of each case tend to require different arguments and assumptions.

In this paper and the companion paper [10], we provide a unified approach to the mean-square analysis of adaptive filtering algorithms. The approach is based on studying the energy flow through each iteration of an adaptive filter, and it relies on a fundamental energy conservation relation (8) that holds for a large class of adaptive filters. The unweighted version of this relation was originally developed in [11]–[12] in the context of robustness analysis of adaptive filters within a deterministic framework. It has since then been used in [13, 14] as a convenient tool for studying the steady-state performance of adaptive filters.

2. DATA MODEL AND ALGORITHMS

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 and the companion paper [10] we study adaptive filters of the form

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

where $\boldsymbol{w}_i$ is an estimate for $\boldsymbol{w}^o$ at iteration i, μ is the step-size, $f[e(i)]$ is a generic error nonlinearity, $\boldsymbol{H}[\mathbf{u}_i]$ is a generic data nonlinearity, and

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

is the estimation error. We confine our attention here to updates that are linear in the error, i.e., we set $f[e(i)] \equiv e(i)$, while [10] studies the dual problem of linear data updates (for which $\boldsymbol{H}[\mathbf{u}_i] \equiv \boldsymbol{I}$).

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.

3. PRELIMINARIES: DEFINITIONS AND NOTATION

Mean square analysis of (2)-(3) is best carried out in terms of the normalized regressor $\overline{\mathbf{u}}_i = \mathbf{u}_i \boldsymbol{H}[\mathbf{u}_i]$ and the 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 $e_p(i) = e_p^I(i)$. The defining relations (2)-(3) can now be rewritten in terms of these quantities as

$$\tilde{\boldsymbol{w}}_{i+1} = \tilde{\boldsymbol{w}}_i - \mu f[e(i)] \overline{\mathbf{u}}_i^T \quad (4)$$

$$e(i) = e_a(i) + v(i) \quad (5)$$

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, as demonstrated by the following properties:

1) Superposition.

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

2) Polarization.

$$(\mathbf{u}_i \boldsymbol{\Sigma}_1 \tilde{\boldsymbol{w}}_i)(\mathbf{u}_i \boldsymbol{\Sigma}_2 \tilde{\boldsymbol{w}}_i) = \|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1 u_i^T u_i \Sigma_2}^2 \quad (6)$$

3) Independence. If $\tilde{\boldsymbol{w}}_i$ and $\mathbf{u}_i$ are independent,

$$E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1 u_i^T u_i \Sigma_2}^2\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1 E[u_i^T u_i] \Sigma_2}^2\right]$$

4) Linear transformation. $\|\boldsymbol{A}\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 = \|\tilde{\boldsymbol{w}}_i\|_{A^T \Sigma A}^2$

5) Blindness to asymmetry.

$$\|\tilde{\boldsymbol{w}}_i\|_A^2 = \|\tilde{\boldsymbol{w}}_i\|_{A^T}^2 = \|\tilde{\boldsymbol{w}}_i\|_{\frac{1}{2}A + \frac{1}{2}A^T}^2$$

6) Notational convention. $\|\tilde{\boldsymbol{w}}_i\|_{\mathrm{vec}(\Sigma_1)}^2 \triangleq \|\tilde{\boldsymbol{w}}_i\|_{\Sigma_1}^2$

With the above notation and definitions at hand, we can easily derive a relation between the energies of $\{\tilde{\boldsymbol{w}}_i, \tilde{\boldsymbol{w}}_{i+1}, e_a^{\Sigma}(i), e_p^{\Sigma}(i)\}$. To this end, premultiply both sides of (4) by $\mathbf{u}_i \boldsymbol{H}[\mathbf{u}_i] \boldsymbol{\Sigma}$ to get

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

Using the defining expressions for $\overline{\mathbf{u}}_i$, $e_a^{(\cdot)}$, $e_p^{(\cdot)}$, and solving for $\mu f[e(i)]$, we get

$$\boxed{\mu f[e(i)] = \frac{e_a^{H\Sigma}(i)}{\|\overline{\mathbf{u}}_i\|_{\Sigma}^2} - \frac{e_p^{H\Sigma}(i)}{\|\overline{\mathbf{u}}_i\|_{\Sigma}^2}} \quad (7)$$

