

Energy Conservation and State-Space Characterization of Adaptive Filter Performance

Ali H. Sayed

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

Abstract—This article provides an overview of an energy-based approach to the study of the steady-state and transient performances of adaptive filters. The analysis employs energy-conservation arguments and is based on studying the energy flow through each iteration of an adaptive filter. Among other results, the approach characterizes the transient behavior of adaptive filters in terms of a linear time-invariant state-space model. The stability and steady-state behavior of the model then translate into stability and mean-square performance results for the adaptive filter. In addition to deriving earlier results in a unified manner, the approach does not restrict the regression data to being Gaussian or white.

I. Introduction

The study of the steady-state and transient performances of adaptive filters can be facilitated by resorting to energy-conservation arguments. These arguments are based on studying the energy flow through each iteration of an adaptive filter, and on showing that the energies of certain a-priori and a-posteriori errors maintain a balance for all time instants [1]–[3]. When examined under expectation, the energy balance leads to a variance relation that characterizes the dynamics of the filter. An advantage of the energy framework is that it does not necessitate an explicit recursion for the covariance matrix of the weight-error vector [4]–[8]; a step that can be cumbersome for adaptive filters with error and data nonlinearities. The energy approach also does not restrict the input data to being Gaussian or white. In this article, we review some of the basic ideas pertaining to this method of analysis and study adaptive filters that employ data nonlinearities in their update equations. A more detailed overview can be found in [9], which also considers adaptive filters with error nonlinearities.

As with most adaptive filter analyses, progress is usually difficult without relying on simplifying assumptions. This is because adaptive filters are, by their very nature, nonlinear and stochastic systems and, therefore, studying their performance can be a formidable task [10]–[12]. The assumptions that we employ tend to be reasonable for sufficiently small step-sizes and for longer filters; both of these conditions are common in applications.

Proc. 34th ISICIE International Symposium on Stochastic Systems Theory and Its Applications, pp. 1–10, Fukuoka, Japan, October 2002. This work was partially supported by NSF under grants ECS-9820765 and CCR-0208573.

II. The Data Model

Consider reference data $\{\mathbf{d}(i)\}$ and regression data $\{\mathbf{u}_i\}$ satisfying a linear regression model of the form

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

for some $M \times 1$ unknown column vector w^o that we wish to estimate. While all vectors in our notation are column vectors, the regressors \mathbf{u}_i in (1) are taken as row vectors so that the inner product between \mathbf{u}_i and w^o is simply written as $\mathbf{u}_i w^o$, with no transposition signs.

In (1), $\mathbf{v}(i)$ denotes measurement noise. Observe that we are using boldface letters to denote random quantities, while the deterministic quantity w^o is written in normal font. In the model (1), the variables $\{\mathbf{d}(i), \mathbf{u}_i, \mathbf{v}(i)\}$ are random and they are assumed to satisfy the following conditions:

- (a) $\{\mathbf{v}(i)\}$ is a zero-mean, independent and identically distributed sequence with variance $E \mathbf{v}^2(i) = \sigma_v^2$.
- (b) $\mathbf{v}(i)$ is independent of \mathbf{u}_j for all i, j .
- (c) The regressor \mathbf{u}_i is zero-mean and has covariance matrix $E \mathbf{u}_i^T \mathbf{u}_i = R_u > 0$.

We focus in this paper on data-normalized adaptive filters for estimating w^o , viz., filters with updates of the form

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]} \mathbf{e}(i), \quad i \geq 0 \quad (3)$$

where

$$\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{u}_i \mathbf{w}_{i-1} \quad (4)$$

is the estimation error at iteration i , and $g[\mathbf{u}_i] > 0$ is some function of \mathbf{u}_i . Typical choices for g are $g[\mathbf{u}] = 1$ (LMS), $g[\mathbf{u}] = \|\mathbf{u}\|^2$ (NLMS), and $g[\mathbf{u}] = \epsilon + \|\mathbf{u}\|^2$ (ϵ -NLMS). The initial condition \mathbf{w}_{-1} of (3) is assumed to be independent of all $\{\mathbf{d}(j), \mathbf{u}_j, \mathbf{v}(j)\}$.

Our first step towards examining the steady-state, convergence and stability behavior of (3) is to establish an energy-conservation relation that holds for a large class of adaptive filters.

III. Energy-Conservation Relation

Let $\tilde{\mathbf{w}}_i = w^o - \mathbf{w}_i$ denote the weight-error vector at iteration i , and define the a-priori and a-posteriori errors:

$$\mathbf{e}_a(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i-1}, \quad \mathbf{e}_p(i) = \mathbf{u}_i \tilde{\mathbf{w}}_i \quad (5)$$

It turns out that the errors $\{\tilde{\mathbf{w}}_i, \tilde{\mathbf{w}}_{i-1}, \mathbf{e}_a(i), \mathbf{e}_p(i)\}$ satisfy a fundamental energy-conservation relation. To see this, we subtract both sides of (3) from w^o to get

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

and then multiply (6) by \mathbf{u}_i from the left to conclude that

$$\mathbf{e}_p(i) = \mathbf{e}_a(i) - \mu \frac{\|\mathbf{u}_i\|^2}{g[\mathbf{u}_i]} \mathbf{e}(i) \quad (7)$$

where the notation $\|\mathbf{u}_i\|^2$ denotes the squared Euclidean norm of \mathbf{u}_i , viz., $\|\mathbf{u}_i\|^2 = \mathbf{u}_i \mathbf{u}_i^T$. Relation (7) can be used to eliminate $\mathbf{e}(i)/g[\mathbf{u}_i]$ from (6). Doing so leads to the equality

$$\|\mathbf{u}_i\|^2 \cdot \tilde{\mathbf{w}}_i + \mathbf{u}_i^T \mathbf{e}_a(i) = \|\mathbf{u}_i\|^2 \cdot \tilde{\mathbf{w}}_{i-1} + \mathbf{u}_i^T \mathbf{e}_p(i) \quad (8)$$

By equating the weighted Euclidean norms of both sides of this equation, we arrive, after a straightforward calculation, at the equality:

$$\boxed{\|\tilde{\mathbf{w}}_i\|^2 + (\|\mathbf{u}_i\|^2)^\dagger \cdot (\mathbf{e}_a(i))^2 = \|\tilde{\mathbf{w}}_{i-1}\|^2 + (\|\mathbf{u}_i\|^2)^\dagger \cdot (\mathbf{e}_p(i))^2} \quad (9)$$

in terms of the pseudo-inverse¹ of the scalar quantity $\|\mathbf{u}_i\|^2$. This energy relation is an exact result that shows how the energies of the weight-error vectors at two successive time instants are related to the energies of the a priori and a posteriori estimation errors.

The result (9) was developed in [1] and subsequently used in a series of works to study the robustness of adaptive filters (e.g., [2], [3], [13], [14]). It was also used in [4]–[6] to study the steady-state and tracking performances of adaptive filters, as we proceed to explain.

IV. Mean-Square Error Performance

In performance analysis we are interested in evaluating the steady-state mean-square error of the filter, defined by

$$\text{MSE} = \lim_{i \rightarrow \infty} \text{E} \mathbf{e}^2(i)$$

Now since

$$\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{u}_i \mathbf{w}_{i-1} = \mathbf{e}_a(i) + \mathbf{v}(i)$$

it follows that

$$\text{E} \mathbf{e}^2(i) = \text{E} \mathbf{e}_a^2(i) + \sigma_v^2$$

so that

$$\boxed{\text{MSE} = \text{EMSE} + \sigma_v^2}$$

where EMSE refers to the filter excess-mean-square error, defined as the limiting value of $\text{E} \mathbf{e}_a^2(i)$. In other words, it is enough for performance evaluation to focus on computing the EMSE. The ratio EMSE/σ_v^2 is called the filter misadjustment.

¹For a scalar a , $a^\dagger = 1/a$ when $a \neq 0$ and $a^\dagger = 0$ when $a = 0$.

