EE210A: Adaptation and Learning

Professor Ali H. Sayed

MEAN-SQUARE PERFORMANCE

Sections in order: 15.1-15.4, 16.1-16.6, 19.3

GRADIENT NOISE

stochastic approximations introduce gradient noise and, consequently, the performance of adaptive filters will degrade in comparison with the performance of the original steepest-descent methods.

The purpose of this chapter, and of the subsequent chapters in this part (*Mean-Square Performance*) and in Part V (*Transient Performance*), is to describe a unifying framework for the evaluation of the performance of adaptive filters. This objective is rather challenging, especially since adaptive filters are, by design, time-variant, stochastic, and nonlinear systems. Their update recursions not only depend on the reference and regression data in a nonlinear and time-variant manner, but the data they employ are also stochastic in nature. For this reason, the study of the performance of adaptive algorithms is a formidable task, so much so that *exact* performance analyses are rare and limited to special cases. It is customary to introduce simplifying assumptions in order to make the performance analyses more tractable. Fortunately, most assumptions tend to lead to reasonable agreements between theory and practice.

15.1 PERFORMANCE MEASURE

We use the LMS filter as a motivation for our explanations.

Thus recall that the steepest-descent iteration (8.20), namely,

$$w_i = w_{i-1} + \mu[R_{du} - R_u w_{i-1}] \tag{15.1}$$

was reduced to the LMS recursion (10.10), i.e.,

$$w_i = w_{i-1} + \mu u_i^*[d(i) - u_i w_{i-1}] \tag{15.2}$$

by replacing the second-order moments $R_{du} = \mathsf{E}\,\boldsymbol{d}\boldsymbol{u}^*$ and $R_u = \mathsf{E}\,\boldsymbol{u}^*\boldsymbol{u}$ by the instantaneous approximations

$$R_{du} \approx d(i)u_i^* \quad \text{and} \quad R_u \approx u_i^* u_i \tag{15.3}$$

Other adaptive algorithms were obtained in Chapter 10 by using similar instantaneous approximations. Recall further that we examined the convergence properties of (15.1) in Chapter 8 in some detail. Specifically, we established in Thm. 8.2 that by choosing the step-size μ such that

$$0 < \mu < 2/\lambda_{\max} \tag{15.4}$$

PERFORMANCE MEASURE

where $\lambda_{\max}$ is the largest eigenvalue of R_u, the successive weight estimates w_i of (15.1) are guaranteed to converge to the solution w^o of the normal equations, i.e., to the vector

$$w^o = R_u^{-1} R_{du} \tag{15.5}$$

that solves the least-mean-squares problem

$$\min_w \ \mathsf{E}\ |\boldsymbol{d} - \boldsymbol{u}w|^2 \tag{15.6}$$

Correspondingly, the learning curve of the steepest-descent method (15.1), namely,

$$\begin{aligned} J(i) &= \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w_{i-1}|^2 \\ &= \sigma_d^2 - R_{du}^* w_{i-1} - w_{i-1}^* R_{du} + w_{i-1}^* R_u w_{i-1} \end{aligned}$$

is also guaranteed to converge to the minimum cost of (15.6), i.e.,

$$J(i) \to J_{\min} \ \stackrel{\Delta}{=} \ \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w^o|^2 \ = \ \sigma_d^2 \ - \ R_{ud} R_u^{-1} R_{du} \tag{15.7}$$

where $\sigma_d^2 = \mathsf{E}\,|\boldsymbol{d}|^2$.

PERFORMANCE MEASURE

Obviously, the behavior of the weight estimates w_i that are generated by the stochastic-gradient approximation (15.2) is more complex than the behavior of the weight estimates w_i that are generated by the steepest-descent method (15.1). This is because the w_i from (15.2) need not converge to w^o anymore due to gradient noise. It is the purpose of Parts IV (*Mean-Square Performance*) and V (*Transient Performance*) to examine the effect of gradient noise on filter performance, not only for LMS but also for several other adaptive filters.

Stochastic Equations

First, however, in all such performance studies, it is necessary to regard (or treat) the update equation of an adaptive filter as a *stochastic* difference equation rather than a *deterministic* difference equation. What this means is that we need to regard the variables that appear in an update equation of the form (15.2) as *random variables*. Recall our convention in this book that random variables are represented in boldface, while realizations of random variables are represented in normal font.

For this reason, we shall write from now on $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$, with boldface letters, instead of $\{d(i), u_i\}$. The notation $\boldsymbol{d}(i)$ refers to a zero-mean random variable with variance σ_d^2, while $\boldsymbol{u}_i$ denotes a zero-mean row vector with covariance matrix R_u,

$$\mathsf{E}\,|\boldsymbol{d}(i)|^2 = \sigma_d^2, \quad \mathsf{E}\,\boldsymbol{u}_i^*\boldsymbol{u}_i = R_u, \quad \mathsf{E}\,\boldsymbol{d}(i)\boldsymbol{u}_i^* = R_{du}$$

In the same vein, we shall replace the weight estimates w_i and w_{i-1} in the update equation of an adaptive algorithm by $\boldsymbol{w}_i$ and $\boldsymbol{w}_{i-1}$, respectively, since, by being functions of $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$, they become random variables as well.

