MEAN-SQUARE ANALYSIS OF NORMALIZED LEAKY ADAPTIVE FILTERS

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

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

[2]Electrical Engineering Department
Stanford University, CA 94305

ABSTRACT

In this paper, we study leaky adaptive algorithms that employ a general scalar or matrix data nonlinearity. We perform mean-square analysis of this class of algorithms without imposing restrictions on the distribution of the input signal. In particular, we derive conditions on the step-size for stability, and provide closed form expressions for the steady-state performance.

1. ADAPTIVE FILTERING MODEL

In this paper, we consider the following class of leaky adaptive filters:

$$\begin{aligned} \boldsymbol{w}_{i+1} &= (1-\alpha\mu)\boldsymbol{w}_i + \mu \mathbf{H}(\mathbf{u_i})\mathbf{u}_i^T e(i) \qquad (1) \\ e(i) &= d(i) - \mathbf{u}_i \boldsymbol{w}_i \qquad (2) \\ d(i) &= \mathbf{u}_i \boldsymbol{w}^o + v(i) \qquad (3) \end{aligned}$$

where $\boldsymbol{w}_i$ is an estimate for $\boldsymbol{w}^o$ at iteration i, μ is the step-size, $\alpha \geq 0$ is the leakage parameter, $\mathbf{u}_i$ is a row regression vector, $v(i)$ is measurement noise, and $\mathbf{H}(\mathbf{u}_i)$ is a matrix data nonlinearity with nonnegative diagonal entries. Usually, $\mathbf{H}(\mathbf{u}_i)$ is a multiple of the identity, say $\mathbf{H}(\mathbf{u}_i) = \frac{1}{g(\mathbf{u}_i)}\boldsymbol{I}$ for some function $g(\cdot)$. Table 1 lists some common examples of data nonlinearities. There are several reasons for incorporating leakage into an adaptive filter update and special cases of (1)–(2) have been studied before in the literature (see, e.g., [1] and [2] and the references therein for motivation and related discussions).

The purpose of this article is to provide a framework for performing mean-square analysis of the general class of leaky algorithms (1)–(2). This is achieved by relying on the energy-conservation approach developed in [3]–[5]. Among other results, the approach avoids imposing conditions on the statistical distribution of the input sequence (see, e.g., [7, 8]). In addition, the approach enables us to perform both mean-square analysis and transient analysis.

This work was partially supported by the National Science Foundation under awards ECS-9820765 and CCR-9732376. The work of T. Y. Al-Naffouri was also partially supported by a fellowship from King Fahd University of Petroleum and Minerals, Saudi Arabia.

Table 1: Examples of data nonlinearities.

ALGORITHM	$\boldsymbol{H}(\mathbf{u}_i)$
NLMS	$\frac{1}{\|\mathbf{u}_i\|^2}\boldsymbol{I}$
ϵ-NLMS	$\frac{1}{\epsilon+\|\mathbf{u}_i\|^2}\boldsymbol{I}$
sign regressor	$\text{diag}\left(\frac{\text{sign}(u_{i_1})}{u_{i_1}}, \ldots, \frac{\text{sign}(u_{i_M})}{u_{i_M}}\right)$
variable steps	$\text{diag}(\mu_1, \mu_2, \ldots, \mu_M)$

2. DEFINITIONS AND NOTATION

Mean-square analysis of (1)–(2) is carried out in terms of the error quantities:

$$\tilde{\boldsymbol{w}}_i \triangleq \boldsymbol{w}^o - \boldsymbol{w}_i \quad \text{and} \quad e_a(i) \triangleq \mathbf{u}_i \tilde{\boldsymbol{w}}_i \qquad (4)$$

and the normalized regressor $\overline{\mathbf{u}}_i = \mathbf{u}_i \mathbf{H}(\mathbf{u}_i)$. These quantities can be used to rewrite the filter relations (1)–(2) as:

$$\begin{aligned} \tilde{\boldsymbol{w}}_{i+1} &= (1-\alpha\mu)\tilde{\boldsymbol{w}}_i - \mu \overline{\mathbf{u}}_i^T e(i) + \alpha\mu \boldsymbol{w}^o \qquad (5) \\ e(i) &= e_a(i) + v(i) \qquad (6) \end{aligned}$$

We shall replace (5) with the more general adaptation

$$\boxed{\tilde{\boldsymbol{w}}_{i+1} = (1-\alpha\mu)\tilde{\boldsymbol{w}}_i - \mu \overline{\mathbf{u}}_i^T e(i) + \beta\mu \boldsymbol{w}^o} \qquad (7)$$

with separate parameters $\{\alpha, \beta\}$.

