# Chapter 1

**In *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*, K. E. Barner and G. Arce, editors, CRC Press,pp. 1–35, 2003.**

# Energy Conservation in Adaptive Filtering

ALI H. SAYED[1]
Electrical Engineering Department
University of California
Los Angeles, CA
Email: sayed@ee.ucla.edu

TAREQ Y. AL-NAFFOURI
Department of Electrical Engineering
Stanford University
Stanford, CA
Email: nafouri@stanford.edu

VITOR H. NASCIMENTO
Department of Electrical Engineering
University of Sao Paulo
Sao Paulo, Brazil
Email: vitor@lps.usp.br

[1]This work was supported in part by the National Science Foundation under grants ECS-9820765 and CCR-0208573. The work of T. Y. Al-Naffouri was also supported by a fellowship from King Fahd University of Petroleum and Minerals, Saudi Arabia.

## 1.1 Introduction

The study of the steady-state and transient performance of adaptive filters is a challenging task due to the nonlinear and stochastic nature of their update equations (see, e.g., [1]–[3]). The purpose of this chapter is to provide an overview of an energy-conservation approach to studying the performance of adaptive filters in a unified manner. The approach is based on showing that certain a-priori and a-posteriori errors maintain an energy balance for all time instants [4]–[6]. When examined under expectation, this energy balance leads to a variance relation that characterizes the dynamics of an adaptive filter [9]–[13]. An advantage of the energy framework is that it allows us to push the algebraic manipulations of variables to a limit, and to eliminate unnecessary cross-terms before appealing to expectations. This is a useful step because it is usually easier to handle random variables algebraically than under expectations, especially for higher-order moments. A second advantage of the energy arguments is that they can be pursued without restricting the distribution of the input data. To illustrate this point, we have opted not to restrict the regression data to being Gaussian or white in most of the discussions below. Instead, all results are derived for arbitrary input distributions. Of course, by specializing the results to particular distributions, some known results from the literature can be recovered as special cases of the general framework.

As with most adaptive filter analyses, progress is difficult without relying on simplifying assumptions. In the initial part of our presentation, we derive exact energy-conservation and variance relations that hold for a large class of adaptive filters and without any approximations. Subsequent discussions will call upon simplifying assumptions in order to make the analysis more tractable. The assumptions tend to be reasonable for small step-sizes and long filters.

## 1.2 The Data Model

Consider reference data $\{\mathbf{d}(i)\}$ and regression data $\{\mathbf{u}_i\}$, assumed related via the linear regression model

$$\boxed{\mathbf{d}(i) = \mathbf{u}_i w^o \; + \; \mathbf{v}(i)} \tag{1.1}$$

for some $M \times 1$ unknown column vector $w^o$ that we wish to estimate. Here $\mathbf{u}_i$ is a regressor, taken as a row vector, and $\mathbf{v}(i)$ is measurement noise. Observe that we are using boldface letters to denote random quantities, which will be our convention throughout this chapter. Also, all vectors in our presentation are column vectors except for the regressor $\mathbf{u}_i$. In this way, the inner product between $\mathbf{u}_i$ and $w^o$ is written simply as $\mathbf{u}_i w^o$ without the need for transposition symbols.

In Eq. (1.1), $\{\mathbf{d}(i), \mathbf{u}_i, \mathbf{v}(i)\}$ are random variables that satisfy the following conditions:

$$\begin{array}{l} \text{a)} \{\mathbf{v}(i)\} \text{ is zero-mean, independent and identically distributed with variance } \mathrm{E}\,\mathbf{v}^2(i) = \sigma_v^2. \\ \text{b) } \mathbf{v}(i) \text{ is independent of } \mathbf{u}_j \text{ for all } i, j. \\ \text{c) The regressor } \mathbf{u}_i \text{ is zero-mean and has covariance matrix } \mathrm{E}\,\mathbf{u}_i^T\mathbf{u}_i = R_u > 0. \end{array} \tag{1.2}$$

We focus in the first part of the chapter on data-normalized adaptive filters for generating estimates for $w^o$, viz., on updates of the form

$$\boxed{\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]}\mathbf{e}(i), \quad i \geq 0} \tag{1.3}$$

where

$$\boxed{\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{u}_i\mathbf{w}_{i-1}} \tag{1.4}$$

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

$$g[\mathbf{u}] = 1 \quad \text{(LMS)}, \qquad g[\mathbf{u}] = \|\mathbf{u}\|^2 \quad \text{(NLMS)}, \qquad g[\mathbf{u}] = \epsilon + \|\mathbf{u}\|^2 \quad (\epsilon\text{-NLMS})$$

The initial condition $\mathbf{w}_{-1}$ of (1.3) is assumed to be independent of all $\{\mathbf{d}(j), \mathbf{u}_j, \mathbf{v}(j)\}$. Later in the chapter we study adaptive filters with error nonlinearities in their update equations — see Eq. (1.52).

Our purpose is to examine the transient and steady-state performance of such data-normalized filters in a unified manner (i.e., uniformly for all $g$). The first step in this regard is to establish an energy-conservation relation that holds for a large class of adaptive filters, and then use it as the basis of all subsequent analysis.

## 1.3 Energy-Conservation Relation

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

$$\mathbf{e}_a^{\Sigma}(i) \triangleq \mathbf{u}_i\Sigma\tilde{\mathbf{w}}_{i-1}, \qquad \mathbf{e}_p^{\Sigma}(i) \triangleq \mathbf{u}_i\Sigma\tilde{\mathbf{w}}_i \tag{1.5}$$

When $\Sigma = I$, we recover the standard definitions

$$\mathbf{e}_a(i) = \mathbf{u}_i\tilde{\mathbf{w}}_{i-1}, \qquad \mathbf{e}_p(i) = \mathbf{u}_i\tilde{\mathbf{w}}_i \tag{1.6}$$

The freedom in selecting $\Sigma$ will be seen to be useful in characterizing several aspects of the dynamic behavior of an adaptive filter. For now, we shall treat $\Sigma$ as an arbitrary weighting matrix.

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 a fundamental energy-conservation relation. To arrive at the relation, we subtract $w^o$ from both sides of (1.3) to get

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]} \mathbf{e}(i) \tag{1.7}$$

and then multiply (1.7) by $\mathbf{u}_i \Sigma$ from the left to conclude that

$$\mathbf{e}_p^{\Sigma}(i) = \mathbf{e}_a^{\Sigma}(i) - \mu \frac{\|\mathbf{u}_i\|_{\Sigma}^2}{g[\mathbf{u}_i]} \mathbf{e}(i) \tag{1.8}$$

where the notation $\|\mathbf{u}_i\|_{\Sigma}^2$ denotes the squared weighted Euclidean norm of $\mathbf{u}_i$, viz.,

$$\|\mathbf{u}_i\|_{\Sigma}^2 \; = \mathbf{u}_i \Sigma \mathbf{u}_i^T$$

Relation (1.8) can be used to express $\mathbf{e}(i)/g[\mathbf{u}_i]$ in terms of $\{\mathbf{e}_p^{\Sigma}(i), \mathbf{e}_a^{\Sigma}(i)\}$ and to eliminate this term from (1.7). Doing so leads to the equality

$$\|\mathbf{u}_i\|_{\Sigma}^2 \cdot \tilde{\mathbf{w}}_i + \mathbf{u}_i^T \mathbf{e}_a^{\Sigma}(i) \;=\; \|\mathbf{u}_i\|_{\Sigma}^2 \cdot \tilde{\mathbf{w}}_{i-1} + \mathbf{u}_i^T \mathbf{e}_p^{\Sigma}(i) \tag{1.9}$$

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

$$\boxed{\|\mathbf{u}_i\|_{\Sigma}^2 \cdot \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 \;+\; \left(\mathbf{e}_a^{\Sigma}(i)\right)^2 = \|\mathbf{u}_i\|_{\Sigma}^2 \cdot \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma}^2 \;+\; \left(\mathbf{e}_p^{\Sigma}(i)\right)^2} \tag{1.10}$$

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.[2] In addition, it follows from $\mathbf{e}(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i-1} + \mathbf{v}(i)$, and from (1.7), that the weight-error vector satisfies

$$\boxed{\tilde{\mathbf{w}}_i = \left(I - \mu \frac{\mathbf{u}_i^T \mathbf{u}_i}{g[\mathbf{u}_i]}\right) \tilde{\mathbf{w}}_{i-1} \;-\; \mu \frac{\mathbf{u}_i^T}{g[\mathbf{u}_i]} \mathbf{v}(i)} \tag{1.11}$$

## 1.4 Weighted Variance Relation

The result (1.10) with $\Sigma = I$ was developed in [4] and subsequently used in a series of works to study the robustness of adaptive filters (e.g., [5]–[8]). It was later used in [9]–[11] to study the steady-state and tracking performance of adaptive filters. The incorporation of a weighting matrix $\Sigma$ in [12, 13] turns out to be useful for transient (convergence and stability) analysis.

In transient analysis we are interested in characterizing the time evolution of the quantity $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2$, for some $\Sigma$ of interest (usually, $\Sigma = I$ or $\Sigma = R_u$). To arrive at this evolution,

[2]Later in Sec. 1.13 we shall provide an interpretation of the energy relation (1.10) in terms of Snell's Law for light propagation.

we use (1.8) to replace $\mathbf{e}_p^{\Sigma}(i)$ in (1.10) in terms of $\mathbf{e}_a^{\Sigma}(i)$ and $\mathbf{e}(i)$. This step yields, after expanding and grouping terms,

$$\begin{aligned}
\|\mathbf{u}_i\|_\Sigma^2 \cdot \|\tilde{\mathbf{w}}_i\|_\Sigma^2 \;=\; & \|\mathbf{u}_i\|_\Sigma^2 \cdot \|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 \;+ \\
& \frac{\mu^2(\|\mathbf{u}_i\|_\Sigma^2)^2}{g^2[\mathbf{u}_i]}\;\|\tilde{\mathbf{w}}_{i-1}\|^2_{\mathbf{u}_i^T\mathbf{u}_i} \;+ \\
& \frac{\mu^2(\|\mathbf{u}_i\|_\Sigma^2)^2}{g^2[\mathbf{u}_i]}\;\mathbf{v}^2(i) \;- \\
& \frac{\mu\|\mathbf{u}_i\|_\Sigma^2}{g[\mathbf{u}_i]}\;\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma\mathbf{u}_i^T\mathbf{u}_i+\mathbf{u}_i^T\mathbf{u}_i\Sigma} \;+ \\
& 2\mu^2\frac{(\|\mathbf{u}_i\|_\Sigma^2)^2}{g^2[\mathbf{u}_i]}\;\mathbf{v}(i)\mathbf{e}_a(i) \;- \\
& 2\mu\frac{\|\mathbf{u}_i\|_\Sigma^2}{g[\mathbf{u}_i]}\mathbf{v}(i)\mathbf{e}_a^{\Sigma}(i)
\end{aligned} \tag{1.12}$$

Assuming the event $\|\mathbf{u}_i\|_\Sigma^2 = 0$ occurs with zero probability, we can eliminate $\|\mathbf{u}_i\|_\Sigma^2$ from both sides of (1.12) and take expectations to arrive at:

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|_\Sigma^2 = \mathrm{E}\left(\|\tilde{\mathbf{w}}_{i-1}\|^2_{\mathbf{\Sigma}'}\right) \;+\; \mu^2\sigma_v^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|_\Sigma^2}{g^2[\mathbf{u}_i]}\right)} \tag{1.13}$$

where the weighting matrix $\mathbf{\Sigma}'$ is defined by

$$\boxed{\mathbf{\Sigma}' = \Sigma - \frac{\mu}{g[\mathbf{u}_i]}\Sigma\mathbf{u}_i^T\mathbf{u}_i \;-\; \frac{\mu}{g[\mathbf{u}_i]}\mathbf{u}_i^T\mathbf{u}_i\Sigma + \frac{\mu^2\|\mathbf{u}_i\|_\Sigma^2}{g^2[\mathbf{u}_i]}\mathbf{u}_i^T\mathbf{u}_i} \tag{1.14}$$

Observe that $\mathbf{\Sigma}'$ is a random matrix due to its dependence on the data (and, hence, the use of the boldface notation for it). The matrix $\Sigma$, on the other hand, is not random.