Assuming filter operation in steady-state, the energy relation (9) can be used to estimate the filter EMSE. Indeed, taking expectations of both sides of (9), we get

$$\text{E} \|\tilde{\mathbf{w}}_i\|^2 + \text{E} [(\|\mathbf{u}_i\|^2)^\dagger \mathbf{e}_a^2(i)] = \text{E} \|\tilde{\mathbf{w}}_{i-1}\|^2 + \text{E} [(\|\mathbf{u}_i\|^2)^\dagger \mathbf{e}_p^2(i)]$$

and since, in steady-state,

$$\text{E} \|\tilde{\mathbf{w}}_i\|^2 = \text{E} \|\tilde{\mathbf{w}}_{i-1}\|^2 \quad \text{as } i \rightarrow \infty$$

we find that the following equality must hold as $i \rightarrow \infty$,

$$\text{E} [(\|\mathbf{u}_i\|^2)^\dagger \mathbf{e}_a^2(i)] = \text{E} [(\|\mathbf{u}_i\|^2)^\dagger \mathbf{e}_p^2(i)]$$

If we replace $\mathbf{e}_p(i)$ by (7) in terms of $\{\mathbf{e}_a(i), \mathbf{v}(i)\}$ and arrange terms, and if we assume that the event $\|\mathbf{u}_i\|^2 = 0$ has zero probability, we arrive at the following steady-state variance identity:

$$\boxed{2\text{E} \left(\frac{\mathbf{e}_a^2(i)}{g[\mathbf{u}_i]} \right) = \mu \text{E} \left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]} \mathbf{e}_a^2(i) \right) + \mu \sigma_v^2 \text{E} \left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]} \right)} \quad (10)$$

This relation is an exact result for filters operating in steady-state. It provides an equation in terms of $\mathbf{e}_a(i)$, which can be used to estimate $\text{E} \mathbf{e}_a^2(i)$. It is at this stage that we need to resort to some simplifying assumptions because (10) is hard to solve analytically even for the simplest of algorithms.

A. Application to LMS

As an illustration, consider the LMS filter for which $g(\mathbf{u}_i) = 1$, i.e.,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i^T \mathbf{e}(i) \quad (11)$$

In this case, relation (10) gives

$$2\zeta = \mu \text{E} (\|\mathbf{u}_i\|^2 \mathbf{e}_a^2(i)) + \mu \sigma_v^2 \text{Tr}(R_u) \quad (12)$$

where we are using the symbol ζ to denote the filter EMSE, $\zeta = \text{E} \mathbf{e}_a^2(\infty)$. Eq. (12) requires that we evaluate the limiting value of $\text{E} \|\mathbf{u}_i\|^2 \mathbf{e}_a^2(i)$. This evaluation is simplified by resorting to some approximations.

For instance, when the step-size μ is sufficiently small, we may ignore the effect of $\text{E} (\|\mathbf{u}_i\|^2 \mathbf{e}_a^2(i))$ to get

$$\boxed{\zeta \approx \frac{\mu \sigma_v^2 \text{Tr}(R_u)}{2}} \quad (13)$$

which is a famous result for LMS. For larger step-sizes, we can resort to the separation assumption:

$$\boxed{\text{At steady-state, } \|\mathbf{u}_i\|^2 \text{ is independent of } \mathbf{e}_a(i)} \quad (14)$$

which allows us to write

$$\text{E} (\|\mathbf{u}_i\|^2 \mathbf{e}_a^2(i)) = (\text{E} \|\mathbf{u}_i\|^2) \cdot (\text{E} \mathbf{e}_a^2(i)) \quad i \rightarrow \infty$$

so that (12) gives

$$\boxed{\zeta \approx \frac{\mu \sigma_v^2 \text{Tr}(R_u)}{2 - \mu \text{Tr}(R_u)}} \quad (15)$$

which is another famous expression for LMS.

The separation assumption (14) is reasonable for longer filters. It can be exact in some scenarios, e.g., when the entries of the regressor belong to a finite alphabet. Assume for instance that \mathbf{u}_i possesses shift-structure, say,

$$\mathbf{u}_i = [\mathbf{u}(i) \quad \mathbf{u}(i-1) \quad \dots \quad \mathbf{u}(i-M+1)]$$

and that each $\mathbf{u}(i)$ is a binary random variable, i.e., it assumes the values ± 1 with probability $1/2$. Then $\|\mathbf{u}_i\|^2 = M$ for all i and is clearly independent of $\mathbf{e}_a(i)$. In this case, the filter EMSE will be exactly equal to $\zeta = \mu\sigma_v^2 M / (2 - \mu M)$.

B. Application to NLMS

Consider now the NLMS filter for which $g(\mathbf{u}_i) = \|\mathbf{u}_i\|^2$,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \frac{\mathbf{u}_i^T}{\|\mathbf{u}_i\|^2} \mathbf{e}(i) \quad (16)$$

In this case, relation (10) gives

$$(2 - \mu) \mathbb{E} \left(\frac{\mathbf{e}_a^2(i)}{\|\mathbf{u}_i\|^2} \right) = \mu\sigma_v^2 \mathbb{E} \left(\frac{1}{\|\mathbf{u}_i\|^2} \right) \quad (17)$$

Using the separation assumption (14) again we have

$$\mathbb{E} \left(\frac{\mathbf{e}_a^2(i)}{\|\mathbf{u}_i\|^2} \right) = \mathbb{E} \mathbf{e}_a^2(i) \cdot \mathbb{E} \left(\frac{1}{\|\mathbf{u}_i\|^2} \right)$$

so that the filter EMSE is given by

$$\zeta \approx \frac{\mu\sigma_v^2}{2 - \mu} \quad (18)$$

Alternatively, if we use the approximation

$$\mathbb{E} \left(\frac{\mathbf{e}_a^2(i)}{\|\mathbf{u}_i\|^2} \right) \approx \frac{\mathbb{E} \mathbf{e}_a^2(i)}{\mathbb{E} \|\mathbf{u}_i\|^2} = \frac{\mathbb{E} \mathbf{e}_a^2(i)}{\text{Tr}(R_u)}$$

then the filter EMSE is approximated by

$$\zeta \approx \frac{\mu\sigma_v^2}{2 - \mu} \text{Tr}(R_u) \mathbb{E} \left(\frac{1}{\|\mathbf{u}_i\|^2} \right) \quad (19)$$

which is a known result for NLMS. Again, when the entries of \mathbf{u}_i arise from a finite alphabet, so that $\|\mathbf{u}_i\|^2$ is equal to a constant, the filter EMSE is exactly given by $\zeta = \mu\sigma_v^2 / (2 - \mu)$.

C. Application to RLS

The energy-conservation approach can be extended to handle more general update recursions. Thus consider now the RLS algorithm,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mathbf{P}_i \mathbf{u}_i^T [\mathbf{d}(i) - \mathbf{u}_i \mathbf{w}_{i-1}], \quad i \geq 0 \quad (20)$$

$$\mathbf{P}_i = \alpha^{-1} \left[\mathbf{P}_{i-1} - \frac{\alpha^{-1} \mathbf{P}_{i-1} \mathbf{u}_i^T \mathbf{u}_i \mathbf{P}_{i-1}}{1 + \alpha^{-1} \mathbf{u}_i \mathbf{P}_{i-1} \mathbf{u}_i^T} \right] \quad (21)$$

with initial condition $\mathbf{P}_{-1} = \epsilon^{-1} I$ and where $0 \ll \alpha \leq 1$ is the forgetting factor. Compared with recursion (3), the

RLS update includes the matrix factor \mathbf{P}_i multiplying \mathbf{u}_i^T from the left. Moreover, \mathbf{P}_i is a function of both current and prior regressors. Still, the energy arguments of Secs. III–IV can be extended to deal with this case. In particular, it is straightforward to verify that (9) is now replaced by

$$\|\tilde{\mathbf{w}}_i\|_{\mathbf{P}_i^{-1}}^2 + \left(\|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \right)^\dagger \mathbf{e}_a^2(i) = \|\tilde{\mathbf{w}}_{i-1}\|_{\mathbf{P}_{i-1}^{-1}}^2 + \left(\|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \right)^\dagger \mathbf{e}_p^2(i) \quad (22)$$

where the notation $\|x\|_{\Sigma}^2$ stands for $\|x\|_{\Sigma}^2 = x^T \Sigma x$.