We now use (7) to eliminate $\mu f[e(i)]$ from (4) and compute the $\boldsymbol{\Sigma}$-weighted norm of both sides of the resulting expression to get

$$\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2 = \left\| \tilde{\boldsymbol{w}}_i - \frac{1}{\|\overline{\mathbf{u}}_i\|_{\Sigma}^2} \left(e_a^{H\Sigma}(i) - e_p^{H\Sigma}(i) \right) \overline{\mathbf{u}}_i^T \right\|_{\Sigma}^2$$

We can equivalently write after some algebra

$$\boxed{\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2 + \frac{\left|e_a^{H\Sigma}(i)\right|^2}{\|\overline{\mathbf{u}}_i\|_{\Sigma}^2} = \|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2 + \frac{\left|e_p^{H\Sigma}(i)\right|^2}{\|\overline{\mathbf{u}}_i\|_{\Sigma}^2}} \quad (8)$$

No assumptions or approximations were used to derive this energy relation, which applies to any adaptation algorithm of the form (2)-(3). Relation (8) will be the starting point for much of the subsequent discussion of this paper and of the companion paper [10].

Remark. This relation plays a role similar to Snell's law in physics, which governs the propagation of a ray from one medium to another with different refraction indices. To see this, consider for simplicity the case $\boldsymbol{\Sigma} = \boldsymbol{I}$ and $\boldsymbol{H}(\mathbf{u}_i) \equiv \boldsymbol{I}$. Let $\{\theta_i, \theta_{i+1}\}$ denote the angles between the regressor $\mathbf{u}_i$ and the weight-error vectors $\{\tilde{\boldsymbol{w}}_i, \tilde{\boldsymbol{w}}_{i+1}\}$. Then (8) reduces to

$$\|\tilde{\boldsymbol{w}}_i\|^2 \sin^2(\theta_i) = \|\tilde{\boldsymbol{w}}_{i+1}\|^2 \sin^2(\theta_{i+1})$$

which suggests that $\{\|\tilde{\boldsymbol{w}}_i\|, \|\tilde{\boldsymbol{w}}_{i+1}\|\}$ play the role of refraction indices and $\mathbf{u}_i$ plays the role of an incident ray.

4. ADAPTIVE FILTERS WITH DATA NONLINEARITIES

In the remainder of this paper, we concentrate on adaptive filters with updates that are nonlinear in the data only, i.e. we set $f[e(i)] \equiv e(i)$. Table 1 lists some common examples of data nonlinearities. The dual case when the update is nonlinear in the error but linear in the data is considered in the second part of this work [10].

In our analysis the following assumptions are needed:

AN The noise $v(i)$ is i.i.d. and independent of the input.

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

ALGORITHM	DATA NONLINEARITIES $\boldsymbol{H}[\mathbf{u}_i]$
LMS	$\boldsymbol{I}$
NLMS	$\frac{1}{\|\mathbf{u}_i\|^2}\boldsymbol{I}$
ϵ-NLMS	$\frac{1}{\epsilon+\|\mathbf{u}_i\|^2}\boldsymbol{I}$
Sign regressor	$\text{diag}\left(\frac{sgn(u_{i_1})}{u_{i_1}},\ldots,\frac{sgn(u_{i_M})}{u_{i_M}}\right)$
Variable step	$\text{diag}(\mu_1,\mu_2,\ldots,\mu_M)$
Power normalized LMS	$\text{diag}\,(p_1(i),p_2(i),\ldots,p_M(i))$ $p_k(i+1)=\beta p_k(i)+(1-\beta)\lvert u_{i_k}\rvert^2$

Notice that we impose no restriction on the input color or statistics.

To perform mean-square analysis, we start with (7), which in the linear error case $f[e(i)]=e(i)$ reads

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

Substituting this expression into the energy relation (8) leads to

$$\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2=\|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2-2\mu e_a^{H\Sigma}(i)e(i)+\mu^2\|\overline{\mathbf{u}}_i\|_{\Sigma}^2e^2(i) \quad (9)$$

By incorporating (5) and assumption AN, (9) becomes under expectation

$$\begin{aligned}E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2\right]-2\mu E\left[e_a^{H\Sigma}(i)e_a(i)\right] \quad (10)\\ &\quad+\mu^2E\left[e_a^2(i)\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\right]+\mu^2\sigma_v^2E\left[\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\right]\end{aligned}$$