### 1.4.1 Independent Regressors

Relations (1.11), (1.13) and (1.14) characterize the dynamic behavior of data-normalized adaptive filters for generic input distributions; they are all exact relations. Still, recursion (1.13) is hard to propagate since it requires the evaluation of the expectation

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

The difficulty is due to the fact that $\mathbf{\Sigma}'$ is a random matrix that depends on $\mathbf{u}_i$, and $\tilde{\mathbf{w}}_{i-1}$ is dependent on prior regressors as well. In order to progress further in the analysis, we shall assume that

$$\boxed{\text{The } \{\mathbf{u}_i\} \text{ are independent and identically distributed}} \tag{1.15}$$

which allows us to deal with $\mathbf{\Sigma}'$ independently from $\tilde{\mathbf{w}}_{i-1}$. This so-called independence assumption is commonly used in the literature. Although rarely applicable, it gives good results for small step-sizes.

Under (1.15), it is easy to verify that $\tilde{\mathbf{w}}_{i-1}$ becomes independent of $\mathbf{\Sigma}'$ and, consequently, that

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

with the weighting matrix $\mathbf{\Sigma}'$ replaced by its mean, which we shall denote by $\Sigma'$. In this way, the variance recursion (1.13) becomes

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{\Sigma} = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma'} \;+\; \mu^2\sigma_v^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2[\mathbf{u}_i]}\right)} \tag{1.16}$$

with deterministic weighting matrices $\{\Sigma, \Sigma'\}$ and where, by evaluating the expectation of (1.14),

$$\boxed{\Sigma' = \Sigma \;-\; \mu\Sigma\mathrm{E}\left(\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right) \;-\; \mu\mathrm{E}\left(\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right)\Sigma \;+\; \mu^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2[\mathbf{u}_i]}\mathbf{u}_i^T\mathbf{u}_i\right)} \tag{1.17}$$

Observe that the expression for $\Sigma'$ is data-dependent *only.*

Finally, taking expectations of both sides of (1.11), and using (1.15), we find that

$$\boxed{\mathrm{E}\,\tilde{\mathbf{w}}_i = \left(I - \mu\mathrm{E}\left(\frac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right)\right)\cdot\mathrm{E}\,\tilde{\mathbf{w}}_{i-1}} \tag{1.18}$$

The expressions (1.16)–(1.18) show that studying the transient behavior of a data-normalized adaptive filter in effect requires evaluating the three multivariate moments:

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

which are functions of $\mathbf{u}_i$ only. In terms of these moments, relations (1.16)–(1.18) can now be used to characterize the dynamic behavior of adaptive filters under the independence assumption (1.15). We start with the mean-square (transient) behavior.

## 1.5 Mean-Square Behavior

Let $\sigma$ denote the $M^2 \times 1$ column vector that is obtained by stacking the columns of $\Sigma$ on top of each other, written as $\sigma = \mathrm{vec}(\Sigma)$. Likewise, let $\sigma' = \mathrm{vec}(\Sigma')$. We shall also use the $\mathrm{vec}^{-1}(\cdot)$ notation and write $\Sigma = \mathrm{vec}^{-1}(\sigma)$ to recover $\Sigma$ from $\sigma$. Similarly, $\Sigma' = \mathrm{vec}^{-1}(\sigma')$.

Then using the Kronecker product notation [15], and the following property, for arbitrary matrices $\{X, Y, Z\}$ of compatible dimensions,

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

we can easily verify that relation (1.17) for $\Sigma'$ transforms into the linear vector relation

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

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

$$\boxed{F = I - \mu A + \mu^2 B} \tag{1.19}$$

in terms of the symmetric matrices $\{A, B\}$,

$$\boxed{\begin{array}{rcl} A & = & (P \otimes I_M) \ + \ (I_M \otimes P) \\ B & = & E\left(\dfrac{\mathbf{u}_i^T\mathbf{u}_i \otimes \mathbf{u}_i^T\mathbf{u}_i}{g^2[\mathbf{u}_i]}\right) \\ P & = & \mathrm{E}\left(\dfrac{\mathbf{u}_i^T\mathbf{u}_i}{g[\mathbf{u}_i]}\right) \end{array}} \tag{1.20}$$

Actually, $A$ is positive-definite (because $P$ is) and $B$ is nonnegative-definite. Using the column notation $\sigma$, and the relation $\sigma' = F\sigma$, we can write (1.16)–(1.17) as

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

which we shall rewrite more succinctly, by dropping the $\mathrm{vec}^{-1}(\cdot)$ notation and keeping the weighting vectors, as

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\sigma = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{F\sigma} \ + \ \mu^2\sigma_v^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_\sigma}{g^2[\mathbf{u}_i]}\right)} \tag{1.21}$$

Now, as mentioned earlier, in transient analysis we are interested in the evolution of $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2$ and $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u}$; the former quantity is the filter mean-square deviation while the second quantity relates to the filter mean-square error (or learning) curve since

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

The quantities $\{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2, \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u}\}$ are in turn special cases of $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\Sigma$ obtained by choosing $\Sigma = I$ or $\Sigma = R_u$. Therefore, in the sequel, we focus on studying the evolution of $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\Sigma$ for arbitrary $\Sigma$.

From (1.21) we see that in order to evaluate $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\sigma$, we need $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{F\sigma}$ with weighting vector $F\sigma$. This term can be deduced from (1.21) by writing it for $\sigma \leftarrow F\sigma$, i.e.,

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

with the weighted term $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{F^2\sigma}$. This term can in turn be deduced from (1.21) by writing it for $\sigma \leftarrow F^2\sigma$. Continuing in this fashion, for successive powers of $F$, we arrive at

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

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

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

$$p(x) = x^{M^2} + p_{M^2-1}x^{M^2-1} + p_{M^2-2}x^{M^2-2} + \ldots + p_1 x + p_0$$

with coefficients $\{p_i\}$. Then, since $p(F) = 0$ in view of the Cayley-Hamilton theorem [15], we have

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

Putting these results together, we conclude that the transient (mean-square) behavior of the filter (1.3) is described by an $M^2-$dimensional state-space model of the form:

$$\boxed{\mathcal{W}_i = \mathcal{F}\mathcal{W}_{i-1} + \mu^2\sigma_v^2\mathcal{Y}} \tag{1.22}$$

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

$$\mathcal{W}_i = \begin{bmatrix} \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{\sigma} \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{F\sigma} \\ \vdots \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{F^{M^2-2}\sigma} \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{F^{M^2-1}\sigma} \end{bmatrix}, \qquad \mathcal{Y} = \begin{bmatrix} \mathrm{E}\,(\|\mathbf{u}_i\|^2_{\sigma}/g^2[\mathbf{u}_i]) \\ \mathrm{E}\,(\|\mathbf{u}_i\|^2_{F\sigma}/g^2[\mathbf{u}_i]) \\ \vdots \\ \mathrm{E}\,\left(\|\mathbf{u}_i\|^2_{F^{M^2-2}\sigma}/g^2[\mathbf{u}_i]\right) \\ \mathrm{E}\,\left(\|\mathbf{u}_i\|^2_{F^{M^2-1}\sigma}/g^2[\mathbf{u}_i]\right) \end{bmatrix} \tag{1.23}$$

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

$$\mathcal{F} = \begin{bmatrix} 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ \vdots & & & & \\ 0 & 0 & 0 & & 1 \\ -p_0 & -p_1 & -p_2 & \ldots & -p_{M^2-1} \end{bmatrix} \tag{1.24}$$

The entries of $\mathcal{Y}$ can be written more compactly as

$$\mathcal{Y} = \mathrm{col}\left\{\mathrm{Tr}(Q\mathrm{vec}^{-1}(F^k\sigma)), \quad k = 0, 1, \ldots, M^2-1\right\}$$

where

$$\boxed{Q = \mathrm{E}\left(\frac{\mathbf{u}_i^T\mathbf{u}_i}{g^2[\mathbf{u}_i]}\right)} \tag{1.25}$$

and the notation $\mathrm{vec}^{-1}(F^k\sigma)$ recovers the weighting matrix that corresponds to the vector $F^k\sigma$.

When $\Sigma = I$, the evolution of the top entry of $\mathcal{W}_i$ in (1.22) describes the mean-square deviation of the filter, i.e., $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2$. If, on the other hand, $\Sigma$ is chosen as $\Sigma = R_u$, the evolution of the top entry of $\mathcal{W}_i$ describes the excess mean-square error (or learning curve) of the filter, i.e., $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u} = \mathrm{E}\,\mathbf{e}_a^2(i)$.

The learning curve can also be characterized more explicitly as follows. Let $r = \mathrm{vec}(R_u)$ and choose $\sigma = r$. Iterating (1.21) we find that

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

that is,

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

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

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

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

$$\boxed{\mathrm{E}\,\mathbf{e}_a^2(i) = \mathrm{E}\,\mathbf{e}_a^2(i-1) \;+\; \|w^o\|^2_{F^{i-1}(F-I)r} \;+\; \mu^2\sigma_v^2\mathrm{Tr}(Q\mathrm{vec}^{-1}(F^{i+1}r))} \tag{1.26}$$

which describes the learning curve of data-normalized adaptive filters as in (1.3). Further discussions on the learning behavior of adaptive filters can be found in [16].

## 1.6 Mean-Square Stability

Recursion (1.22) 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 lie inside the unit circle. But since $\mathcal{F}$ has the form of a companion matrix, its eigenvalues coincide with the roots of $p(x)$, which in turn coincide with the eigenvalues of $F$. Therefore, the mean-square stability of the adaptive filter requires the matrix $F$ in (1.19) to be a stable matrix.

Now it can be verified that matrices $F$ of the form (1.19), for arbitrary $\{A > 0, B \geq 0\}$, are stable for all values of $\mu$ in the range:

$$0 < \mu < \min\left\{\frac{1}{\lambda_{\max}(A^{-1}B)}, \frac{1}{\max\left\{\lambda(H) \in \mathbb{R}^+\right\}}\right\} \tag{1.27}$$

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

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

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 $\mu$ is determined by $1/\lambda_{\max}(A^{-1}B)$ alone.[3]

Likewise, the mean-stability of the filter, as dictated by (1.18), requires the eigenvalues of $(I - \mu P)$ to lie inside the unit circle or, equivalently,

$$\mu < 2/\lambda_{\max}(P) \tag{1.28}$$

Combining (1.27) and (1.28) we conclude that the filter is stable in the mean and mean-square senses for step-sizes in the range

$$\boxed{\mu < \min\left\{\frac{2}{\lambda_{\max}(P)}, \frac{1}{\lambda_{\max}(A^{-1}B)}, \frac{1}{\max\left\{\lambda(H) \in \mathbb{R}^+\right\}}\right\}} \tag{1.29}$$

## 1.7 Steady-State Performance

Steady-state performance results can also be deduced from (1.21). Assuming the filter is operating in steady-state, recursion (1.21) gives in the limit

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

This expression allows us to evaluate the steady-state value of $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_S$ for any weighting matrix $S$, by choosing $\sigma$ such that

$$(I - F)\sigma = \mathrm{vec}(S)$$

i.e.,

$$\sigma = (I - F)^{-1}\mathrm{vec}(S)$$

In particular, the filter excess mean-square error, defined by

$$\mathrm{EMSE} = \lim_{i\to\infty} \mathrm{E}\,\mathbf{e}_a^2(i)$$

corresponds to the choice $S = R_u$ since, by virtue of the independence assumption (1.15), $\mathrm{E}\,\mathbf{e}_a^2(i) = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u}$. In other words, we should select $\sigma$ as

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

[3]The condition involving $\lambda_{\max}(A^{-1}B)$ in (1.27) guarantees that all eigenvalues of $F$ are less than one, while the condition involving $H$ ensures that all eigenvalues of $F$ are larger than $-1$.