The presence of the matrices $\{\mathbf{P}_i^{-1}, \mathbf{P}_i\}$ makes the subsequent performance analysis rather challenging; this is because $\{\mathbf{P}_i^{-1}, \mathbf{P}_i\}$ are dependent not only on \mathbf{u}_i but also on prior regressors $\{\mathbf{u}_j, j \leq i\}$. For this reason, we shall approximate and replace the random variables $\{\mathbf{P}_i^{-1}, \mathbf{P}_i\}$ in steady-state by their respective mean values. From the relation

$$\mathbf{P}_i^{-1} = \alpha^{i+1} \epsilon I + \sum_{j=0}^i \alpha^{i-j} \mathbf{u}_j^* \mathbf{u}_j$$

we find that

$$\lim_{i \rightarrow \infty} \mathbb{E} (\mathbf{P}_i^{-1}) = \frac{R_u}{1 - \alpha} \triangleq P^{-1}$$

The evaluation of $\mathbb{E} \mathbf{P}_i$, on the other hand, is harder and we shall resort to the approximation

$$\mathbb{E} \mathbf{P}_i \approx [\mathbb{E} \mathbf{P}_i^{-1}]^{-1} = P, \quad \text{as } i \rightarrow \infty$$

Now setting

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|_{\mathbf{P}_i^{-1}}^2 \approx \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{P^{-1}}^2, \quad \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{\mathbf{P}_{i-1}^{-1}}^2 \approx \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{P^{-1}}^2$$

and using, in steady-state,

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|_{\mathbf{P}_i^{-1}}^2 = \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{\mathbf{P}_{i-1}^{-1}}^2$$

we find that the energy-relation (22) leads to

$$\mathbb{E} \left[\left(\|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \right)^\dagger \mathbf{e}_a^2(i) \right] = \mathbb{E} \left[\left(\|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \right)^\dagger \mathbf{e}_p^2(i) \right] \quad \text{as } i \rightarrow \infty$$

Using $\mathbf{e}_p(i) = \mathbf{e}_a(i) - \|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \mathbf{e}(i)$ and $\mathbf{e}(i) = \mathbf{e}_a(i) + \mathbf{v}(i)$, we then arrive at the steady-state variance relation:

$$2\mathbb{E} \mathbf{e}_a^2(i) = \sigma_v^2 \mathbb{E} \|\mathbf{u}_i\|_{\mathbf{P}_i}^2 + \mathbb{E} \|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \mathbf{e}_a^2(i) \quad (23)$$

If we now invoke the separation assumption

$$\text{At steady-state, } \|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \text{ is independent of } \mathbf{e}_a(i) \quad (24)$$

replace \mathbf{P}_i by its assumed mean value, and use

$$\mathbb{E} \|\mathbf{u}_i\|_{\mathbf{P}_i}^2 \approx \text{Tr}(R_u P) = (1 - \alpha) M$$

we find that

$$\zeta \approx \frac{\sigma_v^2 (1 - \alpha) M}{2 - (1 - \alpha) M}$$

Usually, the value of α is close to one and, hence,

$$\zeta \approx \frac{\sigma_v^2 (1 - \alpha) M}{2} \quad (25)$$

D. Application to APA

Consider now the APA algorithm,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{U}_i^T (\mathbf{U}_i \mathbf{U}_i^*)^{-1} \mathbf{e}_i, \quad i \geq 0 \quad (26)$$

$$\mathbf{e}_i = \mathbf{d}_i - \mathbf{U}_i \mathbf{w}_{i-1} \quad (27)$$

where

$$\mathbf{U}_i = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{u}_{i-1} \\ \vdots \\ \mathbf{u}_{i-1+K} \end{bmatrix}, \quad \mathbf{d}_i = \begin{bmatrix} \mathbf{d}(i) \\ \mathbf{d}(i-1) \\ \vdots \\ \mathbf{d}(i-1+K) \end{bmatrix}$$

The energy-conservation relation (9) can also be extended to this case, which involves a vector-valued error \mathbf{e}_i . In particular, it is straightforward to verify that (9) is now replaced by

$$\|\tilde{\mathbf{w}}_i\|^2 + \mathbf{e}_{a,i}^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{e}_{a,i} = \|\tilde{\mathbf{w}}_{i-1}\|^2 + \mathbf{e}_{p,i}^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{e}_{p,i} \quad (28)$$

where the a-priori and a-posteriori error vectors are defined by

$$\mathbf{e}_{a,i} = \mathbf{U}_i \tilde{\mathbf{w}}_{i-1}, \quad \mathbf{e}_{p,i} = \mathbf{U}_i \tilde{\mathbf{w}}_i$$

Using the steady-state condition

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|^2 = \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|^2 \quad \text{as } i \rightarrow \infty$$

and the relations

$$\mathbf{e}_{p,i} = \mathbf{e}_{a,i} - \mu \mathbf{e}_i, \quad \mathbf{e}_i = \mathbf{e}_{a,i} + \mathbf{v}_i$$

where $\mathbf{v}_i = \text{col}\{\mathbf{v}(i), \mathbf{v}(i-1), \dots, \mathbf{v}(i-K+1)\}$, we find that (28) reduces under expectation to the following steady-state variance relation:

$$\begin{aligned} & 2\mathbb{E} [\mathbf{e}_{a,i}^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{e}_{a,i}] = \\ & \mu \mathbb{E} [\mathbf{e}_{a,i}^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{e}_{a,i}] + \mu \mathbb{E} [\mathbf{v}_i^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{v}_i] \end{aligned} \quad (29)$$

To proceed we resort to the separation assumption,

$$\begin{aligned} & \text{At steady-state, } \mathbf{U}_i \text{ is independent of } \mathbf{e}_{a,i} \\ & \text{and } \mathbb{E} \mathbf{e}_{a,i} \mathbf{e}_{a,i}^T = (\mathbb{E} \mathbf{e}_a^2(i)) \cdot b_0 b_0^T, \text{ where} \\ & b_0 = \text{col}\{1, 0, \dots, 0\} \end{aligned} \quad (30)$$

which is reasonable at low noise levels, so that

$$\mathbb{E} [\mathbf{e}_{a,i}^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{e}_{a,i}] = \alpha \mathbb{E} \mathbf{e}_a^2(i)$$

where α denotes the expected value of the (0,0) entry of $(\mathbf{U}_i \mathbf{U}_i^T)^{-1}$. Likewise,

$$\begin{aligned} \mathbb{E} [\mathbf{v}_i^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{v}_i] &= \sigma_v^2 \mathbb{E} \text{Tr} [(\mathbf{U}_i \mathbf{U}_i^T)^{-1}] \\ \mathbb{E} [\mathbf{e}_{a,i}^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{e}_{a,i}] &= \alpha \mathbb{E} \mathbf{e}_a^2(i) \end{aligned}$$

as $i \rightarrow \infty$. It then follows that

$$\zeta \approx \frac{\mu \sigma_v^2}{(2-\mu)\alpha} \mathbb{E} \text{Tr} [(\mathbf{U}_i \mathbf{U}_i^T)^{-1}] \quad (31)$$

A related expression can be obtained by using the approximation $\alpha \approx 1/\text{Tr}(R_u)$, which leads to

$$\zeta \approx \frac{\mu \sigma_v^2}{(2-\mu)} \text{Tr}(R_u) \mathbb{E} \text{Tr} [(\mathbf{U}_i \mathbf{U}_i^T)^{-1}] \quad (32)$$

If we further use

$$\mathbb{E} \text{Tr} [(\mathbf{U}_i \mathbf{U}_i^T)^{-1}] \approx \mathbb{E} \left(\frac{K}{\|\mathbf{u}_i\|^2} \right)$$

then

$$\zeta \approx \frac{\mu \sigma_v^2 K}{(2-\mu)} \text{Tr}(R_u) \mathbb{E} \left(\frac{1}{\|\mathbf{u}_i\|^2} \right) \quad (33)$$

V. Tracking Performance

Besides enabling us to evaluate the steady-state performance of the adaptive filters (3) in stationary environments, the energy-conservation approach of Secs. III–IV can also be used to characterize the performance of such filters (and of more general filters) under nonstationary conditions.

Thus consider reference data $\{\mathbf{d}(i)\}$ and regression data $\{\mathbf{u}_i\}$ satisfying the linear relation

$$\mathbf{d}(i) = \mathbf{u}_i \mathbf{w}_i^o + \mathbf{v}(i) \quad (34)$$

where the model \mathbf{w}_i^o is assumed to vary according to the rule

$$\begin{aligned} \mathbf{w}_i^o &= w^o + \boldsymbol{\theta}_i \\ \boldsymbol{\theta}_i &= \alpha \boldsymbol{\theta}_{i-1} + \mathbf{q}_i \end{aligned} \quad (35)$$

where $\boldsymbol{\theta}_i$ denotes a first-order auto-regressive disturbance around w^o , and $|\alpha| \leq 1$.² The sequence $\{\mathbf{q}_i\}$ is zero mean, independent of all $\{\mathbf{v}(j), \mathbf{u}_j\}$ and has covariance matrix \mathbf{Q} . Observe that the model (35) can also be rewritten as³

$$(\mathbf{w}_i^o - w^o) = \alpha (\mathbf{w}_{i-1}^o - w^o) + \mathbf{q}_i \quad (36)$$

Now define the error quantities

$$\begin{aligned} \tilde{\mathbf{w}}_i &= \mathbf{w}_i^o - \mathbf{w}_i \\ \mathbf{e}_a(i) &= \mathbf{u}_i \mathbf{w}_i^o - \mathbf{u}_i \mathbf{w}_{i-1} \\ \mathbf{e}_p(i) &= \mathbf{u}_i \mathbf{w}_i^o - \mathbf{u}_i \mathbf{w}_i \end{aligned}$$

It is then easy to verify, by repeating the arguments of Sec. III, that the energy relation (9) is replaced by

$$\|\tilde{\mathbf{w}}_i - \mathbf{c}_i\|^2 + (\|\mathbf{u}_i\|^2)^\dagger \mathbf{e}_a^2(i) = \|\tilde{\mathbf{w}}_{i-1}\|^2 + (\|\mathbf{u}_i\|^2)^\dagger \mathbf{e}_p^2(i) \quad (37)$$

where

$$\mathbf{c}_i = (\alpha - 1) \boldsymbol{\theta}_{i-1} + \mathbf{q}_i$$

²Although unnecessary, we shall assume for simplicity of presentation in this paper that α is real — see [6] for the general case.