STOCHASTIC EQUATIONS

In this way, the stochastic equation that corresponds to the LMS filter (15.2) would be

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* [\boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}], \qquad \text{(a stochastic equation)}$$

with the initial condition $\boldsymbol{w}_{-1}$ also treated as a random vector. When this equation is implemented as an adaptive algorithm, it would operate on observations $\{d(i), u_i\}$ of the random quantities $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$, in which case the stochastic equation would be replaced by our earlier *deterministic* description (15.2) for LMS, namely,

$$w_i = w_{i-1} + \mu u_i^* [d(i) - u_i w_{i-1}], \qquad \text{(a deterministic equation)}$$

Similar considerations are valid for the update equations of all other adaptive algorithms. In all of them, we replace the deterministic quantities $\{d(i), u_i, w_i, w_{i-1}\}$ by random quantities $\{\boldsymbol{d}(i), \boldsymbol{u}_i, \boldsymbol{w}_i, \boldsymbol{w}_{i-1}\}$ in order to obtain the corresponding stochastic equations.

EXCESS MEAN-SQUARE ERROR (EMSE)

Excess Mean-Square Error and Misadjustment

Now in order to compare the performance of adaptive filters, it is customary to adopt a *common* performance measure across filters (even though different filters may have been derived by minimizing different cost functions). The criterion that is most widely used in the literature of adaptive filtering is the *steady-state mean-square error* (MSE) criterion, which is defined as

$$\boxed{\mathsf{MSE} \triangleq \lim_{i\to\infty} \mathsf{E}\,|\boldsymbol{e}(i)|^2} \tag{15.8}$$

where $\boldsymbol{e}(i)$ denotes the *a priori* output estimation error,

$$\boxed{\boldsymbol{e}(i) \triangleq \boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}} \tag{15.9}$$

Obviously, if the weight estimator $\boldsymbol{w}_{i-1}$ in (15.9) is replaced by the optimal solution w^o of (15.6), then the value of the MSE would coincide with the minimum cost (15.7), namely,

$$J_{\min} = \sigma_d^2 - R_{ud} R_u^{-1} R_{du}$$

EXCESS MEAN-SQUARE ERROR (EMSE)

For this reason, it is common to define the *excess-mean-square error* (EMSE) of an adaptive filter as the difference

$$\boxed{\mathsf{EMSE} \;\stackrel{\Delta}{=}\; \mathsf{MSE} - J_{\min}} \tag{15.10}$$

It is also common to define a relative measure of performance, called *misadjustment*, as

$$\boxed{\mathcal{M} \;\stackrel{\Delta}{=}\; \mathsf{EMSE}/J_{\min}} \tag{15.11}$$

15.2 STATIONARY DATA MODEL

Therefore, given a stochastic difference equation describing an adaptive filter, we are interested in evaluating its EMSE. In order to pursue this objective, and in order to facilitate the ensuing performance analysis, we also need to adopt a model for the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$.

To begin with, recall from the orthogonality principle of linear least-mean-squares estimation (cf. Thm. 4.1) that the solution w^o of (15.6) satisfies the uncorrelatedness property

$$\mathsf{E}\,\boldsymbol{u}_i^*(\boldsymbol{d}(i) - \boldsymbol{u}_i w^o) = 0$$

Let $\boldsymbol{v}(i)$ denote the estimation error (residual), i.e.,

$$\boldsymbol{v}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i w^o$$

Then we can re-express this result by saying that $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ are related via

$$\boldsymbol{d}(i) = \boldsymbol{u}_i w^o + \boldsymbol{v}(i) \tag{15.12}$$

in terms of a signal $\boldsymbol{v}(i)$ that is uncorrelated with $\boldsymbol{u}_i$.

STATIONARY DATA

The variance of $\boldsymbol{v}(i)$ is obviously equal to the minimum cost $J_{\min}$ from (15.7), i.e.,

$$\sigma_v^2 \triangleq \mathsf{E}\,|\boldsymbol{v}(i)|^2 = J_{\min} = \sigma_d^2 - R_{ud}R_u^{-1}R_{du} \qquad (15.13)$$

Linear Regression Model

What the above argument shows is that given any random variables $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ with second-order moments $\{\sigma_d^2, R_u, R_{du}\}$, we can always assume that $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ are related via a linear model of the form (15.12), for some w^o, with the variable $\boldsymbol{v}(i)$ playing the role of a disturbance that is uncorrelated with $\boldsymbol{u}_i$, i.e.,

$$\boxed{\boldsymbol{d}(i) = \boldsymbol{u}_i w^o + \boldsymbol{v}(i), \qquad \mathsf{E}\,|\boldsymbol{v}(i)|^2 = \sigma_v^2, \qquad \mathsf{E}\,\boldsymbol{v}(i)\boldsymbol{u}_i^* = 0} \qquad (15.14)$$

However, in order to make the performance analyses of adaptive filters more tractable, we usually need to adopt the stronger *assumption* that

$$\boxed{\text{The sequence } \{\boldsymbol{v}(i)\} \text{ is i.i.d. and independent of all } \{\boldsymbol{u}_j\}} \qquad (15.15)$$

STATIONARY DATA

Here, the notation i.i.d. stands for "independent and identically distributed". Condition (15.15) on $\boldsymbol{v}(i)$ is an assumption because, as explained above, the signal $\boldsymbol{v}(i)$ in (15.14) is only uncorrelated with $\boldsymbol{u}_i$; it is not necessarily independent of $\boldsymbol{u}_i$, or of all $\{\boldsymbol{u}_j\}$ for that matter. Still, there are situations when conditions (15.14) and (15.15) hold simultaneously, e.g., in the channel estimation application of Sec. 10.5. In that application, it is usually justified to expect the noise sequence $\{\boldsymbol{v}(i)\}$ to be i.i.d. and independent of all other data, including the regression data.