We also find it useful to use the compact notation $\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2 = \tilde{\boldsymbol{w}}_i^T \boldsymbol{\Sigma} \tilde{\boldsymbol{w}}_i$. This notation is convenient because it enables us to transform operations on $\tilde{\boldsymbol{w}}_i$ into operations on the norm subscript, as demonstrated by the following properties. Let a_1 and a_2 be scalars and $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ be symmetric matrices of size M. Then

1) Superposition.

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

2) Polarization.

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

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

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

4) Linear transformation. For any $N\times M$ matrix $\mathbf{A}$, $\|\mathbf{A}\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} = \|\tilde{\boldsymbol{w}}_i\|^2_{A^T\Sigma A}$

5) Blindness to asymmetry. For any square matrix $\mathbf{A}$,

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

6) Notational convention. Using the vector notation, we shall write $\|\tilde{\boldsymbol{w}}_i\|^2_{\text{vec}(\Sigma_1)} \triangleq \|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma_1}$

The analysis in the sequel relies on the following two assumptions:

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

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

Observe that we are not requiring the input to be Gaussian.

3. MEAN-SQUARE PERFORMANCE

To study the mean-square performance of the leaky adaptive filters, we need to develop a recursion for the weight-error energy. We therefore start with recursion (7) and compute the energies of both sides to arrive at, after taking expectations,

$$\begin{aligned}
&E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_\Sigma\right] = (1-\alpha\mu)^2 E\left[\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma\right]\\
&-2\mu(1-\alpha\mu)E\left[\tilde{\boldsymbol{w}}_i^T\mathbf{u}_i^T\overline{\mathbf{u}}_i\mathbf{\Sigma}\tilde{\boldsymbol{w}}_i\right] + \mu^2 E\left[e_a^2(i)\|\overline{\mathbf{u}}_i\|^2_\Sigma\right]\\
&+2\mu\beta(1-\alpha\mu)E\left[\boldsymbol{w}^{o^T}\mathbf{\Sigma}\tilde{\boldsymbol{w}}_i\right] - 2\mu^2\beta E\left[\boldsymbol{w}^{o^T}\mathbf{u}_i^T\overline{\mathbf{u}}_i\mathbf{\Sigma}\tilde{\boldsymbol{w}}_i\right]\\
&+\mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|^2_\Sigma\right] + \mu^2\beta^2\|\boldsymbol{w}^o\|^2_\Sigma
\end{aligned} \tag{8}$$

In the above calculation, we used assumption AN to eliminate three noise cross-terms. The above recursion can be expressed more compactly by using the polarization and asymmetry properties, in addition to the independence assumption, to write

$$E\left[\tilde{\boldsymbol{w}}_i^T\mathbf{u}_i^T\overline{\mathbf{u}}_i\mathbf{\Sigma}\tilde{\boldsymbol{w}}_i\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\frac{1}{2}E[u_i^T\overline{u}_i]\Sigma+\frac{1}{2}\Sigma E[\overline{u}_i^T u_i]}\right] \tag{9}$$

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

and

$$(1-\alpha\mu)E\left[\boldsymbol{w}^{o^T}\mathbf{\Sigma}\tilde{\boldsymbol{w}}_i\right] - \mu E\left[\boldsymbol{w}^{o^T}\mathbf{u}_i^T\overline{\mathbf{u}}_i\mathbf{\Sigma}\boldsymbol{w}_i\right] = \boldsymbol{w}^{o^T}\mathbf{\Sigma}\boldsymbol{J}E\left[\tilde{\boldsymbol{w}}_i\right] \tag{11}$$

where we defined

$$\boxed{\mathbf{J} \triangleq E\left[\boldsymbol{I}-\mu\mathcal{U}_i^T\right], \qquad \mathcal{U}_i \triangleq \alpha\boldsymbol{I} + \mathbf{u}_i^T\overline{\mathbf{u}}_i} \tag{12}$$