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

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

is obtained by setting $S = I$, i.e.,

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

Let $\{\Sigma_{\text{emse}}, \Sigma_{\text{msd}}\}$ denote the weighting matrices that correspond to the vectors $\{\sigma_{\text{emse}}, \sigma_{\text{msd}}\}$, i.e.,

$$\Sigma_{\text{emse}} = \text{vec}^{-1}(\sigma_{\text{emse}}), \qquad \Sigma_{\text{msd}} = \text{vec}^{-1}(\sigma_{\text{msd}})$$

Then we are led to the following expressions for the filter performance:

$$\boxed{\begin{array}{l} \text{EMSE} = \mu^2\sigma_v^2\text{Tr}(Q\Sigma_{\text{emse}}) \\ \\ \text{MSD} = \mu^2\sigma_v^2\text{Tr}(Q\Sigma_{\text{msd}}) \end{array}} \tag{1.30}$$

Alternatively, we can also write

$$\boxed{\begin{array}{lcl} \text{EMSE} = \mu^2\sigma_v^2\text{vec}^T(Q)\sigma_{\text{emse}} & = & \mu^2\sigma_v^2\text{vec}^T(Q)(I-F)^{-1}\text{vec}(R_u) \\ \\ \text{MSD} = \mu^2\sigma_v^2\text{vec}^T(Q)\sigma_{\text{msd}} & = & \mu^2\sigma_v^2\text{vec}^T(Q)(I-F)^{-1}\text{vec}(I) \end{array}} \tag{1.31}$$

While these steady-state results are obtained here as a consequence of the variance relation (1.21), which relies on the independence assumption (1.15), it turns out that steady-state results can also be deduced in an alternative manner that does not rely on using the independence condition. This alternative derivation starts from (1.10) and uses the fact that $\text{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \text{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2$ in steady-state to derive expressions for the filter EMSE; the details are spelled out in [10, 11].

## 1.8 Small Step-Size Approximation

Returning to the expression of $F$ in (1.19), and to the performance results (1.30), we see that they are defined in terms of moment matrices $\{A, B, P, Q\}$. These moments are generally not easy to evaluate for arbitrary input distributions and data nonlinearities $g$. This fact explains why it is common in the literature to resort to Gaussian or whiteness assumptions on the regression data.

In our development so far, all results concerning filter transient performance, stability, and steady-state performance (e.g., (1.22), (1.26), (1.29), and (1.30)) have been derived without restricting the distribution of the regression data to being Gaussian or white. In order to simplify the analysis, we shall keep the input distribution generic and appeal instead to approximations pertaining to the step-size value, to the filter length, and also to a fourth-order moment approximation. In this section, we discuss the small-step size approximation.

To begin with, even though we may not have available explicit values for the moments $\{A, B, P, Q\}$ in general, we can still assert the following. If the distribution of the regression data is such that the matrix $B$ is finite, then there always exists a small enough step-size for which $F$ (and, hence, the filter) is stable. To see this, observe first that the eigenvalues of $I - \mu A$ are given by

$$\{1 - \mu[\lambda_k(P) + \lambda_j(P)]\}$$

for all combinations $1 \leq j, k \leq M$ of the eigenvalues of $P$. Now if $B$ is bounded, then the maximum eigenvalue of $F$ is bounded by

$$\lambda_{\max}(F) \ \leq \ 1 - 2\mu\lambda_{\min}(P) + \mu^2\beta$$

for some finite positive scalar $\beta$ (e.g., $\beta = \lambda_{\max}(B)$). The upper bound on $\lambda_{\max}(F)$ is a quadratic function of $\mu$, and it is easy to verify that the values of this function are less than one for step-sizes in the range $(0, 2\lambda_{\min}(P)/\beta)$. Since $\lambda_{\min}(P)/\beta$ is positive, we conclude that there should exist a small enough $\mu$ such that $F$ is stable and, consequently, the filter is mean-square stable.

Now for such small step-sizes, we may ignore the quadratic term in $\mu$ that appears in (1.17), and approximate the variance relation (1.16)–(1.17) by

$$\boxed{\begin{array}{l} \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\Sigma = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma'} \ + \ \mu^2\sigma_v^2\mathrm{E}\left(\dfrac{\|\mathbf{u}_i\|^2_\Sigma}{g^2[\mathbf{u}_i]}\right) \\ \Sigma' = \Sigma \ - \ \mu\Sigma P \ - \ \mu P\Sigma \end{array}} \tag{1.32}$$

or, equivalently, using the weighting vector notation, by

$$\begin{aligned} \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\sigma \ &= \ \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{F\sigma} \ + \ \mu^2\sigma_v^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_\sigma}{g^2[\mathbf{u}_i]}\right) \\ F \ &= \ I - \mu A \end{aligned}$$

where $P = \mathrm{E}\left(\mathbf{u}_i^T\mathbf{u}_i/g[\mathbf{u}_i]\right)$. Moreover, since $I - F = \mu A$, we can also approximate the EMSE and MSD performances (1.30) of the filter by

$$\boxed{\begin{array}{l} \text{EMSE} \approx \mu\sigma_v^2\mathrm{Tr}(Q\Sigma_{\text{emse}}) \\[1em] \text{MSD} \approx \mu\sigma_v^2\mathrm{Tr}(Q\Sigma_{\text{msd}}) \end{array}} \tag{1.33}$$

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

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

That is, $\{\Sigma_{\text{emse}}, \Sigma_{\text{msd}}\}$ are the unique solutions of the Lyapunov equations

$$P\Sigma_{\text{msd}} + \Sigma_{\text{msd}}P = I \quad \text{and} \quad P\Sigma_{\text{emse}} + \Sigma_{\text{emse}}P = R_u$$

It is easy to verify that $\Sigma_{\text{msd}} = P^{-1}/2$ so that the MSD expression can be written more explicitly as

$$\boxed{\text{MSD} \approx \frac{\mu\sigma_v^2}{2}\text{Tr}(QP^{-1})} \tag{1.34}$$

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

$$\text{EMSE} \approx \frac{\mu\sigma_v^2\text{Tr}(R_u)}{2}, \qquad \text{MSD} \approx \frac{\mu\sigma_v^2 M}{2} \quad \text{(LMS)} \tag{1.35}$$

Using the simplified variance relation (1.32), we can also 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 (1.22). To see this, let $P = U\Delta U^T$ denote the eigen-decomposition of $P > 0$, and introduce the transformed quantities:

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

Then the variance relation (1.32) can be equivalently rewritten as[4]

$$\begin{aligned} &\text{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{\Sigma}} = \text{E}\,\|\overline{\mathbf{w}}_{i-1}\|^2_{\overline{\Sigma}'} \;+\; \mu^2\sigma_v^2\text{E}\left(\frac{\|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}}}{g^2[\overline{\mathbf{u}}_i]}\right) \\ &\overline{\Sigma}' = \overline{\Sigma} \;-\; \mu\overline{\Sigma}\Delta \;-\; \mu\Delta\overline{\Sigma} \end{aligned} \tag{1.36}$$

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

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

Then from (1.36) we find that

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

where $\overline{F}$ is the $M \times M$ matrix

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

Repeating the arguments that led to (1.22) we can then establish that, for sufficiently small step-sizes, the evolution of $\text{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{\sigma}}$ is described by the following $M-$dimensional state-space model:

$$\boxed{\overline{\mathcal{W}}_i = \overline{\mathcal{F}}\,\overline{\mathcal{W}}_{i-1} + \mu^2\sigma_v^2\overline{\mathcal{Y}}} \tag{1.37}$$

[4]Usually, $g[\cdot]$ is invariant under orthogonal transformations, i.e., $g[\mathbf{u}_i] = g[\overline{\mathbf{u}}_i]$. This is the case for LMS, NLMS, and $\epsilon-$NLMS.

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

$$\overline{\mathcal{W}}_i = \begin{bmatrix} \mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{\sigma}} \\ \mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{F}\overline{\sigma}} \\ \vdots \\ \mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{F}^{M-2}\overline{\sigma}} \\ \mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{F}^{M-1}\overline{\sigma}} \end{bmatrix}, \quad \overline{\mathcal{Y}} = \begin{bmatrix} \mathrm{E}\left(\|\overline{\mathbf{u}}_i\|^2_{\overline{\sigma}}/g^2[\overline{\mathbf{u}}_i]\right) \\ \mathrm{E}\left(\|\overline{\mathbf{u}}_i\|^2_{\overline{F}\overline{\sigma}}/g^2[\overline{\mathbf{u}}_i]\right) \\ \vdots \\ \mathrm{E}\left(\|\overline{\mathbf{u}}_i\|^2_{\overline{F}^{M-2}\overline{\sigma}}/g^2[\overline{\mathbf{u}}_i]\right) \\ \mathrm{E}\left(\|\overline{\mathbf{u}}_i\|^2_{\overline{F}^{M-1}\overline{\sigma}}/g^2[\overline{\mathbf{u}}_i]\right) \end{bmatrix} \tag{1.38}$$

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

$$\overline{\mathcal{F}} = \begin{bmatrix} 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ \vdots & & & & \\ 0 & 0 & 0 & & 1 \\ -\overline{p}_0 & -\overline{p}_1 & -\overline{p}_2 & \ldots & -\overline{p}_{M-1} \end{bmatrix} \tag{1.39}$$

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

$$\|\overline{\mathbf{w}}_i\|^2_{\overline{\sigma}} = \|\overline{\mathbf{w}}_i\|^2 = \|U^T\tilde{\mathbf{w}}_i\|^2 = \|\tilde{\mathbf{w}}_i\|^2$$

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

When $P$ and $R_u$ have identical eigenvectors, e.g., as in LMS for which $g[\mathbf{u}] = 1$ and $P = R_u$, then the evolution of the learning curve of the filter can also be read from (1.37). To see this, let $\lambda$ be the column vector consisting of the eigenvalues of $R_u$. Choosing $\overline{\sigma} = \lambda$ gives

$$\|\overline{\mathbf{w}}_i\|^2_{\overline{\sigma}} = \|\overline{\mathbf{w}}_i\|^2_{\lambda} = \overline{\mathbf{w}}_i^T\Lambda\overline{\mathbf{w}}_i = \tilde{\mathbf{w}}_i^T R_u\tilde{\mathbf{w}}_i = \|\tilde{\mathbf{w}}_i\|^2_{R_u}$$

so that the EMSE behavior of the filter can be read from the top entry of the resulting state-vector $\overline{\mathcal{W}}_i$.

## 1.9 Applications to Selected Filters

We now illustrate the application of the results of the earlier sections, as well as some extensions of these results, to selected adaptive filters.

### 1.9.1 The NLMS Algorithm

Our first example derives performance results for NLMS by showing how to relate it to LMS. In NLMS, $g[\mathbf{u}] = \|\mathbf{u}\|^2$, and the filter recursion takes the form

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu\frac{\mathbf{u}_i^T}{\|\mathbf{u}_i\|^2}[\mathbf{d}(i) - \mathbf{u}_i\mathbf{w}_{i-1}]$$

Introduce the transformed variables:

$$\boxed{\check{\mathbf{u}}_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}, \quad \check{\mathbf{d}}(i) = \frac{\mathbf{d}(i)}{\|\mathbf{u}_i\|}, \quad \check{\mathbf{v}}(i) = \frac{\mathbf{v}(i)}{\|\mathbf{u}_i\|}} \tag{1.40}$$

Then the NLMS recursion can be rewritten as

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

with

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

In other words, we find that NLMS can be regarded as an LMS filter with respect to the variables $\{\check{\mathbf{d}}(i), \check{\mathbf{u}}_i\}$. Moreover, these variables satisfy a model similar to that of $\{\mathbf{d}(i), \mathbf{u}_i\}$, as given by (1.1)–(1.2), viz.,

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

where

(a) The sequence $\{\check{\mathbf{v}}(i)\}$ is iid with variance $\mathrm{E}\,\check{\mathbf{v}}^2(i) = \check{\sigma}_v^2 = \sigma_v^2 \mathrm{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)$.