³It is often assumed in the literature that $\alpha = 1$; a choice that is not physically meaningful since the covariance matrix of \mathbf{w}_i^o would grow unboundedly.

Taking expectations of both sides of (37), and using

$$\mathbf{e}_p(i) = \mathbf{e}_a(i) - \mu \frac{\|\mathbf{u}_i\|^2}{g[\mathbf{u}_i]} \mathbf{e}(i)$$

we find that (37) leads to the steady-state variance relation

$$\begin{aligned} 2\mathbb{E} \left(\frac{\mathbf{e}_a^2(i)}{g[\mathbf{u}_i]} \right) &= \mu \mathbb{E} \left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]} \mathbf{e}_a^2(i) \right) + \mu \sigma_v^2 \mathbb{E} \left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]} \right) \\ &\quad + \mu^{-1} \text{Tr}(Q) + \mu^{-1} |1 - \alpha|^2 \text{Tr}(\Theta) \\ &\quad - 2\mu^{-1} \left[(1 - \alpha) \mathbb{E} \left(\boldsymbol{\theta}_{i-1}^T \left(\tilde{\mathbf{w}}_{i-1} - \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]} \mathbf{e}(i) \right) \right) \right] \end{aligned}$$

where

$$\Theta \triangleq \lim_{i \rightarrow \infty} \mathbb{E} \boldsymbol{\theta}_i \boldsymbol{\theta}_i^T = \frac{Q}{1 - |\alpha|^2}$$

A. Application to LMS

Consider again the LMS algorithm for which $g(\mathbf{u}_i) = 1$. Then the above variance relation becomes

$$\begin{aligned} 2\mathbb{E} \mathbf{e}_a^2(i) &= \mu \mathbb{E} (\|\mathbf{u}_i\|^2 \mathbf{e}_a^2(i)) + \mu \sigma_v^2 \mathbb{E} (\|\mathbf{u}_i\|^2) \\ &\quad + \mu^{-1} \text{Tr}(Q) + \mu^{-1} |1 - \alpha|^2 \text{Tr}(\Theta) \\ &\quad - 2\mu^{-1} \left[(1 - \alpha) \mathbb{E} \left(\boldsymbol{\theta}_{i-1}^T (\tilde{\mathbf{w}}_{i-1} - \mu \mathbf{u}_i^T \mathbf{e}(i)) \right) \right] \end{aligned} \quad (38)$$

If we now assume that

$$\text{At steady state, } \tilde{\mathbf{w}}_{i-1} \text{ is independent of } \mathbf{u}_i$$

then it can be verified that

$$\mathbb{E} \tilde{\mathbf{w}}_i = 0 \quad \text{and} \quad \mathbb{E} \tilde{\mathbf{w}}_i \boldsymbol{\theta}_i^T = [\alpha I - (I - \mu R_u)^{-1}]^{-1} C$$

where

$$C = \alpha(1 - \alpha)\Theta - Q$$

Substituting into (38), and assuming the step-size is sufficiently small so that the term $\mathbb{E}(\|\mathbf{u}_i\|^2 \mathbf{e}_a^2(i))$ can be neglected, we can solve for $\mathbb{E} \mathbf{e}_a^2(i)$ and find that

$$\zeta = \frac{\mu \sigma_v^2 \text{Tr}(R_u) + \mu^{-1} \beta}{2}$$

where

$$\begin{aligned} \beta &= \text{Tr}[(I + 2(1 - \alpha)X_\alpha)Q] \\ X_\alpha &= (I - \mu R_u) [\alpha I - (I - \mu R_u)^{-1}]^{-1} \end{aligned}$$

For larger values of μ , we can again resort to the separation assumption (14) to get

$$\zeta = \frac{\mu \sigma_v^2 \text{Tr}(R_u) + \mu^{-1} \beta}{2 - \mu \text{Tr}(R_u)}$$

VI. Transient Analysis

The energy arguments can also be used to study the transient behavior of adaptive filters of the form (3)–(4), viz.,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]} \mathbf{e}(i), \quad i \geq 0 \quad (39)$$

where

$$\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{u}_i \mathbf{w}_{i-1} \quad (40)$$

In order to illustrate this fact, we shall focus on stationary environments of the form (1)–(2), viz.,

$$\mathbf{d}(i) = \mathbf{u}_i w^o + \mathbf{v}(i) \quad (41)$$

where $\{\mathbf{d}(i), \mathbf{u}_i, \mathbf{v}(i)\}$ satisfy (2). It turns out that for transient analysis it is useful to rely on a weighted energy relation instead of (9).

VII. Weighted Energy-Conservation Relation

Let $\tilde{\mathbf{w}}_i = w^o - \mathbf{w}_i$ denote the weight-error vector at iteration i , and choose any $M \times M$ symmetric positive-definite matrix Σ . Define further the weighted a-priori and a-posteriori errors:

$$\mathbf{e}_a^\Sigma(i) \triangleq \mathbf{u}_i \Sigma \tilde{\mathbf{w}}_{i-1}, \quad \mathbf{e}_p^\Sigma(i) \triangleq \mathbf{u}_i \Sigma \tilde{\mathbf{w}}_i \quad (42)$$

As we shall see, the freedom in selecting Σ is useful in characterizing the dynamic behavior of the filter, as well as its steady-state performance [7]–[9].

It turns out that the errors $\{\tilde{\mathbf{w}}_i, \tilde{\mathbf{w}}_{i-1}, \mathbf{e}_a^\Sigma(i), \mathbf{e}_p^\Sigma(i)\}$ satisfy an energy-conservation relation similar to (9). By repeating the arguments of Sec. III we can verify that

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]} \mathbf{e}(i) \quad (43)$$

$$\mathbf{e}_p^\Sigma(i) = \mathbf{e}_a^\Sigma(i) - \mu \frac{\|\mathbf{u}_i\|_\Sigma^2}{g[\mathbf{u}_i]} \mathbf{e}(i) \quad (44)$$

and

$$\|\mathbf{u}_i\|_\Sigma^2 \cdot \|\tilde{\mathbf{w}}_i\|_\Sigma^2 + (\mathbf{e}_a^\Sigma(i))^2 = \|\mathbf{u}_i\|_\Sigma^2 \cdot \|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 + (\mathbf{e}_p^\Sigma(i))^2 \quad (45)$$

This energy relation is an exact result; it is simply the weighted version of (9). In addition, it follows from $\mathbf{e}(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i-1} + \mathbf{v}(i)$, and from (43), that

$$\tilde{\mathbf{w}}_i = \left(I - \mu \frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]} \right) \tilde{\mathbf{w}}_{i-1} - \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]} \mathbf{v}(i) \quad (46)$$

VIII. Weighted Variance Relation

In transient analysis we are interested in studying the evolution of $\mathbb{E} \|\tilde{\mathbf{w}}_i\|_\Sigma^2$, for some Σ of interest (usually, $\Sigma = I$ or $\Sigma = R_u$). In particular, the evolution of $\mathbb{E} \|\tilde{\mathbf{w}}_i\|^2$ corresponds to the filter mean-square deviation while the evolution of $\mathbb{E} \|\tilde{\mathbf{w}}_i\|_{R_u}^2$ relates to the filter mean-square error (or learning) curve since

$$\mathbb{E} \mathbf{e}^2(i) = \mathbb{E} \mathbf{e}_a^2(i) + \sigma_v^2 = \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{R_u}^2 + \sigma_v^2$$

In the sequel, we focus on the evolution of $E\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2$ for arbitrary Σ .