Using the weighted-norm properties, we can rewrite the estimation error expectations in (10) as some weighted norms of $\tilde{\boldsymbol{w}}_i$:

$$2e_a(i)e_a^{H\Sigma}(i) = \|\tilde{\boldsymbol{w}}_i\|^2_{u_i^T\overline{u}_i\Sigma+\Sigma\overline{u}_i^Tu_i} \quad (11)$$

$$e_a^2\|\overline{\mathbf{u}}_i\|_{\Sigma}^2 = \|\tilde{\boldsymbol{w}}_i\|^2_{u_i^T\|\overline{u}_i\|_{\Sigma}^2u_i} \quad (12)$$

Substituting (11)-(12) into (10) and using assumption AI yields

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2\right]=E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma}^2\right]+\mu^2E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{E\left[u_i^T\|\overline{u}_i\|_{\Sigma}^2u_i\right]}\right]$$

$$-\mu E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{E[u_i\overline{u}_i]\Sigma+\Sigma E[\overline{u}_iu_i]}\right]+\mu^2\sigma_v^2E\left[\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\right]$$

or, more compactly,

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_{\Sigma}^2\right]=E\left[\|\tilde{\boldsymbol{w}}_i\|_{\Sigma'}^2\right]+\mu^2\sigma_v^2E\left[\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\right] \quad (13)$$

where

$$\begin{aligned}\boldsymbol{\Sigma}'=\boldsymbol{\Sigma}&-\mu\boldsymbol{\Sigma}E\left[\overline{\mathbf{u}}_i^T\mathbf{u}_i\right]-E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\boldsymbol{\Sigma}\\ &+\mu^2E[\mathbf{u}_i^T\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\mathbf{u}_i] \quad (14)\end{aligned}$$

This recursion represents a linear relation between the elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}'$. This can be further clarified by applying the vec operation to both sides of (14)

$$\begin{aligned}\text{vec}(\boldsymbol{\Sigma}')=\text{vec}(\boldsymbol{\Sigma})&-\mu\text{vec}(\boldsymbol{\Sigma}E\left[\overline{\mathbf{u}}_i^T\mathbf{u}_i\right])-\mu\text{vec}(E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\boldsymbol{\Sigma})\\ &+\mu^2\text{vec}(E[\mathbf{u}_i^T\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\mathbf{u}_i]) \quad (15)\end{aligned}$$

and using the Kronecker product notation to write

$$\text{vec}\left(\boldsymbol{\Sigma}E\left[\overline{\mathbf{u}}_i^T\mathbf{u}_i\right]\right) = \left(E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\otimes\boldsymbol{I}_M\right)\boldsymbol{\sigma} \quad (16)$$

$$\text{vec}\left(E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\boldsymbol{\Sigma}\right) = \left(\boldsymbol{I}_M\otimes E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\right)\boldsymbol{\sigma} \quad (17)$$

and

$$\begin{aligned}\text{vec}\left(E\left[\mathbf{u}_i^T\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\mathbf{u}_i\right]\right) &= E\left[\text{vec}\left(\mathbf{u}_i^T\|\overline{\mathbf{u}}_i\|_{\Sigma}^2\mathbf{u}_i\right)\right]\\ &= E\left[\text{vec}\left(\mathbf{u}_i^T\overline{\mathbf{u}}_i\boldsymbol{\Sigma}\overline{\mathbf{u}}_i^T\mathbf{u}_i\right)\right]\\ &= E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\otimes\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\boldsymbol{\sigma} \quad (18)\end{aligned}$$

where

$$\boldsymbol{\sigma}=\text{vec}(\boldsymbol{\Sigma}) \text{ and } \boldsymbol{\sigma}'=\text{vec}(\boldsymbol{\Sigma}')$$

Substituting (16)-(18) into (15) finally yields

$$\boxed{\boldsymbol{\sigma}'=\boldsymbol{F}\boldsymbol{\sigma}} \quad (19)$$

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

$$\begin{aligned}\boldsymbol{F}\triangleq\boldsymbol{I}_{M^2} &- \mu E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\otimes\boldsymbol{I}_M-\mu\boldsymbol{I}_M\otimes E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\\ &+ \mu^2E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\otimes\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\end{aligned}$$

or

$$\boldsymbol{F}=E\left[\left(\boldsymbol{I}_M-\mu\mathbf{u}_i^T\overline{\mathbf{u}}_i\right)\otimes\left(\boldsymbol{I}_M-\mu\mathbf{u}_i^T\overline{\mathbf{u}}_i\right)\right] \quad (20)$$