Given the above, we shall therefore adopt the following data model in our studies of the performance of adaptive filters. We shall assume that the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy the following conditions:

(a) There exists a vector w^o such that $\boldsymbol{d}(i) = \boldsymbol{u}_i w^o + \boldsymbol{v}(i)$.
(b) The noise sequence $\{\boldsymbol{v}(i)\}$ is i.i.d. with variance $\sigma_v^2 = \mathsf{E}\,|\boldsymbol{v}(i)|^2$.
(c) The noise sequence $\{\boldsymbol{v}(i)\}$ is independent of $\boldsymbol{u}_j$ for all i, j.
(d) The initial condition $\boldsymbol{w}_{-1}$ is independent of all $\{\boldsymbol{d}(j), \boldsymbol{u}_j, \boldsymbol{v}(j)\}$.
(e) The regressor covariance matrix is $R_u = \mathsf{E}\,\boldsymbol{u}_i^* \boldsymbol{u}_i > 0$.
(f) The random variables $\{\boldsymbol{d}(i), \boldsymbol{v}(i), \boldsymbol{u}_i\}$ have zero means.

(15.16)

Useful Independence Results

An important consequence of the data model (15.16) is that, at any particular time instant i, the noise variable $\boldsymbol{v}(i)$ will be independent of all previous weight estimators $\{\boldsymbol{w}_j, j < i\}$. This fact follows easily from examining the update equation of an adaptive filter. Consider, for instance, the LMS recursion

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* [\boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}], \quad \boldsymbol{w}_{-1} = \text{initial condition}$$

By iterating the recursion we find that, at any time instant j, the weight estimator $\boldsymbol{w}_j$ can be expressed as a function of $\boldsymbol{w}_{-1}$, the reference signals $\{\boldsymbol{d}(j), \boldsymbol{d}(j-1), \ldots, \boldsymbol{d}(0)\}$, and the regressors $\{\boldsymbol{u}_j, \boldsymbol{u}_{j-1}, \ldots, \boldsymbol{u}_0\}$. We denote this dependency generically as

$$\boldsymbol{w}_j = \mathcal{F}\,[\,\boldsymbol{w}_{-1};\ \boldsymbol{d}(j), \boldsymbol{d}(j-1), \ldots, \boldsymbol{d}(0);\ \boldsymbol{u}_j, \boldsymbol{u}_{j-1}, \ldots, \boldsymbol{u}_0\,] \tag{15.17}$$

for some function $\mathcal{F}$. A similar dependency holds for other adaptive schemes.

INDEPENDENCE RELATIONS

Now $\boldsymbol{v}(i)$ can be seen to be independent of each one of the terms appearing as an argument of $\mathcal{F}$ in (15.17), so that $\boldsymbol{v}(i)$ will be independent of $\boldsymbol{w}_j$ for all $j < i$. Indeed, the independence of $\boldsymbol{v}(i)$ from $\{\boldsymbol{w}_{-1}, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_0\}$ is obvious by assumption, while its independence from $\{\boldsymbol{d}(j), \ldots, \boldsymbol{d}(0)\}$ can be seen as follows. Consider $\boldsymbol{d}(j)$ for example. Then from $\boldsymbol{d}(j) = \boldsymbol{u}_j w^o + \boldsymbol{v}(j)$ we see that $\boldsymbol{d}(j)$ is a function of $\{\boldsymbol{u}_j, \boldsymbol{v}(j)\}$, both of which are independent of $\boldsymbol{v}(i)$.

Given that $\boldsymbol{v}(i)$ is independent of $\{\boldsymbol{w}_j, j < i\}$, it also follows that $\boldsymbol{v}(i)$ is independent of $\{\tilde{\boldsymbol{w}}_j, j < i\}$, where $\tilde{\boldsymbol{w}}_j$ denotes the weight-error vector:

$$\boxed{\tilde{\boldsymbol{w}}_j \;\stackrel{\Delta}{=}\; w^o - \boldsymbol{w}_j}$$

Moreover, $\boldsymbol{v}(i)$ is also independent of the *a priori* estimation error $\boldsymbol{e}_a(i)$, defined by

$$\boxed{\boldsymbol{e}_a(i) \;\stackrel{\Delta}{=}\; \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1}}$$

This variable measures the difference between $\boldsymbol{u}_i w^o$ and $\boldsymbol{u}_i \boldsymbol{w}_{i-1}$, i.e., it measures how close the estimator $\boldsymbol{u}_i \boldsymbol{w}_{i-1}$ is to the optimal linear estimator of $\boldsymbol{d}(i)$, namely $\hat{\boldsymbol{d}}(i) = \boldsymbol{u}_i w^o$.