(b) The sequence $\mathbf{v}(i)$ is independent of $\mathbf{u}_j$ for all $i \neq j$.

(c) The covariance matrix of $\check{\mathbf{u}}_i$ is $\check{R}_u = \mathrm{E}\,\check{\mathbf{u}}_i^T \check{\mathbf{u}}_i = \mathrm{E}\left(\frac{\mathbf{u}_i^T \mathbf{u}_i}{\|\mathbf{u}_i\|^2}\right) > 0$.

(d) The random variables $\{\check{\mathbf{v}}(i), \check{\mathbf{u}}_i\}$ are zero mean.

These conditions allow us to repeat the previous derivation of the variance and mean relations (1.16)–(1.18) using the transformed variables (1.40). In this way, the performance of NLMS can be deduced from that of LMS. In particular, from (1.35) we get for NLMS:

$$\boxed{\text{MSD} \approx \frac{\mu \check{\sigma}_v^2 M}{2} \ = \ \frac{\mu \sigma_v^2 M}{2} \mathrm{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)} \tag{1.41}$$

and

$$\lim_{i \to \infty} \mathrm{E}\,\check{\mathbf{e}}_a^2(i) \approx \frac{\mu \check{\sigma}_v^2 \mathrm{Tr}(\check{R}_u)}{2} \ = \ \frac{\mu \check{\sigma}_v^2}{2}$$

since $\mathrm{Tr}(\check{R}_u) = 1$, and where $\check{\mathbf{e}}_a(i) = \check{\mathbf{d}}(i) - \check{\mathbf{u}}_i \mathbf{w}_{i-1}$. However, the filter EMSE relates to the limiting value of $\mathrm{E}\,\mathbf{e}_a^2(i)$ and not $\mathrm{E}\,\check{\mathbf{e}}_a^2(i)$. To find this limiting value, we first note from the definitions of $\mathbf{e}_a(i)$ and $\check{\mathbf{e}}_a(i)$ that

$$\frac{1}{\|\mathbf{u}_i\|^2} \cdot \mathbf{e}_a^2(i) \ = \ \check{\mathbf{e}}_a^2(i)$$

Then if we introduce the steady-state separation assumption[5]

$$\mathrm{E}\left(\frac{1}{\|\mathbf{u}_i\|^2} \cdot \mathbf{e}_a^2(i)\right) \approx \frac{\mathrm{E}\,\mathbf{e}_a^2(i)}{\mathrm{E}\,\|\mathbf{u}_i\|^2} \quad \text{as } i \longrightarrow \infty$$

[5] The assumption is reasonable for longer filters.

so that

$$\lim_{i\to\infty} \mathrm{E}\,\mathbf{e}_a^2(i) \;=\; \mathrm{Tr}(R_u)\cdot\left(\lim_{i\to\infty} \mathrm{E}\,\check{\mathbf{e}}_a^2(i)\right)$$

we get

$$\boxed{\mathrm{EMSE} \;=\; \frac{\mu\sigma_v^2 \mathrm{Tr}(R_u)}{2}\mathrm{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)} \tag{1.42}$$

An alternative method to evaluate the steady-state (as well as transient) performance of NLMS is to treat it as a special case of the results developed in Sec. 1.8 by setting $g(\mathbf{u}) = \|\mathbf{u}\|^2$. In this case, the variance relation (1.36) would become

$$\begin{cases} \mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{\Sigma}} & = \;\; \mathrm{E}\,\|\overline{\mathbf{w}}_{i-1}\|^2_{\overline{\Sigma}'} \;+\; \mu^2\sigma_v^2 \mathrm{E}\left[\dfrac{\|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}}}{\|\overline{\mathbf{u}}_i\|^4}\right] \\ \quad\overline{\Sigma}' & = \;\; \overline{\Sigma} - \mu\overline{\Sigma}\Delta - \mu\Delta\overline{\Sigma} \end{cases}$$

Moreover, the EMSE and MSD expressions (1.33) and (1.34) would give

$$\boxed{\begin{array}{l} \mathrm{MSD} = \dfrac{\mu\sigma_v^2 \mathrm{Tr}(QP^{-1})}{2} \\[2ex] \mathrm{EMSE} = \mu\sigma_v^2 \mathrm{Tr}(Q\Sigma_{\mathrm{emse}}) \end{array}} \tag{1.43}$$

where now

$$P = \mathrm{E}\left(\frac{\mathbf{u}_i^T\mathbf{u}_i}{\|\mathbf{u}_i\|^2}\right), \qquad Q = \mathrm{E}\left(\frac{\mathbf{u}_i^T\mathbf{u}_i}{\|\mathbf{u}_i\|^4}\right)$$

and $\Sigma_{\mathrm{emse}}$ is the unique solution of $P\Sigma_{\mathrm{emse}}+\Sigma_{\mathrm{emse}}P = R_u$. Expressions (1.43) are alternatives to (1.41) and (1.42).

### 1.9.2 The RLS Algorithm

Our second example pertains to the recursive least-squares algorithm:

$$\mathbf{w}_i \;=\; \mathbf{w}_{i-1} + \mathbf{P}_i\, \mathbf{u}_i^T[\mathbf{d}(i) - \mathbf{u}_i\mathbf{w}_{i-1}], \quad i \geq 0 \tag{1.44}$$

$$\mathbf{P}_i \;=\; \alpha^{-1}\left[\mathbf{P}_{i-1} - \frac{\alpha^{-1}\mathbf{P}_{i-1}\mathbf{u}_i^T\mathbf{u}_i\mathbf{P}_{i-1}}{1+\alpha^{-1}\mathbf{u}_i\mathbf{P}_{i-1}\mathbf{u}_i^T}\right] \tag{1.45}$$

where the data $\{\mathbf{d}(i), \mathbf{u}_i\}$ satisfy (1.1)–(1.2), and the regressors satisfy the independence assumption (1.15). In the above, $0 \ll \alpha \leq 1$ is a forgetting factor and $\mathbf{P}_{-1} = \epsilon^{-1}I$ for a small positive $\epsilon$.

Compared with the LMS-type recursion (1.3), 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-conservation approach of Secs. 1.3–1.4 can be extended to deal with this more general case. In particular, it is straightforward to verify that (1.10) is now replaced by

$$\|\mathbf{u}_i\|^2_{\mathbf{P}_i\Sigma\mathbf{P}_i}\cdot\|\tilde{\mathbf{w}}_i\|^2_{\Sigma}+(\mathbf{e}_a^{\mathbf{P}_i\Sigma}(i))^2=\|\mathbf{u}_i\|^2_{\mathbf{P}_i\Sigma\mathbf{P}_i}\cdot\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma}+(\mathbf{e}_p^{\mathbf{P}_i\Sigma}(i))^2 \tag{1.46}$$

Under expectation, (1.46) leads to

$$\begin{aligned}\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{\Sigma} &= \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma'} + \sigma_v^2\mathrm{E}\,\|\mathbf{u}_i\|^2_{\mathbf{P}_i\Sigma\mathbf{P}_i}\\ \Sigma' &= \Sigma - \Sigma\mathrm{E}\,(\mathbf{P}_i\mathbf{u}_i^T\mathbf{u}_i) - \mathrm{E}\,(\mathbf{u}_i^T\mathbf{u}_i\mathbf{P}_i)\Sigma + \mathrm{E}\,\left[\|\mathbf{u}_i\|^2_{\mathbf{P}_i\Sigma\mathbf{P}_i}\mathbf{u}_i^T\mathbf{u}_i\right]\end{aligned} \tag{1.47}$$

However, the presence of the matrix $\mathbf{P}_i$ makes the subsequent analysis rather challenging; this is because $\mathbf{P}_i$ is dependent not only on $\mathbf{u}_i$ but also on all prior regressors $\{\mathbf{u}_j, j \le i\}$.

In order to make the analysis more tractable, whenever necessary, we shall approximate and replace the random variable $\mathbf{P}_i$ in steady-state by its respective mean value.[6] Now since

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

we find that, as $i \to \infty$, and since $\alpha < 1$,

$$\lim_{i\to\infty}\mathrm{E}\left(\mathbf{P}_i^{-1}\right) = \frac{R_u}{1-\alpha} \triangleq P^{-1}$$

That is, the mean value of $\mathbf{P}_i^{-1}$ tends to $R_u/(1-\alpha)$. In comparison, the evaluation of the limiting mean value of $\mathbf{P}_i$ is generally harder. For this reason, we shall content ourselves with the approximation

$$\mathrm{E}\mathbf{P}_i \approx \left[\mathrm{E}\mathbf{P}_i^{-1}\right]^{-1} = (1-\alpha)R_u^{-1} = P, \quad \text{as} \quad i\to\infty$$

This is an approximation, of course, because even though $\mathbf{P}_i$ and $\mathbf{P}_i^{-1}$ are the inverses of one another, it does not hold that their means will have the same inverse relation.[7]

Replacing $\mathbf{P}_i$ by $P = (1-\alpha)R_u^{-1}$, we find that the variance relation (1.47) becomes

$$\boxed{\begin{aligned}&\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{\Sigma} = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma'} + \sigma_v^2(1-\alpha)^2\mathrm{E}\,\|\mathbf{u}_i\|^2_{R_u^{-1}\Sigma R_u^{-1}}\\ &\Sigma' = \Sigma - 2(1-\alpha)\Sigma + (1-\alpha)^2\mathrm{E}\left[\|\mathbf{u}_i\|^2_{R_u^{-1}\Sigma R_u^{-1}}\mathbf{u}_i^T\mathbf{u}_i\right]\end{aligned}}$$

Introduce the eigen-decomposition $R_u = U\Lambda U^T$, and define the transformed variables

$$\overline{\mathbf{w}}_i \triangleq U^T\tilde{\mathbf{w}}_i, \qquad \overline{\mathbf{u}}_i \triangleq \mathbf{u}_iU, \qquad \overline{\Sigma} \triangleq U^T\Sigma U$$

[6] This approximation essentially amounts to an ergodicity assumption on the regressors.

[7] It turns out that the approximation is reasonable for Gaussian regressors.