To study such evolutions by means of the energy relation (45), we replace $\mathbf{e}_p^{\Sigma}(i)$ by (44) and expand to find

$$\begin{aligned} \|\mathbf{u}_i\|_{\Sigma}^2 \cdot \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 &= \|\mathbf{u}_i\|_{\Sigma}^2 \cdot \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma}^2 + \\ &\quad \frac{\mu^2(\|\mathbf{u}_i\|_{\Sigma}^2)^2}{g^2[\mathbf{u}_i]} \|\tilde{\mathbf{w}}_{i-1}\|_{\mathbf{u}_i^T \mathbf{u}_i}^2 + \\ &\quad \frac{\mu^2(\|\mathbf{u}_i\|_{\Sigma}^2)^2}{g^2[\mathbf{u}_i]} \mathbf{v}^2(i) - \\ &\quad \frac{\mu\|\mathbf{u}_i\|_{\Sigma}^2}{g[\mathbf{u}_i]} \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma \mathbf{u}_i^T \mathbf{u}_i + \mathbf{u}_i^T \mathbf{u}_i \Sigma}^2 + \\ &\quad 2\mu^2 \frac{(\|\mathbf{u}_i\|_{\Sigma}^2)^2}{g^2[\mathbf{u}_i]} \mathbf{v}(i) \mathbf{e}_a(i) - \\ &\quad 2\mu \frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g[\mathbf{u}_i]} \mathbf{v}(i) \mathbf{e}_a^{\Sigma}(i) \end{aligned} \quad (47)$$

Taking expectations of both sides leads to the variance relation

$$E\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 = E\left(\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2\right) + \mu^2 \sigma_v^2 E\left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]}\right) \quad (48)$$

where

$$\Sigma' = \Sigma - \frac{\mu}{g[\mathbf{u}_i]} \Sigma \mathbf{u}_i^T \mathbf{u}_i - \frac{\mu}{g[\mathbf{u}_i]} \mathbf{u}_i^T \mathbf{u}_i \Sigma + \frac{\mu^2 \|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]} \mathbf{u}_i^T \mathbf{u}_i \quad (49)$$

Observe that Σ' is a random matrix due to its dependence on the data. In contrast to the steady-state variance relation (10), the result (48) holds at every iteration i .

A. Independent Regressors

Relations (46), (48), and (49) describe the dynamic behavior of adaptive filters. However, recursion (48) is particularly hard to propagate. This is because it requires the evaluation of

$$E\left(\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2\right) = E\left(\tilde{\mathbf{w}}_{i-1}^T \Sigma' \tilde{\mathbf{w}}_{i-1}\right)$$

and the weighting matrix Σ' is dependent on \mathbf{u}_i . Not only that, but $\tilde{\mathbf{w}}_{i-1}$ is also dependent on prior regressors. For this reason, at this stage, we shall resort to the independence assumption:

$$\boxed{\text{The } \{\mathbf{u}_i\} \text{ are independent and identically distributed}} \quad (50)$$

Under (50), it holds that

$$E\left(\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2\right) = E\left(\|\tilde{\mathbf{w}}_{i-1}\|_{E[\Sigma']}^2\right)$$

with the weighting matrix Σ' now replaced by its mean, which we denote by Σ' . In this way, the variance recursion (48) becomes

$$E\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 = E\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2 + \mu^2 \sigma_v^2 E\left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]}\right) \quad (51)$$

with deterministic weights $\{\Sigma, \Sigma'\}$ and where, from (49),

$$\Sigma' = \Sigma - \mu \Sigma E\left(\frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]}\right) - \mu E\left(\frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]}\right) \Sigma + \mu^2 E\left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]} \mathbf{u}_i^T \mathbf{u}_i\right) \quad (52)$$

Finally, taking expectations of both sides of (46), and using (50), we also find that

$$E\tilde{\mathbf{w}}_i = \left(I - \mu E\left(\frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]}\right)\right) \cdot E\tilde{\mathbf{w}}_{i-1} \quad (53)$$

Expressions (51)–(53) show that studying the transient behavior of an adaptive filter requires evaluating the multivariate moments:

$$E\left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]}\right), \quad E\left(\frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]}\right) \quad \text{and} \quad E\left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]} \mathbf{u}_i^T \mathbf{u}_i\right)$$

which are only functions of \mathbf{u}_i .

IX. Mean-Square Behavior

Let σ denote the $M^2 \times 1$ column vector that is obtained by stacking the columns of Σ on top of each other, written as $\sigma = \text{vec}(\Sigma)$. Likewise, let $\sigma' = \text{vec}(\Sigma')$. Now using the Kronecker product notation, and the following property, for arbitrary matrices $\{X, Y, Z\}$,

$$\text{vec}(XYZ) = (Z^T \otimes X) \text{vec}(Y)$$

we can verify that relation (52) for Σ' transforms into the vector relation

$$\sigma' = F\sigma$$

where F is $M^2 \times M^2$ and given by

$$F = I - \mu A + \mu^2 B \quad (54)$$

in terms of the symmetric matrices:

$$\begin{aligned} A &= (P \otimes I_M) + (I_M \otimes P) > 0 \\ B &= E\left(\frac{\mathbf{u}_i^T \mathbf{u}_i \otimes \mathbf{u}_i^T \mathbf{u}_i}{g^2[\mathbf{u}_i]}\right) \geq 0 \\ P &= E\left(\frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]}\right) > 0 \end{aligned} \quad (55)$$

Using the column notation σ , and the relation $\sigma' = F\sigma$, we shall rewrite (51)–(52) as

$$E\|\tilde{\mathbf{w}}_i\|_{\sigma}^2 = E\|\tilde{\mathbf{w}}_{i-1}\|_{F\sigma}^2 + \mu^2 \sigma_v^2 E\left(\frac{\|\mathbf{u}_i\|_{\sigma}^2}{g^2[\mathbf{u}_i]}\right) \quad (56)$$

in terms of weighting vectors rather than weighting matrices.

From (56) we see that in order to evaluate $E\|\tilde{\mathbf{w}}_i\|_{\sigma}^2$ we need $E\|\tilde{\mathbf{w}}_i\|_{F\sigma}^2$, with weighting vector $F\sigma$. This term can be deduced from (56) by writing it for $\sigma \leftarrow F\sigma$, i.e.,

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

The term $\mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F^2\sigma}^2$ can in turn be deduced from (56) by writing it for $\sigma \leftarrow F^2\sigma$. Continuing in this fashion, for successive powers of F , we arrive at

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F^{M^2-1}\sigma}^2 = \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{F^{M^2}\sigma}^2 + \mu^2 \sigma_v^2 \mathbb{E} \left(\frac{\|\mathbf{u}_i\|_{F^{M^2-1}\sigma}^2}{g^2[\mathbf{u}_i]} \right)$$

in terms of the M^2 -power of F (recall that F is $M^2 \times M^2$).

Fortunately, this procedure terminates. Let $p(x) = \det(xI - F)$ denote the characteristic polynomial of F , say

$$p(x) = x^{M^2} + p_{M^2-1}x^{M^2-1} + p_{M^2-2}x^{M^2-2} + \dots + p_1x + p_0$$

with coefficients $\{p_i\}$. Then, since $p(F) = 0$ in view of the Cayley-Hamilton theorem, we have

$$\mathbb{E} \|\mathbf{w}_i\|_{F^{M^2}\sigma}^2 = \sum_{k=0}^{M^2-1} -p_k \mathbb{E} \|\mathbf{w}_i\|_{F^k\sigma}^2$$

Putting these results together, we conclude that the transient behavior of the filter is described by an M^2 -dimensional state-space model of the form:

$$\boxed{\mathcal{W}_i = \mathcal{F}\mathcal{W}_{i-1} + \mu^2 \sigma_v^2 \mathcal{Y}} \quad (57)$$

where the $M^2 \times 1$ vectors $\{\mathcal{W}_i, \mathcal{Y}\}$ are defined by

$$\mathcal{W}_i = \begin{bmatrix} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{\sigma}^2 \\ \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F\sigma}^2 \\ \vdots \\ \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F^{M^2-2}\sigma}^2 \\ \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F^{M^2-1}\sigma}^2 \end{bmatrix}, \quad \mathcal{Y} = \begin{bmatrix} \mathbb{E} (\|\mathbf{u}_i\|_{\sigma}^2 / g^2[\mathbf{u}_i]) \\ \mathbb{E} (\|\mathbf{u}_i\|_{F\sigma}^2 / g^2[\mathbf{u}_i]) \\ \vdots \\ \mathbb{E} (\|\mathbf{u}_i\|_{F^{M^2-2}\sigma}^2 / g^2[\mathbf{u}_i]) \\ \mathbb{E} (\|\mathbf{u}_i\|_{F^{M^2-1}\sigma}^2 / g^2[\mathbf{u}_i]) \end{bmatrix} \quad (58)$$

and the $M^2 \times M^2$ coefficient matrix \mathcal{F} is given by

$$\mathcal{F} = \begin{bmatrix} 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ \vdots & & & & \\ 0 & 0 & 0 & & 1 \\ -p_0 & -p_1 & -p_2 & \dots & -p_{M^2-1} \end{bmatrix} \quad (59)$$