Substituting (9)–(11) into (8), yields

$$\begin{aligned}
E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_\Sigma\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma'}\right] + \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|^2_\Sigma\right]\\
&+\mu^2\beta^2\|\boldsymbol{w}^o\|^2_\Sigma + 2\mu\beta\boldsymbol{w}^{o^T}\mathbf{\Sigma}\boldsymbol{J}E\left[\tilde{\boldsymbol{w}}_i\right]
\end{aligned} \tag{13}$$

where $\mathbf{\Sigma}'$ is related to $\mathbf{\Sigma}$ via

$$\begin{aligned}
\mathbf{\Sigma}' &= (1-\alpha\mu)^2\mathbf{\Sigma} - \mu(1-\alpha\mu)\mathbf{\Sigma}E\left[\overline{\mathbf{u}}_i^T\mathbf{u}_i\right]\\
&-\mu(1-\alpha\mu)E\left[\mathbf{u}_i^T\overline{\mathbf{u}}_i\right]\mathbf{\Sigma} + \mu^2 E[\|\overline{\mathbf{u}}_i\|^2_\Sigma\mathbf{u}_i^T\mathbf{u}_i]
\end{aligned} \tag{14}$$

Relations (13)–(14) (or, equivalently, relations (16)–(17) below and ultimately (19)) can be used to characterize the mean-square performance of the adaptive filter. In particular, they can be used to derive conditions for mean-square stability, as well as expressions for the steady-state mean-square error and mean-square deviation of an adaptive filter. To this end, note that the above recursion for $\mathbf{\Sigma}$ can be rewritten more compactly, using the vec operation and the Kronecker product notation, as

$$\boxed{\boldsymbol{\sigma}' = \boldsymbol{F}\boldsymbol{\sigma}} \tag{15}$$

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

$$\boxed{\boldsymbol{F} = E\left[(\boldsymbol{I}-\mu\mathcal{U}_i)\otimes(\boldsymbol{I}-\mu\mathcal{U}_i)\right]} \tag{16}$$

In light of (15), recursion (13) becomes

$$\boxed{\begin{aligned}E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_\sigma\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F\sigma}\right] + \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|^2_\sigma\right] +\\ \mu^2\beta^2\|\boldsymbol{w}^o\|^2_\sigma + 2\mu\beta\boldsymbol{w}^{o^T}\mathbf{\Sigma}\boldsymbol{J}E\left[\tilde{\boldsymbol{w}}_i\right]\end{aligned}} \tag{17}$$

To make this recursion self-contained, we need a recursion for $E[\tilde{\boldsymbol{w}}_i]$, which can be obtained by evaluating the expected value of both sides of (7):

$$\boxed{E[\tilde{\boldsymbol{w}}_i] = \boldsymbol{J}E[\tilde{\boldsymbol{w}}_{i-1}] + \mu\beta\boldsymbol{w}^o} \tag{18}$$

Recursion (18) is what we need to supplement (17) and produce the desired self-contained relation. To this end, let us write (17) explicitly for $\{\boldsymbol{\sigma}, \boldsymbol{F}\boldsymbol{\sigma}, \cdots, \boldsymbol{F}^{M^2-1}\boldsymbol{\sigma}\}$:

$$\begin{cases} E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\sigma}\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F\sigma}\right] + \mu^2\sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{\sigma}\right] \\ & \quad +\mu^2\beta^2\|\boldsymbol{w}^o\|^2_{\sigma} + 2\mu\beta\boldsymbol{f}_0^T E[\tilde{\boldsymbol{w}}_i] \\ E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{F\sigma}\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F^2\sigma}\right] + \mu^2\sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{F\sigma}\right] \\ & \quad +\mu^2\beta^2\|\boldsymbol{w}^o\|^2_{F\sigma} + 2\mu\beta\boldsymbol{f}_1^T E[\tilde{\boldsymbol{w}}_i] \\ & \vdots \\ E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{F^{M^2-1}\sigma}\right] &= E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F^{M^2}\sigma}\right] \\ & \quad +\mu^2\sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{F^{M^2-1}\sigma}\right] \\ & \quad +\mu^2\beta^2\|\boldsymbol{w}^o\|^2_{F^{M^2-1}\sigma} \\ & \quad +2\mu\beta\boldsymbol{f}^T_{M^2-1}E[\tilde{\boldsymbol{w}}_i] \\ &= -p_0 E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\sigma}\right] - \cdots \\ & \quad -p_{M^2-1}E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F^{M^2-1}\sigma}\right] \\ & \quad +\mu^2\sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{F^{M^2-1}\sigma}\right] \\ & \quad +\mu^2\beta^2\|\boldsymbol{w}^o\|^2_{F^{M^2-1}\sigma} \\ & \quad +2\mu\beta\boldsymbol{f}^T_{M^2-1}E[\tilde{\boldsymbol{w}}_i] \end{cases}$$

In the above system of equations, $\boldsymbol{f}_k$ is a vector defined by $\boldsymbol{f}_k = \boldsymbol{J}\mathbf{L}_k\boldsymbol{w}^o$, where $\mathbf{L}_k$ is matrix of size M such that $\text{vec}(\mathbf{L}_k) = \boldsymbol{F}^k\boldsymbol{\sigma}$. The last expresion in the above system is obtained from the previous one by means of the Cayley-Hamilton theorem, which enables us to express $\boldsymbol{F}^{M^2}$ as a linear combination of lower powers:

$$\boldsymbol{F}^{M^2} = -p_0\boldsymbol{I} - p_1\boldsymbol{F} - \cdots - p_{M^2-1}\boldsymbol{F}^{M^2-1}$$

where the p_i's are the coefficients of the characteristic polynomial of $\boldsymbol{F}$, viz., $p(x) = \det(x\boldsymbol{I} - \boldsymbol{F})$.

In summary, recursions (15)–(18) can be combined together into a single matrix recursion in state-space form:

$$\boxed{\begin{bmatrix} \mathcal{W}_{i+1} \\ E[\tilde{\boldsymbol{w}}_{i+1}] \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ O & \boldsymbol{J} \end{bmatrix} \begin{bmatrix} \mathcal{W}_i \\ E[\tilde{\boldsymbol{w}}_i] \end{bmatrix} + \begin{bmatrix} \mathcal{Y} \\ \mu\beta\boldsymbol{w}^o \end{bmatrix}} \tag{19}$$

where the matrices $\{\mathbf{G}_1, \mathbf{G}_2\}$ are defined by

$$\mathbf{G}_1 \triangleq \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ -p_0 & -p_1 & \cdots & -p_{M^2-1} \end{bmatrix} \tag{20}$$

$$\mathbf{G}_2 \triangleq 2\mu\beta \begin{bmatrix} \boldsymbol{f}_0^T \\ \boldsymbol{f}_1^T \\ \vdots \\ \boldsymbol{f}^T_{M^2-1} \end{bmatrix}, \quad \mathcal{W}_i \triangleq \begin{bmatrix} E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\sigma}\right] \\ E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F\sigma}\right] \\ \vdots \\ E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F^{M^2-1}\sigma}\right] \end{bmatrix}$$

and

$$\mathcal{Y} = \mu^2 \begin{bmatrix} \sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{\sigma}\right] + \beta^2\|\boldsymbol{w}^o\|^2_{\sigma} \\ \sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{F\sigma}\right] + \beta^2\|\boldsymbol{w}^o\|^2_{F\sigma} \\ \vdots \\ \sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{F^{M^2-1}\sigma}\right] + \beta^2\|\boldsymbol{w}^o\|^2_{F^{M^2-1}\sigma} \end{bmatrix}$$

The state recursion (19) characterizes the transient behavior of the leaky adaptive filters (1)–(2). It can now be used to study mean-square stability and mean-square error performance.