INDEPENDENCE RELATIONS

Lemma 15.1 (Useful properties) From the data model (15.16), it follows that $\boldsymbol{v}(i)$ is independent of each of the following:

$$\{\boldsymbol{w}_j \ \text{ for } j < i\}, \quad \{\tilde{\boldsymbol{w}}_j \ \text{ for } j < i\}, \quad \text{and} \quad \boldsymbol{e}_a(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1}$$

***Alternative Expression for the* EMSE**

Using model (15.16), and the independence results of Lemma 15.1, we can determine a more compact expression for the EMSE of an adaptive filter. Recall from (15.8) and (15.10) that, by definition,

$$\text{EMSE} = \lim_{i\to\infty} \mathsf{E}\,|\boldsymbol{e}(i)|^2 \ - \ J_{\min} \tag{15.18}$$

where, as we already know from (15.13), $J_{\min} = \sigma_v^2$.

EMSE

Using

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

and the linear model (15.16), we find that

$$\boldsymbol{e}(i) = \boldsymbol{v}(i) + \boldsymbol{u}_i(w^o - \boldsymbol{w}_{i-1})$$

That is,

$$\boxed{\boldsymbol{e}(i) = \boldsymbol{v}(i) + \boldsymbol{e}_a(i)} \qquad (15.19)$$

Now the independence of $\boldsymbol{v}(i)$ and $\boldsymbol{e}_a(i)$, as stated in Lemma 15.1, gives

$$\mathsf{E}\,|\boldsymbol{e}(i)|^2 \;=\; \mathsf{E}\,|\boldsymbol{v}(i)|^2 \;+\; \mathsf{E}\,|\boldsymbol{e}_a(i)|^2 \;=\; \sigma_v^2 \;+\; \mathsf{E}\,|\boldsymbol{e}_a(i)|^2 \qquad (15.20)$$

so that substituting into (15.18) we get

$$\boxed{\mathsf{EMSE} = \lim_{i\to\infty} \mathsf{E}\,|\boldsymbol{e}_a(i)|^2} \qquad (15.21)$$

EMSE

In other words, the EMSE can be computed by evaluating the steady-state mean-square value of the *a priori* estimation error $e_a(i)$. We shall use this alternative representation to evaluate the EMSE of several adaptive algorithms in this chapter. Likewise, from (15.8) and (15.20), we have

$$\boxed{\text{MSE} \;=\; \text{EMSE} \;+\; \sigma_v^2} \qquad (15.22)$$

TABLE 15.1 Definitions of several estimation errors.

Error	Definition	Interpretation
$e(i)$	$d(i) - u_i w_{i-1}$	*a priori* output estimation error
$r(i)$	$d(i) - u_i w_i$	*a posteriori* output estimation error
$\tilde{w}_i$	$w^o - w_i$	weight error vector
$e_a(i)$	$u_i \tilde{w}_{i-1}$	*a priori* estimation error
$e_p(i)$	$u_i \tilde{w}_i$	*a posteriori* estimation error

15.3 ENERGY CONSERVATION RELATION

Our approach to the performance analysis of adaptive filters in Parts IV (*Mean-Square Performance*) and V (*Transient Performance*) is based on an energy conservation relation that holds for general data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ (it does *not* even require the assumptions (15.16)).

In order to motivate this relation, we consider adaptive filter updates of the generic form:

$$\boxed{\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu\, \boldsymbol{u}_i^*\, g[\boldsymbol{e}(i)], \quad \boldsymbol{w}_{-1} = \text{initial condition}} \tag{15.23}$$

where $g[\cdot]$ denotes some function of the *a priori* output error signal,

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

Updates of this form are said to correspond to filters with *error nonlinearities*.

ENERGY CONSERVATION

We can also study update equations with *data* (as opposed to error) nonlinearities, say of the form

$$w_i = w_{i-1} + \mu\, g[u_i]u_i^* e(i), \qquad w_{-1} = \text{initial condition}$$

for some positive function $g[\cdot]$ of the regression data, $g[u_i] > 0$ (the function $g[\cdot]$ could also be matrix-valued).

TABLE 15.2 Examples of error functions for several adaptive algorithms: ϵ is a small positive number, $0 \leq \delta \leq 1$, and $p(i)$ is a positive quantity whose computation will be explained later.

Algorithm	Error function
LMS	$g[e(i)] = e(i)$
ϵ-NLMS	$g[e(i)] = e(i)/(\epsilon + \|u_i\|^2)$
ϵ-NLMS with power normalization	$g[e(i)] = e(i)/(\epsilon + p(i))$
LMF	$g[e(i)] = e(i)\|e(i)\|^2$
LMMN	$g[e(i)] = e(i)\left[\delta + (1-\delta)\|e(i)\|^2\right]$
sign-error LMS	$g[e(i)] = \text{csgn}[e(i)]$

ERROR RELATIONS

The update recursion (15.23) can be rewritten in terms of the weight-error vector

$$\tilde{\boldsymbol{w}}_i = w^o - \boldsymbol{w}_i$$

Subtracting both sides of (15.23) from w^o we get

$$w^o - \boldsymbol{w}_i = w^o - \boldsymbol{w}_{i-1} - \mu\, \boldsymbol{u}_i^*\, g[\boldsymbol{e}(i)], \qquad \tilde{\boldsymbol{w}}_{-1} = \text{initial condition}$$

or, equivalently,

$$\boxed{\tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i-1} - \mu \boldsymbol{u}_i^* g[\boldsymbol{e}(i)]} \qquad (15.24)$$