The learning curve of the filter can be characterized more explicitly as follows. Let $r = \text{vec}(R_u)$ and choose $\sigma = r$. Iterating (56) we find that

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|_r^2 = \|\tilde{\mathbf{w}}_{-1}\|_{F^{i+1}r}^2 + \mu^2 \sigma_v^2 \mathbb{E} \left[\frac{\|\mathbf{u}_i\|_{(I+F+\dots+F^i)r}^2}{g^2[\mathbf{u}_i]} \right]$$

that is,

$$\mathbb{E} \|\tilde{\mathbf{w}}_i\|_r^2 = \|\tilde{\mathbf{w}}_{-1}\|_{a_i}^2 + \mu^2 \sigma_v^2 b(i)$$

where the vector a_i and the scalar $b(i)$ satisfy the recursions

$$\begin{aligned} a_i &= F a_{i-1}, \quad a_{-1} = r \\ b(i) &= b(i-1) + \mathbb{E} \left[\frac{\|\mathbf{u}_i\|_{a_{i-1}}^2}{g^2[\mathbf{u}_i]} \right], \quad b(-1) = 0 \end{aligned}$$

Usually, $\mathbf{w}_{-1} = 0$ so that $\tilde{\mathbf{w}}_{-1} = w^o$. Using the definitions for $\{a_i, b(i)\}$, it is easy to verify that

$$\boxed{\mathbb{E} \mathbf{e}_a^2(i) = \mathbb{E} \mathbf{e}_a^2(i-1) + \|w^o\|_{F^{i-1}(F-I)r}^2 + \mu^2 \sigma_v^2 \mathbb{E} \left(\frac{\|\mathbf{u}_i\|_{F^{i-1}r}^2}{g^2[\mathbf{u}_i]} \right)}$$

which describes the learning curve of a data-normalized adaptive filter.

X. Mean-Square Stability

Recursion (57) shows that the adaptive filter will be mean-square stable if, and only if, the matrix \mathcal{F} is a stable matrix, i.e., all its eigenvalues should lie inside the unit circle. But since \mathcal{F} and F have identical eigenvalues, we conclude that F must be stable. It can be verified that matrices F of the form (54), for arbitrary $\{A > 0, B \geq 0\}$, are stable for any step-size μ in the range:

$$0 < \mu < \min \left\{ \frac{1}{\lambda_{\max}(A^{-1}B)}, \frac{1}{\max\{\lambda(H) \in \mathbb{R}^+\}} \right\} \quad (60)$$

where the second condition is in terms of the largest positive real eigenvalue of the block matrix,

$$H \triangleq \begin{bmatrix} A/2 & -B/2 \\ I_{M^2} & 0 \end{bmatrix}$$

when it exists. Since H is not symmetric, its eigenvalues may not be positive or even real. If H does not have any real positive eigenvalue, then the upper bound on μ is determined by $1/\lambda_{\max}(A^{-1}B)$ alone. Likewise, the mean-stability of the filter, as dictated by (53), requires

$$\mu < 2/\lambda_{\max}(P) \quad (61)$$

XI. Steady-State Performance

In Sec. IV we evaluated the steady-state performance of filters (3) by relying on the variance relation (10). Obviously, steady-state results can also be deduced from (56) by letting $i \rightarrow \infty$, as we show below. However, it is worth pointing out that any steady-state result that is obtained from (56) would be bound by the same assumptions that were used to establish (56), such as the independence assumption (50). In contrast, the steady-state results deduced earlier in Sec. IV were derived under weaker conditions on the data.

To see how steady-state results can be obtained from (56), assume filter operation in steady-state. Then recursion (56) gives in the limit

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{(I-F)\sigma}^2 = \mu^2 \sigma_v^2 \mathbb{E} \left[\frac{\|\mathbf{u}_i\|_{\sigma}^2}{g^2[\mathbf{u}_i]} \right] \quad (62)$$

Now recall that the filter excess mean-square error is defined by

$$\text{EMSE} = \lim_{i \rightarrow \infty} \mathbb{E} \mathbf{e}_a^2(i) = \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{R_u}^2$$

This suggests that we should select σ in (62) as

$$\sigma_{\text{emse}} = (I - F)^{-1} \text{vec}(R_u)$$

On the other hand, the filter mean-square deviation, defined as

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

is obtained by choosing

$$\sigma_{\text{msd}} = (I - F)^{-1} \text{vec}(I)$$

Let $\{\Sigma_{\text{emse}}, \Sigma_{\text{msd}}\}$ denote the weighting matrices that correspond to $\{\sigma_{\text{emse}}, \sigma_{\text{msd}}\}$. Then we conclude that

$$\boxed{\begin{aligned} \text{EMSE} &= \mu^2 \sigma_v^2 \text{Tr}(Q \Sigma_{\text{emse}}) \\ \text{MSD} &= \mu^2 \sigma_v^2 \text{Tr}(Q \Sigma_{\text{msd}}) \end{aligned}} \quad (63)$$

XII. Small Step-Size Approximation

A simplification is possible when the step-size is sufficiently small. Observe that the matrix F in (54), as well as the filter EMSE and MSD in (63), are defined in terms of the matrices $\{A, B, P, Q\}$. These moments are hard to evaluate for arbitrary input distributions and arbitrary functions g . However, for small step-sizes, we may ignore the quadratic term in μ^2 that appears in (52), and approximate the variance relation (51)–(52) by

$$\boxed{\begin{aligned} \text{E} \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 &= \text{E} \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2 + \mu^2 \sigma_v^2 \text{E} \left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]} \right) \\ \Sigma' &= \Sigma - \mu \Sigma P - \mu P \Sigma \end{aligned}} \quad (64)$$

where $P = \text{E}(\mathbf{u}_i^T \mathbf{u}_i / g[\mathbf{u}_i])$. Moreover, since now $F = I - \mu A$, we can also approximate the EMSE and MSD performances (63) of the filter by

$$\boxed{\begin{aligned} \text{EMSE} &\approx \mu \sigma_v^2 \text{Tr}(Q \Sigma_{\text{emse}}) \\ \text{MSD} &\approx \mu \sigma_v^2 \text{Tr}(Q \Sigma_{\text{msd}}) \end{aligned}} \quad (65)$$

where $\{\Sigma_{\text{emse}}, \Sigma_{\text{msd}}\}$ denote the weighting matrices that correspond to the vectors

$$\sigma_{\text{emse}} = A^{-1} \text{vec}(R_u), \quad \sigma_{\text{msd}} = A^{-1} \text{vec}(I)$$

For example, in the special case of LMS, for which $g[\mathbf{u}_i] = 1$ and $P = R_u = Q$, the above expressions give for small step-sizes:

$$\text{EMSE} \approx \frac{\mu \sigma_v^2 \text{Tr}(R_u)}{2}, \quad \text{MSD} \approx \frac{\mu \sigma_v^2 M}{2} \quad (\text{LMS}) \quad (66)$$

Actually, using the simplified variance relation (64), we can describe the dynamic behavior of the mean-square deviation of the filter by means of an M -dimensional state-space model, as opposed to the M^2 -dimensional model (57). To see this, let $P = U \Delta U^T$ denote the eigen-decomposition of $P > 0$, and introduce the transformed quantities:

$$\bar{\mathbf{w}}_i = U^T \tilde{\mathbf{w}}_i, \quad \bar{\mathbf{u}}_i = \mathbf{u}_i U, \quad \bar{\Sigma} = U^T \Sigma U, \quad \bar{Q} = U^T Q U$$

Then the variance relation (64) can be equivalently rewritten as⁴

$$\begin{aligned} \text{E} \|\bar{\mathbf{w}}_i\|_{\bar{\Sigma}}^2 &= \text{E} \|\bar{\mathbf{w}}_{i-1}\|_{\bar{\Sigma}'}^2 + \mu^2 \sigma_v^2 \text{E} \left(\frac{\|\bar{\mathbf{u}}_i\|_{\bar{\Sigma}}^2}{g^2[\bar{\mathbf{u}}_i]} \right) \\ \bar{\Sigma}' &= \bar{\Sigma} - \mu \bar{\Sigma} \Delta - \mu \Delta \bar{\Sigma} \end{aligned} \quad (67)$$

The expression for $\bar{\Sigma}'$ shows that it will be diagonal as long as $\bar{\Sigma}$ is diagonal. Therefore, since we are free to choose Σ (and, consequently, $\bar{\Sigma}$), we can assume that $\bar{\Sigma}'$ is diagonal. In this way, $\{\bar{\Sigma}, \bar{\Sigma}'\}$ will be fully characterized by their diagonal entries. Thus let $\{\bar{\sigma}, \bar{\sigma}'\}$ denote $M \times 1$ vectors that collect the diagonal entries of $\{\bar{\Sigma}, \bar{\Sigma}'\}$, i.e.,

$$\bar{\sigma} = \text{diag}(\bar{\Sigma}), \quad \bar{\sigma}' = \text{diag}(\bar{\Sigma}')$$