Therefore, in terms of the vec notation, (13) takes the form

$$\boxed{E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_{\sigma}^2\right]=E\left[\|\tilde{\boldsymbol{w}}_i\|_{F\sigma}^2\right]+\mu^2\sigma_v^2E\left[\|\overline{\mathbf{u}}_i\|_{\sigma}^2\right]} \quad (21)$$

This recursion is not self-contained, i.e., it can not be propagated in time. We can resolve this issue by expanding the dimension of the recursion and using the Caley-Hamilton theorem to construct the following state-space model (which characterizes the dynamical behavior of adaptive filters with data nonlinearities):

$$\boxed{\mathcal{W}_{i+1}=\mathcal{A}\mathcal{W}_i+\mu^2\mathcal{Y}} \quad (22)$$

where

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

and

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

where the $\{p_i\}$ denote the coefficients of the characteristic polynomial of $\boldsymbol{F}$:

$$\begin{aligned} p(x) &\triangleq \det(x\boldsymbol{I} - \boldsymbol{F}) \\ &= p_0 + p_1 x + \cdots + p_{M^2-1} x^{M^2-1} + x^{M^2} \end{aligned}$$

We are now ready to construct the learning curves and to perform stability and steady-state analysis.

4.1. Learning Curves

The learning curve of an adaptive filter describes the evolution of the variance $E[e_a^2(i)]$ as a function of time. By the polarization property, we have

$$E[e_a^2(i)] = E[(\mathbf{u}_i \tilde{\boldsymbol{w}}_i)^2] \;=\; E[\|\tilde{\boldsymbol{w}}_i\|_R^2]$$

which suggests that the learning curve can be evaluated by computing $E[\|\tilde{\boldsymbol{w}}_i\|_R^2]$ for each i. This task can be accomplished recursively from relation (21) by iterating it and setting $\boldsymbol{\sigma} = \mathrm{vec}(\boldsymbol{R})$ (which we shall denote by $\boldsymbol{r}$). This yields

$$\begin{aligned} E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_r^2\right] &= E\left[\|\tilde{\boldsymbol{w}}_0\|_{F^{i+1}r}^2\right] + \\ &\qquad \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|_{(I+F+\cdots+F^i)r}^2\right] \end{aligned}$$

That is,

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_r^2\right] = E\left[\|\tilde{\boldsymbol{w}}_0\|_{a_i}^2\right] + \mu^2\sigma_v^2 b_i \qquad (23)$$

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

$$\boldsymbol{a}_i = \boldsymbol{F}\boldsymbol{a}_{i-1}, \quad \boldsymbol{a}_{-1} = \boldsymbol{r} \qquad (24)$$

$$b_i = b_{i-1} + E\left[\|\overline{\mathbf{u}}_i\|_{a_{i-1}}^2\right], \quad b_{-1} = 0 \qquad (25)$$

We thus have an expression for $\|\tilde{\boldsymbol{w}}_i\|_r^2$ in terms of the initial weight-error vector $\tilde{\boldsymbol{w}}_0$. Using the superposition property of weighted norms, and the above definitions for $\{\boldsymbol{a}_i, b_i\}$, it is easy to show that

$$\begin{aligned} E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|_r^2\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|_r^2\right] + E\left[\|\tilde{\boldsymbol{w}}_0\|_{F^i(F-I)r}^2\right] \\ +\mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|_{F^i r}^2\right] \end{aligned}$$

In other words,

$$\boxed{\begin{aligned} E\left[e_a^2(i+1)\right] = E\left[e_a^2(i)\right] + E\left[\|\tilde{\boldsymbol{w}}_0\|_{F^i(F-I)r}^2\right] \\ +\mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|_{F^i r}^2\right] \end{aligned}} \qquad (26)$$

5. STABILITY AND STEADY-STATE ANALYSIS

Starting from (21) or (22), it is easy to characterize stability and steady-state behavior.