In addition, if we multiply both sides of (15.24) by $\boldsymbol{u}_i$ from the left we find that the *a priori* and *a posteriori* estimation errors $\{\boldsymbol{e}_a(i), \boldsymbol{e}_p(i)\}$ are related via:

$$\boxed{\boldsymbol{e}_p(i) = \boldsymbol{e}_a(i) - \mu \|\boldsymbol{u}_i\|^2 g[\boldsymbol{e}(i)]} \qquad (15.25)$$

where $\{\boldsymbol{e}_p(i), \boldsymbol{e}_a(i)\}$ were defined in Table 15.1 as

$$\boxed{\boldsymbol{e}_a(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1}, \qquad \boldsymbol{e}_p(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_i} \qquad (15.26)$$

PERFORMANCE METRICS

Expressions (15.24)–(15.25) provide an alternative description of the adaptive filter (15.23) in terms of the error quantities $\{e_a(i), e_p(i), \tilde{w}_i, \tilde{w}_{i-1}, e(i)\}$. This description is useful since we are often interested in questions related to the behavior of these errors, such as:

1. Steady-state behavior, which relates to determining the steady-state values of $\mathsf{E}\,\|\tilde{w}_i\|^2$, $\mathsf{E}\,|e_a(i)|^2$, and $\mathsf{E}\,|e(i)|^2$.

2. Stability, which relates to determining the range of values of the step-size μ over which the variances $\mathsf{E}\,|e_a(i)|^2$ and $\mathsf{E}\,\|\tilde{w}_i\|^2$ remain bounded.

3. Transient behavior, which relates to studying the time evolution of the curves $\mathsf{E}\,|e_a(i)|^2$ and $\{\mathsf{E}\,\tilde{w}_i, \mathsf{E}\,\|\tilde{w}_i\|^2\}$.

In order to address questions of this kind, we shall rely on an energy equality that relates the squared norms of the errors $\{e_a(i), e_p(i), \tilde{w}_{i-1}, \tilde{w}_i\}$.

ALGEBRAIC DERIVATION

To derive the energy relation, we first combine (15.24)–(15.25) to eliminate the error nonlinearity $g[\cdot]$ from (15.24), i.e., we solve for $g[\cdot]$ from (15.25) and then substitute into (15.24), as done below. What this initial step means is that the resulting energy relation will hold irrespective of the error nonlinearity. We distinguish between two cases:

1. $\underline{\boldsymbol{u}_i = 0}$. This is a degenerate situation. In this case, it is obvious from (15.24) and (15.25) that $\tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i-1}$ and $\boldsymbol{e}_a(i) = \boldsymbol{e}_p(i)$ so that

$$\|\tilde{\boldsymbol{w}}_i\|^2 = \|\tilde{\boldsymbol{w}}_{i-1}\|^2 \quad \text{and} \quad |\boldsymbol{e}_a(i)|^2 = |\boldsymbol{e}_p(i)|^2 \tag{15.27}$$

2. $\underline{\boldsymbol{u}_i \neq 0}$. In this case, we use (15.25) to solve for $g[\boldsymbol{e}(i)]$,

$$g[\boldsymbol{e}(i)] = \frac{1}{\mu\|\boldsymbol{u}_i\|^2}[\boldsymbol{e}_a(i) - \boldsymbol{e}_p(i)]$$

and substitute into (15.24) to obtain

$$\tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i-1} - \frac{\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2}[\boldsymbol{e}_a(i) - \boldsymbol{e}_p(i)] \tag{15.28}$$

ALGEBRAIC DERIVATION

This relation involves the four errors $\{\tilde{\boldsymbol{w}}_i, \tilde{\boldsymbol{w}}_{i-1}, \boldsymbol{e}_a(i), \boldsymbol{e}_p(i)\}$; observe that even the step-size μ is cancelled out. Expression (15.28) can be rearranged as

$$\tilde{\boldsymbol{w}}_i + \frac{\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2}\boldsymbol{e}_a(i) \;=\; \tilde{\boldsymbol{w}}_{i-1} + \frac{\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2}\boldsymbol{e}_p(i) \tag{15.29}$$

On each side of this identity we have a combination of *a priori* and *a posteriori* errors. By evaluating the energies (i.e., the squared Euclidean norms) of both sides we find, after a straightforward calculation, that the following energy equality holds:

$$\|\tilde{\boldsymbol{w}}_i\|^2 + \frac{1}{\|\boldsymbol{u}_i\|^2}|\boldsymbol{e}_a(i)|^2 = \|\tilde{\boldsymbol{w}}_{i-1}\|^2 + \frac{1}{\|\boldsymbol{u}_i\|^2}|\boldsymbol{e}_p(i)|^2 \tag{15.30}$$

Interesting enough, this equality simply amounts to adding the energies of the individual terms of (15.29); the cross-terms cancel out. This is one advantage of working with the energy relation (15.30): irrelevant cross-terms are eliminated so that one does not need to worry later about evaluating their expectations.