3.1. Stability

From (19), we see that stability is achieved if, and only if, both $\mathbf{G}_1$ and $\boldsymbol{J}$ are stable matrices. However, since $\mathbf{G}_1$ and $\boldsymbol{F}$ have the same eigenvalues, this condition corresponds to requiring that $\boldsymbol{F}$ and $\boldsymbol{J}$ be stable matrices. By inspecting the defining expressions (12) for $\boldsymbol{J}$ and (16) for $\boldsymbol{F}$, we can show that

$$\boldsymbol{J} \text{ is stable} \quad \Leftrightarrow \quad \mu < \frac{2}{\lambda_{\max}(E[\mathcal{U}_i])} \tag{21}$$

$$\boldsymbol{F} \text{ is stable} \quad \Leftrightarrow \quad \mu < \frac{1}{\lambda_{\max}(\mathbf{A}^{-1}\mathbf{B})} \tag{22}$$

where $\mathbf{A} = E[\mathcal{U}_i]\otimes\boldsymbol{I} + \boldsymbol{I}\otimes E[\mathcal{U}_i]$ and $\mathbf{B} = E[\mathcal{U}_i\otimes\mathcal{U}_i]$.

3.2. Steady-State Error

Steady-state performance can be obtained directly from recursion (17). So, assuming the filter is stable, we get $E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_{\sigma}\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\sigma}\right]$ as $i \to \infty$. Therefore, in the limit, relations (17) and (18) lead to

$$\lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{\sigma}\right] = \lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F\sigma}\right] + \mu^2\sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{\sigma}\right] + \mu^2\beta^2\|\boldsymbol{w}^o\|^2_{\sigma} + 2\mu\beta\boldsymbol{w}^{o^T}\boldsymbol{\Sigma}\boldsymbol{J}\lim_{i\to\infty} E[\tilde{\boldsymbol{w}}_i]$$

and

$$\lim_{i\to\infty} E[\tilde{\boldsymbol{w}}_i] = \mu\beta(\boldsymbol{I} - \boldsymbol{J})^{-1}\boldsymbol{w}^o$$

or, equivalently,

$$\boxed{\begin{gathered}\lim_{i\to\infty} E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{(I-F)\sigma}\right] = \\ \mu^2\sigma_v^2 E\left[\|\bar{\mathbf{u}}_i\|^2_{\sigma}\right] + \mu^2\beta^2\|\boldsymbol{w}^o\|^2_{\Sigma\left(I+2J(I-J)^{-1}\right)}\end{gathered}} \tag{23}$$

This expression allows us to evaluate the steady-state weight-error energy for any choice of a symmetric weight $\mathbf{\Sigma}$. In particular, we can get the mean-square error by choosing $\mathbf{\Sigma} = \mathbf{R}$, i.e., by choosing $\boldsymbol{\sigma}$ such that $(\boldsymbol{I} - \boldsymbol{F})\boldsymbol{\sigma} = \text{vec}(\mathbf{R})$. This leads to the expression

$$\begin{aligned}\lim_{i\to\infty} E\left[e_a^2(i)\right] &= \mu^2\sigma_v^2 E\left[\|\overline{\mathbf{u}}_i\|^2_{(I-F)^{-1}\text{vec}(R)}\right] \\ &\quad +\mu^2\beta^2\|\boldsymbol{w}^o\|^2_{\Sigma\left(I+2J(I-J)^{-1}\right)}\end{aligned}$$

where $\text{vec}(\mathbf{\Sigma}) = (\boldsymbol{I} - \boldsymbol{F})^{-1}\text{vec}(\boldsymbol{R})$. Similarly, the mean-square deviation is obtained by choosing $\boldsymbol{\sigma}$ (and hence $\mathbf{\Sigma}$) such that $(\boldsymbol{I} - \boldsymbol{F})\boldsymbol{\sigma} = \text{vec}(\mathbf{I})$.