Assume further, for the sake of illustration, that the regressors $\{\mathbf{u}_i\}$ are Gaussian. Then

$$\mathrm{E}\left[\|\overline{\mathbf{u}}_i\|^2_{\Lambda^{-1}\overline{\Sigma}\Lambda^{-1}}\overline{\mathbf{u}}_i^T\overline{\mathbf{u}}_i\right] \;=\; 2\Lambda\mathrm{Tr}(\Lambda^{-1}\overline{\Sigma}) \;+\; \overline{\Sigma}$$

and the variance relation becomes

$$\boxed{\begin{array}{l} \mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{\Sigma}} = \mathrm{E}\,\|\overline{\mathbf{w}}_{i-1}\|^2_{\overline{\Sigma}'} \;+\; \sigma_v^2(1-\alpha)^2\mathrm{E}\,\|\overline{\mathbf{u}}_i\|^2_{\Lambda^{-1}\overline{\Sigma}\Lambda^{-1}} \\ \\ \overline{\Sigma}' = \alpha^2\overline{\Sigma} \;+\; 2(1-\alpha)^2\Lambda\mathrm{Tr}(\Lambda^{-1}\overline{\Sigma}) \end{array}}$$

It follows that $\overline{\Sigma}'$ will be diagonal if $\overline{\Sigma}$ is. If we further introduce the $M-$dimensional column vectors

$$\lambda = \mathrm{diag}\{\Lambda\}, \qquad a = \mathrm{diag}\{\Lambda^{-1}\}, \qquad \overline{\sigma} = \mathrm{diag}\{\overline{\Sigma}\}$$

then the above recursion for $\overline{\Sigma}'$ is equivalent to

$$\overline{\sigma}' = \overline{F}\overline{\sigma} \quad \text{where} \quad \overline{F} = \alpha^2 I + 2(1-\alpha)^2\lambda a^T$$

Let $\overline{\Sigma}_{\mathrm{msd}}$ denote the weighting matrix that corresponds to the vector

$$\overline{\sigma}_{\mathrm{msd}} \;=\; (I-\overline{F})^{-1}\mathrm{diag}(I)$$

Let also $\overline{\Sigma}_{\mathrm{emse}}$ denote the weighting matrix that corresponds to the vector

$$\overline{\sigma}_{\mathrm{emse}} \;=\; (I-\overline{F})^{-1}\lambda$$

Then since

$$\begin{aligned} \mathrm{MSD} &= \sigma_v^2(1-\alpha)^2\mathrm{E}\,\|\overline{\mathbf{u}}_i\|^2_{\Lambda^{-1}\overline{\Sigma}_{\mathrm{msd}}\Lambda^{-1}} \\ \mathrm{EMSE} &= \sigma_v^2(1-\alpha)^2\mathrm{E}\,\|\overline{\mathbf{u}}_i\|^2_{\Lambda^{-1}\overline{\Sigma}_{\mathrm{emse}}\Lambda^{-1}} \end{aligned}$$

we can verify after some algebra that

$$\boxed{\begin{array}{l} \mathrm{MSD} = \dfrac{\sigma_v^2\sum_{k=1}^{M}(1/\lambda_k)}{\frac{1+\alpha}{1-\alpha} - 2M} \\ \\ \mathrm{EMSE} = \dfrac{\sigma_v^2 M}{\frac{1+\alpha}{1-\alpha} - 2M} \end{array}} \qquad (1.48)$$

### 1.9.3 Leaky-LMS

Our third example extends the energy-conservation and variance relations of Secs. 1.3 and 1.4 to leaky-LMS updates of the form:

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

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

where $\alpha$ is a positive scalar. The data $\{\mathbf{d}(i), \mathbf{u}_i\}$ are still assumed to satisfy (1.1)–(1.2), with the regressors satisfying the independence assumption (1.15).

Repeating the arguments of Secs. 1.3–1.4, it is straightforward to verify that the variance and mean relations (1.16)–(1.18) extend to the following (see [17]):

$$\boxed{\begin{array}{rcl} \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\Sigma & = & \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma'} \;+\; \mu^2\sigma_v^2 \mathrm{E}\,\|\mathbf{u}_i\|^2_\Sigma \;+\; 2\alpha\mu\,(w^o)^T\,\Sigma J \mathrm{E}\,\tilde{\mathbf{w}}_{i-1} \;+\; \alpha^2\mu^2\|w^o\|^2_\Sigma \\ \Sigma' & = & \Sigma - \mu\big(\mathrm{E}\,\mathbf{U}_i\big)\Sigma - \mu\Sigma\big(\mathrm{E}\,\mathbf{U}_i\big) \;+\; \mu^2 \mathrm{E}\,\big(\mathbf{U}_i \Sigma \mathbf{U}_i\big) \\ \mathrm{E}\,\tilde{\mathbf{w}}_i & = & J\mathrm{E}\,\tilde{\mathbf{w}}_{i-1} \;+\; \alpha\mu w^o \end{array}}$$

where

$$\mathbf{U}_i = \alpha I + \mathbf{u}_i^T \mathbf{u}_i, \qquad J = \mathrm{E}\,(I - \mu \mathbf{U}_i) \;=\; (1 - \alpha\mu)I - \mu R_u$$

Frequently $w_{-1} = 0$, so that $\mathrm{E}\,\tilde{w}_{-1} = w^o$. We will make this assumption to simplify the analysis, although it is not necessary for stability or steady-state results.

Therefore, by iterating the recursion for $\mathrm{E}\,\tilde{\mathbf{w}}_i$ we can verify that

$$\mathrm{E}\,\tilde{\mathbf{w}}_{i-1} \;=\; C_i w^o, \quad i \geq 0$$

where

$$C_i = J^i + \alpha\mu(I + J + \ldots + J^{i-1})$$

It then follows that the term below, which appears in the recursion for $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\Sigma$, can be expressed in terms of $\|w^o\|^2$ as

$$2\alpha\mu\,(w^o)^T\,\Sigma J \mathrm{E}\,\tilde{\mathbf{w}}_{i-1} \;=\; \alpha\mu\|w^o\|^2_{\Sigma J C_i + C_i J \Sigma}$$

Now repeating the arguments of Sec. 1.5 we can verify that the transient behavior of the leaky filter is characterized by the following state-space model:

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

where $\mathcal{W}_i$ is the $M^2-$dimensional vector

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

and $\mathcal{F}$ is the $M^2 \times M^2$ companion matrix

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

with

$$p(x) \triangleq \det(xI - F) = x^{M^2} + \sum_{k=0}^{M^2-1} p_k x^k$$

denoting the characteristic polynomial of the matrix

$$F = I - \mu A + \mu^2 B$$

where

$$\begin{aligned} A &= (\mathrm{E}\,\mathbf{U}_i \otimes I) + (I \otimes \mathrm{E}\,\mathbf{U}_i) \\ B &= \mathrm{E}\,(\mathbf{U}_i \otimes \mathbf{U}_i) \end{aligned}$$

Moreover,

$$\mathcal{Y}_i = \mu\sigma_v^2 \begin{bmatrix} \mathrm{E}\,|\mathbf{u}_i\|_\sigma^2 \\ \mathrm{E}\,\|\mathbf{u}_i\|_{F\sigma}^2 \\ \mathrm{E}\,\|\mathbf{u}_i\|_{F^2\sigma}^2 \\ \vdots \\ \mathrm{E}\,\|\mathbf{u}_i\|_{F^{M^2-1}\sigma}^2 \end{bmatrix} + \alpha \begin{bmatrix} \|w^o\|_{(\alpha\mu I + S_i)\sigma}^2 \\ \|w^o\|_{(\alpha\mu I + S_i)F\sigma}^2 \\ \|w^o\|_{(\alpha\mu I + S_i)F^2\sigma}^2 \\ \vdots \\ \|w^o\|_{(\alpha\mu I + S_i)F^{M^2-1}\sigma}^2 \end{bmatrix}$$

where $S_i$ is the $M^2 \times M^2$ matrix

$$S_i \triangleq (JC_i \otimes I_M) + (I_M \otimes C_i J)$$

It follows that the filter is stable in the mean and mean-square senses for step-sizes in the range

$$\mu < \min\left\{\frac{2}{\alpha + \lambda_{\max}(R_u)}, \frac{1}{\lambda_{\max}(A^{-1}B)}, \frac{1}{\max\left\{\lambda(H) \in \mathbb{R}^+\right\}}\right\}$$

where

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

It also follows that in steady-state,

$$\boxed{\begin{array}{rcl} \lim_{i\to\infty} \mathrm{E}\,\tilde{\mathbf{w}}_i & = & \alpha(\alpha I + R_u)^{-1} w^o \\ \text{MSD} & = & \mu^2\sigma_v^2 \mathrm{E}\left(\|\mathbf{u}_i\|^2_{(I-F)^{-1}\mathrm{vec}(I)}\right) \;+\; \alpha^2\mu^2 \|w^o\|^2_{T(I-F)^{-1}\mathrm{vec}(I)} \\ \text{EMSE} & = & \mu^2\sigma_v^2 \mathrm{E}\left(\|\mathbf{u}_i\|^2_{(I-F)^{-1}\mathrm{vec}(R_u)}\right) \;+\; \alpha^2\mu^2 \|w^o\|^2_{T(I-F)^{-1}\mathrm{vec}(R_u)} \end{array}} \tag{1.49}$$

where $T$ is the $M^2 \times M^2$ matrix

$$T = I \;+\; \left((I-J)^{-1}J \otimes I\right) \;+\; \left(I \otimes (I-J)^{-1}J\right)$$

# 1.10 Fourth-Order Moment Approximation

Instead of the small-step size approximation of Sec. 1.8, we can choose to approximate the fourth-order moment that appears in the expression for $\Sigma'$ in (1.17) as

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

where $P = \mathrm{E}\left(\mathbf{u}_i^T\mathbf{u}_i/g[\mathbf{u}_i]\right)$. In this way, expression (1.17) for $\Sigma'$ would become

$$\boxed{\Sigma' = \Sigma \;-\; \mu\Sigma P \;-\; \mu P\Sigma \;+\; \mu^2 P\mathrm{Tr}(P\Sigma)} \tag{1.50}$$

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:

$$\overline{\mathbf{w}}_i = U^T\tilde{\mathbf{w}}_i, \qquad \overline{\mathbf{u}}_i = \mathbf{u}_i U, \qquad \overline{\Sigma} = U^T\Sigma U$$

Then the variance relations (1.16) and (1.50) can be equivalently rewritten as

$$\begin{aligned} \mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{\Sigma}} &= \mathrm{E}\,\|\overline{\mathbf{w}}_{i-1}\|^2_{\overline{\Sigma}'} \;+\; \mu^2\sigma_v^2\mathrm{E}\left(\frac{\|\overline{\mathbf{u}}_i\|^2_{\overline{\Sigma}}}{g^2[\overline{\mathbf{u}}_i]}\right) \\ \overline{\Sigma}' &= \overline{\Sigma} \;-\; \mu\overline{\Sigma}\Delta \;-\; \mu\Delta\overline{\Sigma} + \mu^2\Delta\mathrm{Tr}(\overline{\Sigma}\Delta) \end{aligned} \tag{1.51}$$

The expression for $\overline{\Sigma}'$ shows that it will be diagonal as long as $\overline{\Sigma}$ is diagonal. Thus let again

$$\overline{\sigma} = \mathrm{diag}(\overline{\Sigma}), \qquad \overline{\sigma}' = \mathrm{diag}(\overline{\Sigma}')$$

Then from (1.51) we find that

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

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

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

where $\delta = \mathrm{diag}(\Delta)$. Repeating the arguments that led to (1.22) we can establish that, under the assumed fourth-order moment approximation, the evolution of $\mathrm{E}\,\|\overline{\mathbf{w}}_i\|^2_{\overline{\sigma}}$ is described by an $M-$dimensional state-space model similar to (1.37).

## 1.11 Long Filter Approximation

In addition to the small step-size and fourth-order moment approximations of Secs. 1.8 and 1.10, we can also resort to a long filter approximation and derive simplified transient and steady-state performance results for data-normalized filters of the form (1.3). We postpone this discussion until Sec. 1.12.5, whereby the simplified results will be obtained as a special case of the theory we develop below for adaptive filters with error nonlinearities.

## 1.12 Adaptive Filters with Error Nonlinearities

The analysis in the earlier sections focused on data-normalized adaptive filters of the form (1.3). We now extend the energy-based arguments to filters with error nonlinearities in their update equations. This class of filters is usually more challenging to study. For this reason, we shall resort to a long filter assumption in order to make the analysis more tractable, as we explain in the sequel.