ALGEBRAIC DERIVATION

The results in both cases of zero and nonzero regression vectors can be combined together by using a common notation. Define $\bar{\mu}(i) = \left(\|\boldsymbol{u}_i\|^2\right)^\dagger$, in terms of the pseudo-inverse operation. Recall that the pseudo-inverse of a nonzero scalar is equal to its inverse value, while the pseudo-inverse of zero is equal to zero. That is,

$$\boxed{\bar{\mu}(i) \triangleq \begin{cases} 1/\|\boldsymbol{u}_i\|^2 & \text{if } \boldsymbol{u}_i \neq 0 \\ 0 & \text{otherwise} \end{cases}} \tag{15.31}$$

Using $\bar{\mu}(i)$, we can combine (15.27) and (15.30) into a single identity as

$$\boxed{\|\tilde{\boldsymbol{w}}_i\|^2 + \bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \|\tilde{\boldsymbol{w}}_{i-1}\|^2 + \bar{\mu}(i)|\boldsymbol{e}_p(i)|^2} \tag{15.32}$$

We can alternatively express (15.32) as

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

ENERGY CONSERVATION RELATION

Theorem 15.1 (Energy conservation relation) For adaptive filters of the form (15.23), and for any data $\{d(i), \boldsymbol{u}_i\}$, it always holds that

$$\|\tilde{\boldsymbol{w}}_i\|^2 + \bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \|\tilde{\boldsymbol{w}}_{i-1}\|^2 + \bar{\mu}(i)|\boldsymbol{e}_p(i)|^2$$

where $\boldsymbol{e}_a(i) = \boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}$, $\boldsymbol{e}_p(i) = \boldsymbol{u}_i\tilde{\boldsymbol{w}}_i$, $\tilde{\boldsymbol{w}}_i = w^o - \boldsymbol{w}_i$, and $\bar{\mu}(i)$ is defined as in (15.31).

The important fact to emphasize here is that *no approximations* have been used to establish the energy relation (15.32); it is an exact relation 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.

15.4 VARIANCE RELATION

Relation (15.32) has important ramifications in the study of adaptive filters. In this chapter, and the remaining chapters of this part, we shall focus on its significance to the steady-state performance, tracking analysis, and finite-precision analysis of adaptive filters. In Part V (*Transient Performance*) we shall apply it to transient analysis, and in Part XI (*Robust Filters*) we shall examine its significance to robustness analysis. In the course of these discussions, it will become clear that the energy-conservation relation (15.32) provides a unifying framework for the performance analysis of adaptive filters.

With regards to steady-state performance, which is the subject matter of this chapter, it has been common in the literature to study the steady-state performance of an adaptive filter as the limiting behavior of its transient performance (which is concerned with the study of the time evolution of $\mathsf{E}\|\tilde{w}_i\|^2$). As we shall see in Part V (*Transient Performance*), transient analysis is a more demanding task to pursue and it tends to require a handful of additional assumptions and restrictions on the data. In this way, steady-state results that are obtained as the limiting behavior of a transient analysis would be governed by the same restrictions on the data. In our treatment, on the other hand, we separate the study of the steady-state performance of an adaptive filter from the study of its transient performance. In so doing, it becomes possible to pursue the steady-state analysis in several instances under weaker assumptions than those required by a full blown transient analysis.

Steady-State Filter Operation

To initiate our steady-state performance studies, we first explain what is meant by an adaptive filter operating in *steady-state*.

Definition 15.1 (Steady-state filter) An adaptive filter will be said to operate in steady-state if it holds that

$$\mathsf{E}\,\tilde{w}_i \longrightarrow s, \quad \text{as } i \to \infty \tag{15.33}$$

$$\mathsf{E}\,\tilde{w}_i\tilde{w}_i^* \longrightarrow C, \quad \text{as } i \to \infty \tag{15.34}$$

where s and C are some finite constants (usually $s = 0$).

Although, strictly speaking, the first and second-order moments of the error vector need not tend to constant limit values, the definition is an acceptable approximation for small step-sizes. In particular, it follows that

$$\mathsf{E}\,\|\tilde{w}_i\|^2 = \mathsf{E}\,\|\tilde{w}_{i-1}\|^2 = c, \quad \text{as } i \to \infty \tag{15.35}$$

where $c = \mathsf{Tr}(C)$.

Variance Relation for Steady-State Performance

In order to explain how (15.32) is useful in evaluating the steady-state performance of an adaptive filter, we recall from (15.21) that we are interested in evaluating the steady-state variance of $e_a(i)$. Now taking expectations of both sides of (15.32) we get

$$\boxed{\mathsf{E}\,\|\tilde{w}_i\|^2 + \mathsf{E}\,\bar{\mu}(i)|e_a(i)|^2 = \mathsf{E}\,\|\tilde{w}_{i-1}\|^2 + \mathsf{E}\,\bar{\mu}(i)|e_p(i)|^2} \qquad (15.36)$$

where the expectation is with respect to the distributions of the random variables $\{d(i), u_i\}$. Taking the limit of (15.36) as $i \to \infty$ and using the steady-state condition (15.35), we obtain

$$\mathsf{E}\,\bar{\mu}(i)|e_a(i)|^2 = \mathsf{E}\,\bar{\mu}(i)|e_p(i)|^2, \quad \text{as } i \to \infty \qquad (15.37)$$

This equality is in terms of $\{e_a(i), e_p(i)\}$. However, from (15.25) we know how $e_p(i)$ is related to $e_a(i)$. Substituting into (15.37) we get

$$\mathsf{E}\,\bar{\mu}(i)|e_a(i)|^2 = \mathsf{E}\,\bar{\mu}(i)\left|e_a(i) - \mu\|u_i\|^2 g[e(i)]\right|^2, \qquad \text{as } i \to \infty \qquad (15.38)$$