Then from (67) we find that

$$\bar{\sigma}' = \bar{F} \bar{\sigma}$$

where \bar{F} is $M \times M$ and given by

$$\bar{F} = I - \mu \bar{A}, \quad \bar{A} = 2\Delta$$

Repeating the arguments that led to (57) we can then establish that, for sufficiently small step-sizes, the evolution of $\text{E} \|\bar{\mathbf{w}}_i\|_{\bar{\sigma}}^2$ is described by the following M -dimensional state-space model:

$$\boxed{\bar{\mathcal{W}}_i = \bar{\mathcal{F}} \bar{\mathcal{W}}_{i-1} + \mu^2 \sigma_v^2 \bar{\mathcal{Y}}}$$
 (68)

where the $M \times 1$ vectors $\{\bar{\mathcal{W}}_i, \bar{\mathcal{Y}}\}$ are defined by

$$\bar{\mathcal{W}}_i = \begin{bmatrix} \text{E} \|\bar{\mathbf{w}}_i\|_{\bar{\sigma}}^2 \\ \text{E} \|\bar{\mathbf{w}}_i\|_{\bar{\sigma}}^2 / \bar{F} \bar{\sigma} \\ \vdots \\ \text{E} \|\bar{\mathbf{w}}_i\|_{\bar{\sigma}}^2 / \bar{F}^{M-2} \bar{\sigma} \\ \text{E} \|\bar{\mathbf{w}}_i\|_{\bar{\sigma}}^2 / \bar{F}^{M-1} \bar{\sigma} \end{bmatrix}, \quad \bar{\mathcal{Y}} = \begin{bmatrix} \text{E} (\|\bar{\mathbf{u}}_i\|_{\bar{\sigma}}^2 / g^2[\bar{\mathbf{u}}_i]) \\ \text{E} (\|\bar{\mathbf{u}}_i\|_{\bar{\sigma}}^2 / g^2[\bar{\mathbf{u}}_i]) \\ \vdots \\ \text{E} (\|\bar{\mathbf{u}}_i\|_{\bar{\sigma}}^2 / g^2[\bar{\mathbf{u}}_i]) \\ \text{E} (\|\bar{\mathbf{u}}_i\|_{\bar{\sigma}}^2 / g^2[\bar{\mathbf{u}}_i]) \end{bmatrix} \quad (69)$$

and the $M \times M$ coefficient matrix $\bar{\mathcal{F}}$ is given by

$$\bar{\mathcal{F}} = \begin{bmatrix} 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ \vdots & & & & \\ 0 & 0 & 0 & & 1 \\ -\bar{p}_0 & -\bar{p}_1 & -\bar{p}_2 & \cdots & -\bar{p}_{M-1} \end{bmatrix} \quad (70)$$

where the $\{\bar{p}_i\}$ are the coefficients of the characteristic polynomial of \bar{F} . If we select $\bar{\sigma} = \text{vec}(I)$ then

$$\|\bar{\mathbf{w}}_i\|_{\bar{\sigma}}^2 = \|\bar{\mathbf{w}}_i\|^2 = \|\tilde{\mathbf{w}}_i\|^2$$

since U is orthogonal. In this case, the top entry of the resulting state vector $\bar{\mathcal{W}}_i$ will describe the evolution of the filter MSD.

⁴Usually, $g[\cdot]$ is invariant under orthogonal transformations, i.e., $g[\mathbf{u}_i] = g[\bar{\mathbf{u}}_i]$.

XIII. Fourth-Order Moment Approximation

Besides the small-step size approximation of Sec. XII, we could have instead chosen to approximate the fourth-order moment that appears in the expression for Σ' in (52) as

$$\mathbb{E} \left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g^2[\mathbf{u}_i]} \mathbf{u}_i^T \mathbf{u}_i \right) \approx \mathbb{E} \left(\frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]} \right) \cdot \mathbb{E} \left(\frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g[\mathbf{u}_i]} \right) = P \text{Tr}(\Sigma P)$$

where $P = \mathbb{E}(\mathbf{u}_i^T \mathbf{u}_i / g[\mathbf{u}_i])$. In this way, expression (52) for Σ' would become

$$\Sigma' = \Sigma - \mu \Sigma P - \mu P \Sigma + \mu^2 P \text{Tr}(P \Sigma) \quad (71)$$

which is fully characterized in terms of the single moment P . If we now let $P = U \Delta U^T$ denote the eigen-decomposition of $P > 0$, and introduce the transformed quantities

$$\bar{\mathbf{w}}_i = U^T \tilde{\mathbf{w}}_i, \quad \bar{\mathbf{u}}_i = \mathbf{u}_i U, \quad \bar{\Sigma} = U^T \Sigma U$$

Then the variance relations (51) and (71) can be equivalently rewritten as

$$\begin{aligned} \mathbb{E} \|\bar{\mathbf{w}}_i\|_{\bar{\Sigma}}^2 &= \mathbb{E} \|\bar{\mathbf{w}}_{i-1}\|_{\bar{\Sigma}'}^2 + \mu^2 \sigma_v^2 \mathbb{E} \left(\frac{\|\bar{\mathbf{u}}_i\|_{\bar{\Sigma}}^2}{g^2[\bar{\mathbf{u}}_i]} \right) \\ \bar{\Sigma}' &= \bar{\Sigma} - \mu \bar{\Sigma} \bar{\Delta} - \mu \bar{\Delta} \bar{\Sigma} + \mu^2 \bar{\Delta} \text{Tr}(\bar{\Sigma} \bar{\Delta}) \end{aligned} \quad (72)$$

Again, $\bar{\Sigma}'$ will be diagonal as long as $\bar{\Sigma}$ is. Thus let

$$\bar{\sigma} = \text{diag}(\bar{\Sigma}), \quad \bar{\sigma}' = \text{diag}(\bar{\Sigma}')$$

Then from (72) we find that $\bar{\sigma}' = \bar{F} \bar{\sigma}$, where \bar{F} is $M \times M$ and given by

$$\bar{F} = I - \mu \bar{A} + \mu^2 \bar{B}, \quad \bar{A} = 2\bar{\Delta}, \quad \bar{B} = \mu^2 \delta \delta^T$$

where $\delta = \text{diag}(\bar{\Delta})$. Repeating the arguments that led to (57) we can establish that, under the assumed fourth-order moment approximation, the evolution of $\mathbb{E} \|\bar{\mathbf{w}}_i\|_{\bar{\Sigma}}^2$ is described by an M -dimensional state-space model similar to (68).

XIV. Transient Behavior of APA

As with the steady-state analysis of Sec. IV, the transient analysis of Secs. VI–XIII can also be extended to treat more general adaptive updates than (3), e.g., the APA algorithm (26)–(27). To illustrate this fact, define the weighted a-priori and a-posteriori error vectors,

$$\mathbf{e}_{a,i}^{\Sigma} = \mathbf{U}_i \Sigma \tilde{\mathbf{w}}_{i-1}, \quad \mathbf{e}_{p,i}^{\Sigma} = \mathbf{U}_i \Sigma \tilde{\mathbf{w}}_i$$

Then it follows from the APA recursion that the weight-error vector satisfies

$$\tilde{\mathbf{w}}_i = (I - \mu \mathbf{U}_i^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{U}_i) \tilde{\mathbf{w}}_{i-1} - \mu \mathbf{U}_i^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{v}_i$$

while the energy relation (28) extends to

$$\begin{aligned} \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 + (\mathbf{e}_{a,i}^{\Sigma})^T (\mathbf{U}_i \Sigma \mathbf{U}_i^T)^{-1} \mathbf{e}_{a,i}^{\Sigma} &= \\ \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma}^2 + (\mathbf{e}_{p,i}^{\Sigma})^T (\mathbf{U}_i \Sigma \mathbf{U}_i^T)^{-1} \mathbf{e}_{p,i}^{\Sigma} & \end{aligned}$$

If we assume the sequence $\{\mathbf{U}_i\}$ is iid, and follow the arguments of Sec. VIII, we can verify by taking expectations of both sides of the above equality that the variance relation (48) is replaced by

$$\begin{aligned} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 &= \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2 + \mu^2 \mathbb{E}(\mathbf{v}_i^T \mathbf{A}_i^{\Sigma} \mathbf{v}_i) \\ \Sigma' &= \Sigma - \mu \Sigma \mathbb{E} \mathbf{P}_i - \mu \mathbf{P}_i \Sigma + \mu^2 \mathbb{E}[\mathbf{P}_i \Sigma \mathbf{P}_i] \\ \mathbf{P}_i &= \mathbf{U}_i^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{U}_i \\ \mathbf{A}_i^{\Sigma} &= (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \mathbf{U}_i \Sigma \mathbf{U}_i^T (\mathbf{U}_i \mathbf{U}_i^T)^{-1} \end{aligned} \quad (73)$$