Thus consider filter updates of the form

$$\boxed{\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i^T f[\mathbf{e}(i)], \quad i \geq 0} \tag{1.52}$$

where

$$\boxed{\mathbf{e}(i) = \mathbf{d}(i) - \mathbf{u}_i \mathbf{w}_{i-1}} \tag{1.53}$$

is the estimation error at iteration $i$, and $f$ is some function of $\mathbf{e}(i)$. Typical choices for $f$ are

$$f[\mathbf{e}] = \mathbf{e} \quad \text{(LMS)}, \qquad f[\mathbf{e}] = \text{sign}(\mathbf{e}) \quad \text{(sign-LMS)}, \qquad f[\mathbf{e}] = \mathbf{e}^3 \quad \text{(LMF)}$$

The initial condition $\mathbf{w}_{-1}$ of (1.52) is assumed to be independent of all $\{\mathbf{d}(j), \mathbf{u}_j, \mathbf{v}(j)\}$.

The same argument that was employed in Sec. 1.3 can be repeated here to verify that the energy relation (1.10) still holds. Indeed, subtracting $w^o$ from both sides of (1.52) we get

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu \mathbf{u}_i^T f[\mathbf{e}(i)] \tag{1.54}$$

and multiplying (1.54) by $\mathbf{u}_i\Sigma$ from the left we find that

$$\mathbf{e}_p^{\Sigma}(i) = \mathbf{e}_a^{\Sigma}(i) - \mu \|\mathbf{u}_i\|_{\Sigma}^2 f[\mathbf{e}(i)] \tag{1.55}$$

Relation (1.55) can be used to express $f[\mathbf{e}(i)]$ in terms of $\{\mathbf{e}_p^{\Sigma}(i), \mathbf{e}_a^{\Sigma}(i)\}$ and to eliminate it from (1.54). Doing so leads to the equality

$$\|\mathbf{u}_i\|_{\Sigma}^2 \cdot \tilde{\mathbf{w}}_i + \mathbf{u}_i^T \mathbf{e}_a^{\Sigma}(i) \;=\; \|\mathbf{u}_i\|_{\Sigma}^2 \cdot \tilde{\mathbf{w}}_{i-1} + \mathbf{u}_i^T \mathbf{e}_p^{\Sigma}(i) \tag{1.56}$$

and by equating the weighted Euclidean norms of both sides of this equation we arrive again at (1.10), which is repeated here for ease of reference,

$$\boxed{\|\mathbf{u}_i\|_{\Sigma}^2 \cdot \|\tilde{\mathbf{w}}_i\|_{\Sigma}^2 \;+\; \left(\mathbf{e}_a^{\Sigma}(i)\right)^2 = \|\mathbf{u}_i\|_{\Sigma}^2 \cdot \|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma}^2 \;+\; \left(\mathbf{e}_p^{\Sigma}(i)\right)^2} \tag{1.57}$$

### 1.12.1 Variance Relation For Error Nonlinearities

Now recall that in transient analysis we are interested in characterizing the time evolution of the quantity $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|_\Sigma^2$, for some $\Sigma$ of interest (usually, $\Sigma = I$ or $\Sigma = R_u$). To characterize this evolution, we replace $\mathbf{e}_p^\Sigma(i)$ in (1.57) by its expression (1.55) in terms of $\mathbf{e}_a^\Sigma(i)$ and $\mathbf{e}(i)$ to get

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

Assuming the event $\|\mathbf{u}_i\|_\Sigma^2 = 0$ occurs with zero probability, we can eliminate $\|\mathbf{u}_i\|_\Sigma^2$ from both sides and take expectations to arrive at:

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|_\Sigma^2 = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 \;-\; 2\mu\mathrm{E}\left(\mathbf{e}_a^\Sigma(i) f[\mathbf{e}(i)]\right) \;+\; \mu^2\mathrm{E}\left(\|\mathbf{u}_i\|_\Sigma^2 f^2[\mathbf{e}(i)]\right)} \tag{1.58}$$

which is the equivalent of (1.13) for filters with error nonlinearities. Observe, however, that the weighting matrix for $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|_\Sigma^2$ and $\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2$ are still identical since we did not substitute $\{\mathbf{e}_a(i), \mathbf{e}(i)\}$ by their expressions in terms of $\tilde{\mathbf{w}}_{i-1}$. The reason we did not do so here is because of the nonlinear error function $f$. Instead, to proceed, we shall show how to evaluate the expectations

$$\mathrm{E}\left(\mathbf{e}_a^\Sigma(i) f[\mathbf{e}(i)]\right) \quad \text{and} \quad \mathrm{E}\left(\|\mathbf{u}_i\|_\Sigma^2 f^2[\mathbf{e}(i)]\right) \tag{1.59}$$

These expectations are generally hard to compute because of $f$. In order to facilitate their evaluation, we shall assume that the filter is long enough in order to justify, by central limit theorem arguments, that

$$\boxed{\mathbf{e}_a(i) \text{ and } \mathbf{e}_a^\Sigma(i) \text{ are jointly Gaussian random variables}} \tag{1.60}$$

**Evaluation of $\mathrm{E}\left(\mathbf{e}_a^\Sigma f[\mathbf{e}]\right)$**

Using (1.60) we can evaluate the first expectation, $\mathrm{E}\left(\mathbf{e}_a^\Sigma(i) f[\mathbf{e}(i)]\right)$, by appealing to Price's theorem [14]. The theorem states that if $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian random variables that are independent from a third random variable $\mathbf{z}$, then

$$\mathrm{E}\,\mathbf{x}k(\mathbf{y}+\mathbf{z}) = \frac{\mathrm{E}\,\mathbf{x}\mathbf{y}}{\mathrm{E}\,\mathbf{y}^2}\mathrm{E}\,\mathbf{y}k(\mathbf{y}+\mathbf{z})$$

where $k(\cdot)$ is some function of $\mathbf{y}+\mathbf{z}$. Using this result, together with the equality $\mathbf{e}(i) = \mathbf{e}_a(i) + \mathbf{v}(i)$, we get

$$\mathrm{E}\,\mathbf{e}_a^\Sigma(i) f[\mathbf{e}(i)] \;=\; \mathrm{E}\,\mathbf{e}_a^\Sigma(i)\mathbf{e}_a(i)\frac{\mathrm{E}\,\mathbf{e}_a(i) f[\mathbf{e}(i)]}{\mathrm{E}\,\mathbf{e}_a^2(i)} \;\stackrel{\Delta}{=}\; \left(\mathrm{E}\,\mathbf{e}_a^\Sigma(i)\mathbf{e}_a(i)\right)\cdot h_G$$

where the function $h_G$ is defined by

$$\boxed{h_G \;\stackrel{\Delta}{=}\; \frac{\mathrm{E}\,\mathbf{e}_a(i) f[\mathbf{e}(i)]}{\mathrm{E}\,\mathbf{e}_a^2(i)}} \tag{1.61}$$

Clearly, since $\mathbf{e}_a(i)$ is Gaussian, the expectation $\mathrm{E}\,\mathbf{e}_a(i)f[\mathbf{e}(i)]$ depends on $\mathbf{e}_a(i)$ only through its second moment, $\mathrm{E}\,\mathbf{e}_a^2(i)$. This means that $h_G$ itself is only a function of $\mathrm{E}\,\mathbf{e}_a^2(i)$. The function $h_G[\cdot]$ can be evaluated for different choices of the error nonlinearity $f[\cdot]$, as shown in Tab. 1.1.

### Evaluation of $\mathrm{E}\,(\|\mathbf{u}_i\|_\Sigma^2 f^2[\mathbf{e}])$

In order to evaluate the second expectation, $\mathrm{E}\,(\|\mathbf{u}_i\|_\Sigma^2 f^2[\mathbf{e}(i)])$, we resort to a separation assumption, viz., we assume that the filter is long enough so that

$$\boxed{\|\mathbf{u}_i\|_\Sigma^2 \text{ and } f^2[\mathbf{e}(i)] \text{ are uncorrelated}} \tag{1.62}$$

This assumption allows us to write

$$\mathrm{E}\,\left(\|\mathbf{u}_i\|_\Sigma^2 f^2[\mathbf{e}(i)]\right) = \left(\mathrm{E}\,\|\mathbf{u}_i\|_\Sigma^2\right)\cdot\left(\mathrm{E}\,f^2[\mathbf{e}(i)]\right) \stackrel{\Delta}{=} \left(\mathrm{E}\,\|\mathbf{u}_i\|_\Sigma^2\right)\cdot h_U$$

where the function $h_U$ is defined by

$$\boxed{h_U \stackrel{\Delta}{=} \mathrm{E}\,f^2[\mathbf{e}(i)]} \tag{1.63}$$

Again, since $\mathbf{e}_a(i)$ is Gaussian and independent of the noise, the function $h_U$ is a function of $E\mathbf{e}_a^2(i)$ only. The function $h_U$ can also be evaluated for different error nonlinearities, as shown in Tab. 1.1.

Table 1.1: *Expressions for $h_G$ and $h_U$ for some error nonlinearities. In the least-mean-fourth (LMF) case, we assume Gaussian noise for simplicity.*

| **Algorithm** | **Error nonlinearity** | $\{h_G, h_U\}$ |
|---|---|---|
| LMS | $f[\mathbf{e}] = \mathbf{e}$ | $h_G = 1$<br>$h_U = \mathrm{E}\,\mathbf{e}_a^2(i) + \sigma_v^2$ |
| sign-LMS | $f[\mathbf{e}] = \mathrm{sign}[\mathbf{e}]$ | $h_G = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{\mathrm{E}\,\mathbf{e}_a^2(i)+\sigma_v^2}}$<br>$h_U = 1$ |
| LMF | $f[\mathbf{e}] = \mathbf{e}^3$ | $h_G = 3(\mathrm{E}\,\mathbf{e}_a^2(i) + \sigma_v^2)$<br>$h_U = 15(\mathrm{E}\,\mathbf{e}_a^2(i) + \sigma_v^2)^3$ |

### 1.12.2 Independent Regressors

Using the definitions of $h_U$ and $h_G$, we can rewrite the variance relation (1.58) more compactly as

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\Sigma = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_\Sigma \;-\; 2\mu h_G \mathrm{E}\left(\mathbf{e}_a^\Sigma(i)\mathbf{e}_a(i)\right) \;+\; \mu^2 h_U \mathrm{Tr}(R_u\Sigma)} \tag{1.64}$$

As it stands, this relation is still hard to propagate since it requires the evaluation of $\mathrm{E}\,\mathbf{e}_a^\Sigma(i)\mathbf{e}_a(i)$, and this expectation is not trivial in general. This is because of possible dependencies among the successive regressors $\{\mathbf{u}_i\}$. However, if we again resort to the independence assumption (1.15), then it is easy to verify that

$$\mathrm{E}\,\mathbf{e}_a^\Sigma(i)\mathbf{e}_a(i) \;=\; \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma R_u}$$

so that (1.64) becomes

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_\Sigma = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_\Sigma \;-\; 2\mu h_G \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma R_u} \;+\; \mu^2 h_U \mathrm{Tr}(R_u\Sigma)} \tag{1.65}$$

We now illustrate the application of this result by considering two cases separately. We start with the simpler case of white input data followed by correlated data.

### 1.12.3 White Regression Data

Assume first that $R_u = \sigma_u^2 I$ and select $\Sigma = I$. Then (1.65) becomes

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;-\; 2\mu h_G \sigma_u^2 \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; \mu^2 M \sigma_u^2 h_U} \tag{1.66}$$

Note that all terms on the right-hand side are dependent on $E\|\tilde{\mathbf{w}}_{i-1}\|^2$ only; this is because $h_G$ and $h_U$ are functions of $E\mathbf{e}_a^2(i)$ and, for white input data, $E\mathbf{e}_a^2(i) = \sigma_u^2 E\|\tilde{\mathbf{w}}_{i-1}\|^2$. We therefore find that recursion (1.66) characterizes the evolution of $E\|\tilde{\mathbf{w}}_i\|^2$. Two special cases help demonstrate this fact.