STEADY-STATE PERFORMANCE

Expanding the term on the right-hand side and simplifying leads to (we are omitting the argument of g for compactness of notation):

$$\begin{aligned}\bar{\mu}(i)\left|\boldsymbol{e}_a(i)-\mu\|\boldsymbol{u}_i\|^2 g\right|^2 &= \bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 + \mu^2\|\boldsymbol{u}_i\|^2|g|^2 - \mu\boldsymbol{e}_a(i)g^* - \mu\boldsymbol{e}_a^*(i)g \\ &= \bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 + \mu^2\|\boldsymbol{u}_i\|^2|g|^2 - 2\mu\mathrm{Re}\left(\boldsymbol{e}_a^*(i)g\right) \quad (15.39)\end{aligned}$$

where in the first equality we used the fact that

$$\bar{\mu}(i)\|\boldsymbol{u}_i\|^4 = \|\boldsymbol{u}_i\|^2 \quad \text{and} \quad \bar{\mu}(i)\|\boldsymbol{u}_i\|^2\boldsymbol{e}_a(i)g^* = \boldsymbol{e}_a(i)g^*$$

for all $\boldsymbol{u}_i$ (whether $\boldsymbol{u}_i = 0$ or otherwise). Taking expectation of the right-hand side of (15.39) and substituting into (15.38) we obtain

$$\boxed{\mu\mathsf{E}\left(\|\boldsymbol{u}_i\|^2\cdot|g[\boldsymbol{e}(i)]|^2\right) = 2\mathrm{Re}\left(\mathsf{E}\,\boldsymbol{e}_a^*(i)g[\boldsymbol{e}(i)]\right), \quad \text{as } i\to\infty} \qquad (15.40)$$

in terms of the real part of $\boldsymbol{e}_a^*(i)g[\boldsymbol{e}(i)]$.

VARIANCE RELATION

Theorem 15.2 (Variance relation) For adaptive filters of the form (15.23) and for any data $\{d(i), u_i\}$, assuming filter operation in steady-state, the following relation holds:

$$\mu \mathsf{E}\left(\|u_i\|^2 \cdot |g[e(i)]|^2\right) = 2\mathsf{Re}\left(\mathsf{E}\, e_a^*(i) g[e(i)]\right), \quad \text{as } i \to \infty$$

For real-valued data, this variance relation becomes

$$\mu \mathsf{E}\left(\|u_i\|^2 \cdot g^2[e(i)]\right) = 2\mathsf{E}\, e_a(i) g[e(i)], \quad \text{as } i \to \infty$$

We remark again that the variance relation (15.40) is *exact*, since it holds without any approximations or assumptions (except for the assumption that the filter is operating in steady-state, which is necessary if one is interested in evaluating the steady-state performance of a filter). We refer to (15.40) as a variance relation since it will be our starting point for evaluating the variance $\mathsf{E}\,|e_a(i)|^2$ for different adaptive filters. The results of both Thms. 15.1 and 15.2 do *not* require the analysis model (15.16); they hold for general data $\{d(i), u_i\}$.

MEAN-SQUARE ANALYSIS

Relevance to Mean-Square Performance Analysis

However, when model (15.16) is assumed, then we know from (15.19) that $e(i)$ can be expressed in terms of $e_a(i)$ as

$$e(i) = e_a(i) + v(i)$$

In this way, relation (15.40) becomes an identity involving $e_a(i)$ and, in principle, it could be solved to evaluate the EMSE of an adaptive filter, i.e., to compute $\mathsf{E}\,|e_a(\infty)|^2$. We say "in principle" because, although (15.40) is an exact result, different choices for the error function $g[\cdot]$ can make the solution for $\mathsf{E}\,|e_a(\infty)|^2$ easier for some cases than others. It is at this stage that simplifying assumptions become necessary. We shall illustrate this point for several adaptive filters in the sections that follow.

MEAN-SQUARE ANALYSIS

In order to simplify the notation, we shall employ the symbol ζ to refer to the EMSE of an adaptive filter, i.e.,

$$\zeta = \mathsf{E}\,|e_a(\infty)|^2$$

For example, the EMSE of LMS will be denoted by ζ^{LMS}. Its misadjustment will be denoted by $\mathcal{M}^{\mathsf{LMS}}$. Similar notation will be used for other algorithms. In view of the analysis model (15.16), which enables us to identify $J_{\min}$ as σ_v^2, we obtain from (15.11) that the misadjustment of an adaptive filter is related to its EMSE via

$$\boxed{\mathcal{M} = \mathsf{EMSE}/\sigma_v^2}$$

We limit our derivations in the sequel to determining expressions for the EMSE of several adaptive filters. Expressions for the misadjustment would follow by dividing the result by σ_v^2.