5.1. Stability

By inspecting (21) or (22), it becomes clear that the recursion is stable if, and only if, the matrix $\boldsymbol{F}$ is stable. [The matrix $\boldsymbol{F}$ is actually positive-definite because it is (the expectation of) a Kronecker product of a matrix with itself.] Thus, for stability, the step size μ should be chosen such that $\lambda(\boldsymbol{F}) < 1$. Equivalently, by writing $\boldsymbol{F}$ as $\boldsymbol{F} = \boldsymbol{I} - \mu\boldsymbol{A} + \mu^2\boldsymbol{B}$ with

$$\boldsymbol{A} = \boldsymbol{I}\otimes E\left[\mathbf{u}_i\overline{\mathbf{u}}_i\right] + E\left[\mathbf{u}_i\overline{\mathbf{u}}_i\right]\otimes\boldsymbol{I} \qquad (27)$$

$$\boldsymbol{B} = E\left[\mathbf{u}_i\overline{\mathbf{u}}_i\otimes\mathbf{u}_i\overline{\mathbf{u}}_i\right] \qquad (28)$$

we can show that $\boldsymbol{F}$ (and hence the filter) is stable if and only if

$$\boxed{0 < \mu < \frac{1}{\lambda_{\max}\left(\boldsymbol{A}^{-1}\boldsymbol{B}\right)}} \qquad (29)$$

5.2. Steady-State Behavior

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

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

Thus, in the limit, (21) takes the form

$$\lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|_\sigma^2\right] - \lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|_{F\sigma}^2\right] = \lim_{i\to\infty} \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|_\sigma^2\right]$$

By incorporating the superposition property, and the fact that the input is stationary, we can alternatively write,

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

or, upon replacing $\boldsymbol{\sigma}$ by $(\boldsymbol{I}-\boldsymbol{F})^{-1}\boldsymbol{\sigma}$,

$$\boxed{\lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|_\sigma^2\right] = \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|_{(I-F)^{-1}\sigma}^2\right]} \qquad (30)$$

This gives an expression for the steady-state error energy $\lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|_{\sigma}^2\right]$ for any weight vector $\boldsymbol{\sigma}$. In particular, to evaluate the mean-square deviation (MSD),

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

we set $\boldsymbol{\sigma} = \mathrm{vec}(\boldsymbol{I})$ in (30) to get

$$\boxed{\mathsf{MSD} = \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|^2_{(I-F)^{-1}\mathrm{vec}(I)}\right]} \qquad (31)$$

Similarly, to calculate the mean-square error (MSE)

$$\mathsf{MSE} \triangleq \lim_{i\to\infty} E\left[e_a^2(i)\right] = \lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|^2_R\right]$$

we replace $\boldsymbol{\sigma}$ in (30) by $\mathrm{vec}(\boldsymbol{R})$ which yields

$$\boxed{\mathsf{MSE} = \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|^2_{(I-F)^{-1}\mathrm{vec}(R)}\right]} \qquad (32)$$

6. CONCLUSION

This paper is the first part of a unified study on mean-square analysis of adaptive filtering algorithms. The focus here was on the class of adaptive algorithms employing a general data nonlinearity. Our approach, which is based on energy conservation arguments, does not impose restrictions on the color or statistics of the input sequence. Expressions for the steady-state mean-square performance, and necessary and sufficient conditions for mean-square stability, were derived. In the companion article [10], we extend the discussion to the more demanding case of adaptive filters with error nonlinearities.

7. REFERENCES

[1] 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.

[2] 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.

[3] N. J. Bershad. Analysis of the normalized LMS algorithm with Gaussian inputs. *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, pp. 793–806, 1986.

[4] A. Feuer and E. Weinstein. Convergence analysis of LMS filters with uncorrelated Gaussian data, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 222–229, 1985.

[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] M. Tarrab and A. Feuer. Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data, *IEEE Trans. on Information Theory*, vol. 34, no. 4, pp. 680-690, Jul. 1988.

[7] E. Eweda. Analysis and design of signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data, *IEEE Transactions on Circuits and Systems*, vol. 37, no. 11, Nov. 1990.

[8] M. Rupp. The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes, *IEEE Transactions on Signal Processing*, vol. 41, no. 3, pp. 1149–1160, Mar. 1993.

[9] D. T. M. Slock. On the convergence behavior of the LMS and the normalized LMS algorithms. *IEEE Trans. on Signal Processing*, vol. 41, no. 9, pp. 2811–2825, Sep. 1993.

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

[11] 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.

[12] A. H. Sayed and M. Rupp. Robustness issues in adaptive filtering, in *DSP Handbook*, Chapter 20, CRC Press, 1998.

[13] 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, January 2000.

[14] 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.