#### Transient Behavior of LMS

When $f[\mathbf{e}] = \mathbf{e}$ we obtain the LMS algorithm,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i^T \mathbf{e}(i) \tag{1.67}$$

Using the following expressions from Tab. 1.1,

$$h_U \;=\; \sigma_u^2 \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 + \sigma_v^2, \qquad h_G \;=\; 1$$

we obtain

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \left(1 - 2\mu\sigma_u^2 + \mu^2\sigma_u^4 M\right)\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; \mu^2 M \sigma_u^2 \sigma_v^2 \tag{1.68}$$

which is a linear recursion in $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2$; it characterizes the transient behavior of LMS for white input data.

**Transient Behavior of sign-LMS**

When $f[\mathbf{e}] = \text{sign}(\mathbf{e})$ we obtain the sign-LMS algorithm,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i^T \mathsf{sign}[\mathbf{e}(i)] \tag{1.69}$$

Using the following expressions from Tab. 1.1,

$$h_U \;=\; 1, \quad h_G \;=\; \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{\sigma_u^2 \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 + \sigma_v^2}}$$

we obtain

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \left(1 - \sqrt{\frac{8}{\pi}} \frac{\mu\sigma_u^2}{\sqrt{\sigma_u^2 \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 + \sigma_v^2}}\right) \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; \mu^2 M \sigma_u^2 \tag{1.70}$$

which is now a nonlinear recursion in $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2$; it characterizes the transient behavior of sign-LMS for white input data.

**Transient Behavior of LMF**

When $f[\mathbf{e}] = \mathbf{e}^3$ we obtain the LMF algorithm,

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu \mathbf{u}_i^T \mathbf{e}^3(i) \tag{1.71}$$

Using the following expressions from Tab. 1,

$$h_G \;=\; 3\big(\mathrm{E}\,|\mathbf{e}_a(i)|^2 + \sigma_v^2\big), \quad h_U = 15\big(\;\mathrm{E}\,|\mathbf{e}_a(i)|^2 \;+\; \sigma_v^2\;\big)^3$$

we get

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = f\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; 15\mu^2 M \sigma_u^2 \sigma_v^6} \tag{1.72}$$

where

$$f = \big[1 + \mu\sigma_u^2\sigma_v^2(45\mu M\sigma_u^2\sigma_v^2 - 2) \;+\; \mu\sigma_u^4(45\mu M\sigma_u^2\sigma_v^2 - 2)\big]\,\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 + 15\mu^2\sigma_u^8 M \big(\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2\big)^2$$

which is a nonlinear recursion in $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2$; it characterizes the transient behavior of LMF for white input data.

## 1.12.4 Correlated Regression Data

When the input data is correlated, different weighting matrices will appear on both sides of the variance relation (1.65). Indeed, writing (1.65) for $\Sigma = I$ yields

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;-\; 2\mu h_G \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u} \;+\; \mu^2 \mathrm{Tr}(R_u) \cdot h_U$$

with the weighted term $\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u}$. This term can be deduced from (1.65) by writing it for $\Sigma = R_u$, which leads to

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u} = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u} \;-\; 2\mu h_G \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u^2} \;+\; \mu^2 h_U \mathrm{Tr}(R_u^2)$$

with the weighted term $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^2}$. This term can in turn be deduced from (1.65) by writing it for $\Sigma = R_u^2$. Continuing in this fashion, for successive powers of $R_u$, we arrive at

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^{M-1}} = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u^{M-1}} \;-\; 2\mu h_G \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u^M} \;+\; \mu^2 h_U \mathrm{Tr}(R_u^M)$$

As before, this procedure terminates. To see this, let $p(x) = \det(xI - R_u)$ denote the characteristic polynomial of $R_u$, say

$$p(x) = x^M + p_{M-1}x^{M-1} + p_{M-2}x^{M-2} + \ldots + p_1 x + p_0$$

Then, since $p(R_u) = 0$ in view of the Cayley-Hamilton theorem, we have

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R^M} \;=\; -p_0 \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 \;-\; p_1 \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u} \;-\; \ldots \;-\; p_{M-1} \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^{M-1}}$$

This result indicates that the weighted term $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R^M}$ is fully determined by the prior weighted terms.

Putting these results together, we find that the transient behavior of the filter (1.52) is now described by a nonlinear $M-$dimensional state-space model of the form

$$\boxed{\mathcal{W}_i = \mathcal{F}\mathcal{W}_{i-1} + \mu^2 h_U \mathcal{Y}} \tag{1.73}$$

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

$$\mathcal{W}_i \;\triangleq\; \begin{bmatrix} \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u} \\ \vdots \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^{M-2}} \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^{M-1}} \end{bmatrix}, \qquad \mathcal{Y} \;\triangleq\; \begin{bmatrix} \mathrm{Tr}(R_u) \\ \mathrm{Tr}(R_u^2) \\ \vdots \\ \mathrm{Tr}(R_u^{M-1}) \\ \mathrm{Tr}(R_u^M) \end{bmatrix} \tag{1.74}$$

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

$$\mathcal{F} \;\triangleq\; \begin{bmatrix} 1 & -2\mu h_G & & & \\ 0 & 1 & -2\mu h_G & & \\ 0 & 0 & 1 & -2\mu h_G & \\ \vdots & & & & \\ 0 & 0 & & 1 & -2\mu h_G \\ 2\mu p_0 h_G & 2\mu p_1 h_G & \ldots & 2\mu p_{M-2} h_G & 1 + 2\mu p_{M-1} h_G \end{bmatrix}$$

The evolution of the top entry of $\mathcal{W}_i$ describes the mean-square deviation of the filter, $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2$, while the evolution of the second entry of $\mathcal{W}_i$ relates the learning behavior of the filter since

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

### 1.12.5 Long Filter Approximation

The earlier results on filters with error nonlinearities can be used to provide an alternative simplified analysis of adaptive filters with data nonlinearities as in (1.3); just like we did in Secs. 1.8 and 1.10 by resorting to simplifications that resulted from the small step-size and fourth-order moment approximations.

Indeed, starting from (1.10), substituting $\mathbf{e}_p^{\Sigma}(i)$ in terms of $\{\mathbf{e}_a^{\Sigma}(i), \mathbf{e}(i)\}$ from (1.8), and taking expectations, we arrive at the variance relation

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{\Sigma} = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma} \;-\; 2\mu\mathrm{E}\left(\frac{\mathbf{e}_a^{\Sigma}(i)\mathbf{e}(i)}{g[\mathbf{u}_i]}\right) \;+\; \mu^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_{\Sigma}\mathbf{e}^2(i)}{g^2[\mathbf{u}_i]}\right)} \tag{1.75}$$

This relation is equivalent to (1.13), except that in (1.13) we proceeded further and expressed the terms $\mathbf{e}_a^{\Sigma}(i)\mathbf{e}(i)$ and $\mathbf{e}^2(i)$ as weighted norms of $\tilde{\mathbf{w}}_{i-1}$. Relation (1.75) has the same form as the variance relation (1.58) used for filters with error nonlinearities. Observe in particular that the function $\mathbf{e}/g[\mathbf{u}]$ in data-normalized filters plays the role of $f[\mathbf{e}]$ in nonlinear error filters.

Now by following the arguments of Sec. 1.12.1, and under the following assumptions:

$$\boxed{\begin{array}{l} \mathbf{e}_a(i) \text{ and } \mathbf{e}_a^{\Sigma}(i) \text{ are jointly Gaussian random variables.} \\[1em] \|\mathbf{u}_i\|^2_{\Sigma} \text{ and } g[\mathbf{u}_i] \text{ are independent of } \mathbf{e}(i). \\[1em] \text{The regressors } \mathbf{u}_i \text{ are independent and identically distributed.} \end{array}} \tag{1.76}$$

we can evaluate the expectations

$$\mathrm{E}\left(\frac{\mathbf{e}_a^{\Sigma}(i)\mathbf{e}(i)}{g[\mathbf{u}_i]}\right) \qquad \text{and} \qquad \mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_{\Sigma}\mathbf{e}^2(i)}{g^2[\mathbf{u}_i]}\right)$$

and conclude that the variance relation (1.75) reduces to

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{\Sigma} = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma} \;-\; 2\mu h_G\mathrm{E}\left(\mathbf{e}_a^{\Sigma}(i)\mathbf{e}_a(i)\right) \;+\; \mu^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2[\mathbf{u}_i]}\right)(\mathrm{E}\,\mathbf{e}_a^2(i) + \sigma_v^2)$$

where now

$$h_G \;\triangleq\; \frac{\mathrm{E}\left(\mathbf{e}_a^2(i)/g[\mathbf{u}_i]\right)}{\mathrm{E}\,\mathbf{e}_a^2(i)} \;=\; \mathrm{E}\left(\frac{1}{g[\mathbf{u}_i]}\right) \tag{1.77}$$

in view of the independence assumptions in (1.76).

If we again use $\mathrm{E}\,\mathbf{e}_a^2(i) = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|_{R_u}$, then we arrive at

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{\Sigma} = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma} \;-\; \mu h_G\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{\Sigma R_u + R_u\Sigma} \;+\; \mu^2\mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2_{\Sigma}}{g^2[\mathbf{u}_i]}\right)(\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u} + \sigma_v^2)} \tag{1.78}$$

which is the extension of (1.65) to data-normalized filters. We now illustrate the application of this result to the transient analysis of some data-normalized adaptive filters.

**White Regression Data**

Assume first that $R_u = \sigma_u^2 I$ and select $\Sigma = I$. Then (1.78) becomes

$$\boxed{\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \left(1 - 2\mu\sigma_u^2 h_G + \mu^2\sigma_u^2 \mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]}\right)\right) \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; \mu^2\sigma_v^2 \mathrm{E}\left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]}\right)} \tag{1.79}$$

For the special case of LMS, when $g[\mathbf{u}] = 1$, $h_G$ in (1.77) becomes $h_G = 1$ and (1.79) reduces to

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \left(1 - 2\mu\sigma_u^2 + \mu^2\sigma_u^4 M\right) \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 + \; \mu^2 M \sigma_u^2 \sigma_v^2 \tag{1.80}$$

This is the same recursion we obtained before for LMS when trained with white input data.

For the special case of NLMS, $g[\mathbf{u}] = \|\mathbf{u}\|^2$, and relation (1.79) reduces to

$$\boxed{\|\tilde{\mathbf{w}}_i\|^2 = \left(1 - 2\mu\sigma_u^2 \mathrm{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right) + \mu^2\sigma_u^2 \mathrm{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)\right) \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; \mu^2\sigma_v^2 \mathrm{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)} \tag{1.81}$$

**Correlated Regression Data**

When the input data are correlated, different weighting matrices will appear on both sides of the variance relation (1.78). Indeed, writing (1.78) for $\Sigma = I$ yields

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

with the weighted term $\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|_{R_u}$ . This term can be deduced from (1.78) by writing it for $\Sigma = R_u$, which leads to

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

with the weighted term $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^2}$ and so forth. The procedure terminates and leads to the following state-space model:

$$\boxed{\mathcal{W}_i = \left(\mathcal{F} + \mu^2 \mathcal{Y} e_2^T\right)\mathcal{W}_{i-1} + \mu^2\sigma_v^2 \mathcal{Y}} \tag{1.82}$$

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

$$\mathcal{W}_i \;\triangleq\; \begin{bmatrix} \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u} \\ \vdots \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^{M-2}} \\ \mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2_{R_u^{M-1}} \end{bmatrix}, \qquad \mathcal{Y} \;\triangleq\; \begin{bmatrix} \mathrm{E}\left(\|\mathbf{u}_i\|^2/g^2[\mathbf{u}_i]\right) \\ \mathrm{E}\left(\|\mathbf{u}_i\|^2_{R_u}/g^2[\mathbf{u}_i]\right) \\ \vdots \\ \mathrm{E}\left(\|\mathbf{u}_i\|^2_{R_u^{M-2}}/g^2[\mathbf{u}_i]\right) \\ \mathrm{E}\left(\|\mathbf{u}_i\|^2_{R_u^{M-1}}/g^2[\mathbf{u}_i]\right) \end{bmatrix} \tag{1.83}$$

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

$$\mathcal{F} \triangleq \begin{bmatrix} 1 & -2\mu h_G & & & \\ 0 & 1 & -2\mu h_G & & \\ 0 & 0 & 1 & -2\mu h_G & \\ \vdots & & & & \\ 0 & 0 & & 1 & -2\mu h_G \\ 2\mu p_0 h_G & 2\mu p_1 h_G & \dots & 2\mu p_{M-2} h_G & 1 + 2\mu p_{M-1} h_G \end{bmatrix}$$

and

$$e_2 = \text{col}\{0, 1, 0, \dots, 0\}$$

Also,

$$h_G = \text{E}\left(\frac{1}{g[\mathbf{u}_i]}\right)$$

The evolution of the top entry of $\mathcal{W}_i$ describes the mean-square deviation of the filter, $\text{E}\,\|\tilde{\mathbf{w}}_i\|^2$, while the evolution of the second entry of $\mathcal{W}_i$ relates to the learning behavior of the filter. The model (1.82) is an alternative to (1.22) for adaptive filters with data nonlinearities; it is based on assumptions (1.76).