Moreover,

$$\mathbb{E} \tilde{\mathbf{w}}_i = (I - \mu \mathbb{E}(\mathbf{P}_i)) \mathbb{E} \tilde{\mathbf{w}}_{i-1}$$

As in Sec. IX, we can now conclude that the transient behavior of APA is described by the M^2 -dimensional recursion

$$\mathcal{W}_i = \mathcal{F} \mathcal{W}_{i-1} + \mu^2 \sigma_v^2 \mathcal{Y}$$

where \mathcal{F} is the $M^2 \times M^2$ companion matrix associated with

$$\mathcal{F} \triangleq I - \mu(\mathbb{E}[\mathbf{P}_i] \otimes I) - \mu(I \otimes \mathbb{E}[\mathbf{P}_i]) + \mu^2 \mathbb{E}[\mathbf{P}_i \otimes \mathbf{P}_i]$$

Also,

$$\mathcal{W}_i \triangleq \begin{bmatrix} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{\sigma}^2 \\ \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F\sigma}^2 \\ \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F^2\sigma}^2 \\ \vdots \\ \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{F^{(M^2-1)}\sigma}^2 \end{bmatrix}$$

and the k -th entry of \mathcal{Y} is

$$[\mathcal{Y}]_k = \gamma^T F^k \sigma, \quad k = 0, \dots, M^2 - 1$$

where $\gamma = \text{vec}(\mathbb{E} \mathbf{U}_i^T (\mathbf{U}_i \mathbf{U}_i^T)^{-2} \mathbf{U}_i)$. It follows that APA is mean-square stable for step-sizes in the range

$$0 < \mu < \min \left\{ \frac{2}{\lambda_{\max}(\mathbb{E} \mathbf{P}_i)}, \frac{1}{\lambda_{\max}(A^{-1}B)}, \frac{1}{\max\{\lambda(H) \in \mathbb{R}^+\}} \right\}$$

where

$$A = (\mathbb{E}[\mathbf{P}_i] \otimes I) + (I \otimes \mathbb{E}[\mathbf{P}_i]), \quad B = \mathbb{E}(\mathbf{P}_i \otimes \mathbf{P}_i)$$

and

$$H = \begin{bmatrix} A/2 & -B/2 \\ I & 0 \end{bmatrix}$$

Moreover, the MSD and EMSE of APA can be obtained from the limiting behavior of the variance relation in (73) as

$$\begin{aligned} \text{MSD} &= \mu^2 \sigma_v^2 \gamma^T (I - F)^{-1} \text{vec}(I) \\ \text{EMSE} &= \mu^2 \sigma_v^2 \gamma^T (I - F)^{-1} \text{vec}(R_u) \end{aligned}$$

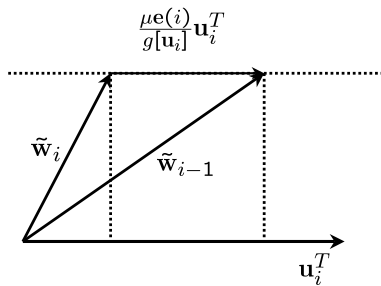


Fig. 1. Geometric interpretation of the energy relation (9).

XV. Geometric Interpretation

We end our discussions with a geometric interpretation of the energy relation (9). Observe from (6) that

$$\tilde{\mathbf{w}}_{i-1} - \tilde{\mathbf{w}}_i = \left(\frac{\mu \mathbf{e}(i)}{g[\mathbf{u}_i]} \right) \mathbf{u}_i^T$$

This means that the difference vector ($\tilde{\mathbf{w}}_{i-1} - \tilde{\mathbf{w}}_i$) is always parallel to the transposed regression vector, \mathbf{u}_i^T , as shown in the figure. Therefore, the distances from the vertices of $\{\tilde{\mathbf{w}}_i, \tilde{\mathbf{w}}_{i-1}\}$ to \mathbf{u}_i^T should coincide.

Let $\{\theta_{i-1}, \theta_i\}$ denote the acute angles between $\{\tilde{\mathbf{w}}_{i-1}, \tilde{\mathbf{w}}_i\}$ and \mathbf{u}_i^T . Then the squared norm of the projection of $\tilde{\mathbf{w}}_i$ onto the vertical direction is given by

$$\|\tilde{\mathbf{w}}_i\|^2 \sin^2(\theta_i) = \|\tilde{\mathbf{w}}_i\|^2 \left[1 - \frac{(\mathbf{u}_i \tilde{\mathbf{w}}_i)^2}{\|\tilde{\mathbf{w}}_i\|^2 \cdot \|\mathbf{u}_i\|^2} \right]$$

Likewise, the squared norm of the projection of $\tilde{\mathbf{w}}_{i-1}$ onto the vertical direction is given by

$$\|\tilde{\mathbf{w}}_{i-1}\|^2 \sin^2(\theta_{i-1}) = \|\tilde{\mathbf{w}}_{i-1}\|^2 \left[1 - \frac{(\mathbf{u}_i \tilde{\mathbf{w}}_{i-1})^2}{\|\tilde{\mathbf{w}}_{i-1}\|^2 \cdot \|\mathbf{u}_i\|^2} \right]$$

Equating these norms we have

$$\|\tilde{\mathbf{w}}_i\|^2 \left[1 - \frac{e_a^2(i)}{\|\tilde{\mathbf{w}}_i\|^2 \cdot \|\mathbf{u}_i\|^2} \right] = \|\tilde{\mathbf{w}}_{i-1}\|^2 \left[1 - \frac{e_p^2(i)}{\|\tilde{\mathbf{w}}_{i-1}\|^2 \cdot \|\mathbf{u}_i\|^2} \right]$$

which, after expanding, agrees with the energy relation (9). Note further that the equality of the vertical norms implies that

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

and this result brings forth a connection between the energy relation (9) and Snell's law in optics.

XVI. Concluding Remarks

This article described an energy-conservation approach to the analysis of adaptive filter performance. By studying the energy balance at each iteration, the dynamic behavior of an adaptive filter can be characterized in terms of a variance relation (e.g., (51)) and, subsequently, in terms of a state-space model (e.g., (57)). The approach does not restrict the input data to being Gaussian or white. Besides providing information about the stability and convergence

behavior of the filter, the energy-conservation arguments also help characterize the steady-state performance of the filter. While the transient analysis of Secs. IX–XIII relies on the independence assumption (50), steady-state results can be obtained without relying on this assumption, as shown in Secs. IV.

Acknowledgment

The results described in this article are parts of joint works with several students and associates over the last few years: M. Rupp (Sec. III on the energy-conservation relation) N. R. Yousef (Secs. IV and V on steady-state and tracking performances), T. Y. Al-Naffouri (Secs. VI–XI on transient analysis), V. H. Nascimento (Sec. X on mean-square stability), and H. Shin (Secs. IV-D and XIV on affine projection algorithms).

REFERENCES

- [1] 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, 1995.
- [2] A. H. Sayed and M. Rupp, "Robustness issues in adaptive filtering," in *DSP Handbook*, Chapter 20, CRC Press, 1998.
- [3] 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, Jun. 1996.
- [4] 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, Jan. 2000.
- [5] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. Signal Processing*, vol. 49, no. 2, pp. 314–324, Feb. 2001.
- [6] N. R. Yousef and A. H. Sayed, "Ability of adaptive filters to track carrier offsets and random channel nonstationarities," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1533–1544, July 2002.
- [7] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," to appear in *IEEE Transactions on Signal Processing*, 2003.
- [8] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," to appear in *IEEE Transactions on Signal Processing*, 2003.
- [9] A. H. Sayed, T. Y. Al-Naffouri, and V. H. Nascimento, "Energy conservation in adaptive filtering," in *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*, K. Barner et al., editors, CRC Press, 2003.
- [10] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, NJ, 1996.
- [11] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice Hall, NJ, 1985.
- [12] O. Macchi, *Adaptive Processing: The LMS Approach with Applications in Transmission*, Wiley, NY, 1995.
- [13] A. H. Sayed and M. Rupp, "An l_2 -stable feedback structure for nonlinear adaptive filtering and identification," *Automatica*, vol. 33, no. 1, pp. 13–30, Jan. 1997.
- [14] M. Rupp and A. H. Sayed, "On the convergence of blind adaptive equalizers for constant-modulus signals," *IEEE Transactions on Communications*, vol. 48, no. 5, pp. 795–803, May 2000.