4. TRACKING ANALYSIS

The results of the previous section can be specialized for non-leaky normalized filters by setting $\alpha = \beta = 0$. More importantly, the analysis can be used to infer (almost immediately) the tracking performance of normalized adaptive filters. In the tracking case, $\boldsymbol{w}^o$ is no more constant but undergoes random perturbations, say

$$\boldsymbol{w}^o_{i+1} = \boldsymbol{w}^o_i + \boldsymbol{q}_i$$

As in the leaky case, we still carry out the derivation in terms of the error quantities $e_a(i)$ and $\tilde{\boldsymbol{w}}_i$ as defined in (4), with $\boldsymbol{w}^o$ replaced by the now time varying $\boldsymbol{w}^o_i$. To perform mean-square analysis in the tracking case, we rely on assumptions AN and AI, in addition to the following assumption:

AT The sequence of tracking errors $\{\boldsymbol{q}_i\}$ is zero-mean and stationary, and is independent of the input $\mathbf{u}_i$ and the additive noise $v(i)$.

Now consider the adaptation equation (1) for $\alpha = 0$, rewritten in terms of $\overline{\mathbf{u}}_i$, $e_a(i)$, and $\tilde{\boldsymbol{w}}_i$:

$$\tilde{\boldsymbol{w}}_{i+1} = \tilde{\boldsymbol{w}}_i - \mu e(i)\overline{\mathbf{u}}_i^T + \boldsymbol{q}_i \qquad (24)$$

Notice that this is the same as (5) for $\alpha = 0$, $\beta = 1/\mu$, and $\boldsymbol{w}^o = \boldsymbol{q}_i$. We can similarly argue that the mean-square behavior is also described by (17) for the same values of α and β and for[1]

$$\|\boldsymbol{w}^o\|^2_\sigma = E\left[\|\boldsymbol{q}\|^2_\sigma\right], \text{ and } \boldsymbol{w}^o = E\left[\boldsymbol{q}_i\right] = 0$$

That is, we now have

$$E\left[\|\tilde{\boldsymbol{w}}_{i+1}\|^2_\sigma\right] = E\left[\|\tilde{\boldsymbol{w}}_i\|^2_{F\sigma}\right] + \mu^2\sigma_v^2 E\left[\|\mathbf{u}_i\|^2_\sigma\right] + E\left[\|\boldsymbol{q}_i\|^2_\sigma\right] \quad (25)$$

Stability and steady-state behavior can now be deduced from (25). In particular, (mean-square) stability is guaranteed if, and only if, $\boldsymbol{F}$ is a stable matrix (see (22)) where now

$$\mathcal{U}_i = \alpha\boldsymbol{I} + \mathbf{u}_i^T\overline{\mathbf{u}}_i\big|_{\alpha=0} = \mathbf{u}_i^T\overline{\mathbf{u}}_i$$

Moreover, by an approach similar to that of the previous section, we can derive the following expression for the steady-state error

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

[1]For completeness, we point out that the mean weight-error behavior can similarly be obtained from (18) with $\alpha = 0, \beta = 1/\mu$, and $\boldsymbol{w}_o = E[\boldsymbol{q}_i] = 0$.

5. CONCLUSION

In this paper, we performed mean-square analysis of leaky normalized adaptive filters. We showed how the analysis can be further used to infer the tracking performance of normalized adaptive filters. Our study applies to a large class of data nonlinearities and does not impose Gaussian assumptions on the data.

6. REFERENCES

[1] K. Mayyas and T. Aboulnasr, "Leaky LMS algorithm: MSE analysis for Gaussian data," *IEEE Trans. on Signal Processing*, vol. 45, no. 4, pp. 927–934, Apr. 1997.

[2] V. H. Nascimento and A. H. Sayed, "Unbiased and stable leakage-based adaptive filters," *IEEE Transactions on Signal Processing*, vol. 47, no. 12, pp. 3261–3276, Dec. 1999.

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

[4] 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.

[5] 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, February 2001. [See also *Proc. 4th IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, vol. 2, pp. 699–703, Antalya, Turkey, June 1999.]

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

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

[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, March 1993.