### Steady-State Performance

The variance relation (1.78) can also be used to approximate the steady-state performance of data-normalized adaptive filters. Writing it for $\Sigma = I$,

$$\text{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \text{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;-\; 2\mu h_G \text{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u} \;+\; \mu^2 \text{E}\left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]}\right)(\text{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u} + \sigma_v^2) \qquad (1.84)$$

and setting, in steady-state,

$$\lim_{i\to\infty} \text{E}\,\|\tilde{\mathbf{w}}_i\|^2 \;=\; \lim_{i\to\infty} \text{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2$$

we obtain

$$0 = -2\mu \text{E}\left(\frac{1}{g[\mathbf{u}_i]}\right)\text{EMSE} \;+\; \mu^2 \text{E}\left(\frac{\|\mathbf{u}_i\|^2}{g^2[\mathbf{u}_i]}\right)(\text{EMSE} + \sigma_v^2)$$

so that the excess mean-square error, $\text{E}\,\mathbf{e}_a^2(\infty)$, is given by

$$\boxed{\text{EMSE} = \frac{\mu\sigma_v^2 \text{Tr}(Q)}{2\text{E}\left(1/g[\mathbf{u}_i]\right) - \mu \text{Tr}(Q)}} \qquad (1.85)$$

where $Q = \text{E}\left(\mathbf{u}_i^T \mathbf{u}_i / g^2[\mathbf{u}_i]\right)$. For LMS we have $g[\mathbf{u}] = 1$ and $Q = R_u$, and the above expression reduces to

$$\text{EMSE} = \frac{\mu\sigma_v^2 \text{Tr}(R_u)}{2 - \mu \text{Tr}(R_u)} \qquad \text{(LMS)}$$

For NLMS we have $g[\mathbf{u}] = \|\mathbf{u}\|^2$ and $Q = \mathrm{E}\left(\mathbf{u}_i^T \mathbf{u}_i / \|\mathbf{u}_i\|^4\right)$, so that

$$\text{EMSE} \approx \frac{\mu\sigma_v^2}{2-\mu} \qquad \text{(NLMS)}$$

**Stability**

The recursion (1.84) can be rearranged as

$$\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2 = \mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; \mu\left(\mu\mathrm{Tr}(Q) - 2h_G\right)\mathrm{E}\,\|\tilde{\mathbf{w}}_{i-1}\|^2_{R_u} \;+\; \mu^2\sigma_v^2\mathrm{Tr}(Q)$$

It is now easy to see that $\mathrm{E}\,\|\tilde{\mathbf{w}}_i\|^2$ converges for step-sizes satisfying

$$\mu\mathrm{Tr}(Q) - 2h_G \;<\; 0$$

or, equivalently,

$$0 < \mu < \frac{2h_G}{\mathrm{Tr}(Q)} \;=\; 2\mathrm{E}\left(\frac{1}{g[\mathbf{u}_i]}\right)\frac{1}{\mathrm{Tr}(Q)}$$

For LMS, this simplified analysis results in the condition $\mu < 2/\mathrm{Tr}(R_u)$. For NLMS, $\mathrm{Tr}(Q) = E(1/\|\mathbf{u}_i\|^2)$ and the condition on $\mu$ becomes $\mu < 2$.

## 1.13 An Interpretation of the Energy Relation

We end our discussions in this chapter by making a connection between the energy relation (1.10) and Snell's law of optics. We reconsider (1.10) and assume first that $\Sigma = I$ so that

$$\|\mathbf{u}_i\|^2 \cdot \|\tilde{\mathbf{w}}_i\|^2 \;+\; \mathbf{e}_a^2(i) = \|\mathbf{u}_i\|^2 \cdot \|\tilde{\mathbf{w}}_{i-1}\|^2 \;+\; \mathbf{e}_p^2(i) \tag{1.86}$$

Let $\theta_i$ denote the acute angle between the column vectors $\{\tilde{\mathbf{w}}_i, \mathbf{u}_i^T\}$. Likewise, let $\theta_{i-1}$ denote the acute angle between $\{\tilde{\mathbf{w}}_{i-1}, \mathbf{u}_i^T\}$. Then

$$\mathbf{e}_a^2(i) = \|\mathbf{u}_i\|^2 \cdot \|\tilde{\mathbf{w}}_{i-1}\|^2 \cdot \cos^2(\theta_{i-1}), \qquad \text{and} \qquad \mathbf{e}_p^2(i) = \|\mathbf{u}_i\|^2 \cdot \|\tilde{\mathbf{w}}_i\|^2 \cdot \cos^2(\theta_i)$$

Substituting into (1.86) and collecting terms we find that it reduces to

$$\boxed{\|\tilde{\mathbf{w}}_{i-1}\|^2 \sin^2(\theta_{i-1}) \;=\; \|\tilde{\mathbf{w}}_i\|^2 \sin^2(\theta_i)} \tag{1.87}$$

Equality (1.87) resembles a famous result in optics, known as Snell's law, which relates the refraction indices of two mediums with the sines of the incident and refracted rays of light, viz.,

$$\eta_1 \sin\theta_1 = \eta_2 \sin\theta_2$$

where $\theta_1$ and $\theta_2$ are the angles of incidence and refraction, respectively; both angles are measured relative to the direction that is orthogonal to the surface separating both mediums. This analogy suggests that we can relate the operation of an adaptive filter, at each iteration,

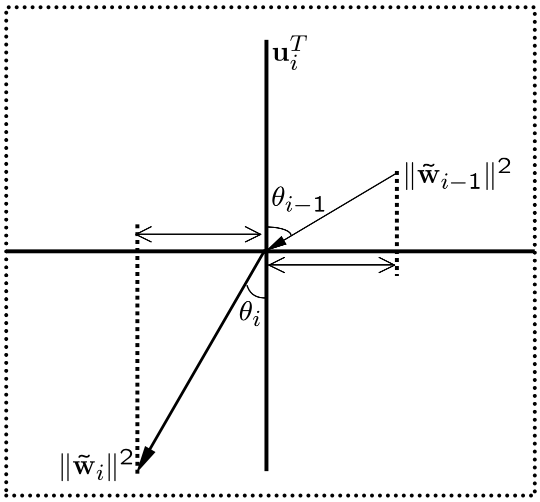


Figure 1.1: *An interpretation of the energy-conservation relation (1.10) by means of an analogy with Snell's law in optics.*

to that of a fictitious ray travelling from one medium to another. The magnitudes $\|\tilde{\mathbf{w}}_{i-1}\|$ and $\|\tilde{\mathbf{w}}_i\|$ play the role of refraction indices of the mediums, while $\{\theta_{i-1}, \theta_i\}$ play the role of the incidence and refraction angles of the ray. Alternatively, we can interpret the result (1.87) as shown in Fig. 1.1. An incident vector of norm $\|\tilde{\mathbf{w}}_{i-1}\|$ impinges on the separation layer at an angle $\theta_{i-1}$ with respect to $\mathbf{u}_i^T$, while a refracted vector of norm $\|\tilde{\mathbf{w}}_i\|$ leaves the layer at an angle $\theta_i$, also with respect to $\mathbf{u}_i^T$. Relation (1.87) then amounts to saying that the projections of these vectors along the horizontal direction should have equal norms.

More generally, when a positive-definite weighting matrix $\Sigma$ is present in (1.10), we let $\{\theta_i, \theta_{i-1}\}$ denote acute angles whose squared cosines are given by

$$\cos^2_\Sigma(\theta_{i-1}) \;\triangleq\; \frac{\left(\mathbf{e}_a^\Sigma(i)\right)^2}{\|\tilde{\mathbf{w}}_{i-1}\|^2_\Sigma \cdot \|\mathbf{u}_i\|^2_\Sigma}, \qquad \cos^2_\Sigma(\theta_i) \;\triangleq\; \frac{\left(\mathbf{e}_p^\Sigma(i)\right)^2}{\|\tilde{\mathbf{w}}_i\|^2_\Sigma \cdot \|\mathbf{u}_i\|^2_\Sigma} \tag{1.88}$$

The subscript $\Sigma$ in $\cos_\Sigma(\cdot)$ indicates that a weighting matrix $\Sigma$ is used in computing it. With this notation, it is straightforward to verify that the energy-relation (1.57) becomes

$$\boxed{\|\tilde{\mathbf{w}}_{i-1}\|^2_\Sigma \sin^2_\Sigma(\theta_{i-1}) \;=\; \|\tilde{\mathbf{w}}_i\|^2_\Sigma \sin^2_\Sigma(\theta_i)} \tag{1.89}$$

which is a natural extension of (1.87).

## 1.14 Concluding Remarks

This chapter describes an energy-conservation approach to studying the performance of adaptive filters. 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., (1.16), (1.64), and (1.65)) and, subsequently, in terms of a state-space model (e.g., (1.22) and (1.73)). The approach does not restrict the input data to Gaussian or white distributions. 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 analysis in this chapter relied on the independence assumption (1.15), steady-state results can be obtained without relying on this assumption (see, e.g., [10, 11]).

# Bibliography


[1] S. Haykin, *Adaptive Filter Theory,* Prentice-Hall, NJ, 1996.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Prentice Hall, NJ, 1985.

[3] O. Macchi, *Adaptive Processing: The LMS Approach with Applications in Transmission,* Wiley, NY, 1995.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] T. Y. Al-Naffouri and A. H. Sayed,“Transient analysis of data-normalized adaptive filters,” to appear in *IEEE Transactions on Signal Processing*, Mar. 2003.

[13] T. Y. Al-Naffouri and A. H. Sayed, “Transient analysis of adaptive filters with error nonlinearities,” to appear in *IEEE Transactions on Signal Processing*, Mar. 2003.

[14] R. Price, “A useful theorem for nonlinear devices having Gaussian inputs,” *IEEE Trans. Information Theory,* vol. 4, pp. 69-72, Jun. 1958.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, NY, 1985.

[16] V. H. Nascimento and A. H. Sayed, “On the learning mechanism of adaptive filters,” *IEEE Transactions on Signal Processing,* vol. 48, no. 6, pp. 1609–1625, June 2000.

[17] A. H. Sayed and T. Y. Al-Naffouri, “Mean-square analysis of normalized leaky adaptive filters,” *Proc. ICASSP*, vol. 6, pp. 3873–3876, Salt Lake City, Utah, May 2001.