16.1 VARIANCE RELATION

Thus assume that the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy model (15.16) and consider the LMS recursion

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* \boldsymbol{e}(i) \tag{16.1}$$

for which

$$g[\boldsymbol{e}(i)] = \boldsymbol{e}(i) = \boldsymbol{e}_a(i) + \boldsymbol{v}(i) \tag{16.2}$$

Relation (15.40) then becomes

$$\mu \mathsf{E}\, \|\boldsymbol{u}_i\|^2 |\boldsymbol{e}_a(i) + \boldsymbol{v}(i)|^2 = 2\mathsf{Re}\big(\mathsf{E}\, \boldsymbol{e}_a^*(i)[\boldsymbol{e}_a(i) + \boldsymbol{v}(i)]\big), \quad i \to \infty \tag{16.3}$$

Several terms in this equality get cancelled. We shall carry out the calculations rather slowly in this section for illustration purposes only. Later, when similar calculations are called upon, we shall be less detailed.

VARIANCE RELATION

To begin with, the expression on the left-hand side of (16.3) expands to

$$\begin{aligned}\mu\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)+\boldsymbol{v}(i)|^2 &= \mu\mathsf{E}\,\|\boldsymbol{u}_i\|^2\left[\,|\boldsymbol{e}_a(i)|^2+|\boldsymbol{v}(i)|^2+\boldsymbol{e}_a^*(i)\boldsymbol{v}(i)+\boldsymbol{e}_a(i)\boldsymbol{v}^*(i)\,\right]\\ &= \mu\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \;+\; \mu\sigma_v^2\mathsf{E}\,\|\boldsymbol{u}_i\|^2\\ &= \mu\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \;+\; \mu\sigma_v^2\mathsf{Tr}(R_u) \qquad (16.4)\end{aligned}$$

where we used the fact that $\boldsymbol{v}(i)$ is independent of both $\boldsymbol{u}_i$ and $\boldsymbol{e}_a(i)$ (recall Lemma 15.1), so that the cross-terms involving $\{\boldsymbol{v}(i),\boldsymbol{e}_a(i),\boldsymbol{u}_i\}$ cancel out. We also used the fact that

$$\mathsf{E}\,\|\boldsymbol{u}_i\|^2=\mathsf{Tr}(R_u) \quad \text{and} \quad \mathsf{E}\,|\boldsymbol{v}(i)|^2=\sigma_v^2$$

Similarly, the expression on the right-hand side of (16.3) simplifies to $2\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$, which is simply $2\zeta^{\mathsf{LMS}}$ as $i\rightarrow\infty$. Therefore, equality (16.3) amounts to

$$\boxed{\zeta^{\mathsf{LMS}}=\frac{\mu}{2}\left[\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \;+\; \sigma_v^2\mathsf{Tr}(R_u)\right], \quad \text{as } i\rightarrow\infty} \qquad (16.5)$$

This expression has been arrived at without approximations. Still, it requires that we evaluate the steady-state value of the expectation $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ in order to arrive at the EMSE

$$\zeta^{\text{LMS}} = \frac{\mu}{2}\left[\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \;+\; \sigma_v^2 \mathsf{Tr}(R_u)\right], \quad \text{as } i \to \infty \tag{16.5}$$

16.2 SMALL STEP-SIZES

Expression (16.5) suggests that small step-sizes lead to small $\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$ in steady-state and, consequently, to a high likelihood of small values for $\boldsymbol{e}_a(i)$ itself. So assume μ is small enough so that, in *steady-state*, the contribution of the term $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ can be neglected, say

$$\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \;\ll\; \sigma_v^2 \mathsf{Tr}(R_u)$$

Then, we find from (16.5) that the EMSE can be approximated by

$$\zeta^{\text{LMS}} = \frac{\mu\sigma_v^2 \mathsf{Tr}(R_u)}{2} \qquad \text{(for sufficiently small } \mu) \tag{16.6}$$

16.3 SEPARATION PRINCIPLE

If the step-size is not sufficiently small, but still small enough to guarantee filter convergence — as will be discussed in Chapter 24, we can derive an alternative approximation for the EMSE from (16.5); the resulting expression will hold over a wider range of step-sizes. To do so, here and in several other places in this chapter and in subsequent chapters, we shall rely on the following assumption:

$$\boxed{\text{At steady-state, } \quad \|\boldsymbol{u}_i\|^2 \quad \text{is independent of} \quad \boldsymbol{e}_a(i)} \tag{16.7}$$

We shall refer to this condition as the *separation* assumption or the separation principle. Alternatively, we could assume instead that

$$\boxed{\text{At steady-state, } \quad \|\boldsymbol{u}_i\|^2 \quad \text{is independent of} \quad \boldsymbol{e}(i)} \tag{16.8}$$

with $\boldsymbol{e}_a(i)$ replaced by $\boldsymbol{e}(i)$. This condition is equivalent to (16.7) since $\boldsymbol{e}(i) = \boldsymbol{e}_a(i) + \boldsymbol{v}(i)$ and $\|\boldsymbol{u}_i\|^2$ is independent of $\boldsymbol{v}(i)$ (as follows from Lemma 15.1).

SEPARATION PRINCIPLE

$$\boxed{\text{At steady-state, } \|u_i\|^2 \text{ is independent of } e_a(i)} \qquad (16.7)$$

Of course, assumption (16.7) is only exact in some special cases, e.g., when the successive regressors have constant Euclidean norms, since then $\|u_i\|^2$ becomes a constant; this situation occurs when the entries of u_i arise from a finite alphabet with constant magnitude — see Prob. IV.2. More generally, the assumption is reasonable at *steady-state* since the behavior of $e_a(i)$ in the limit is likely to be less sensitive to the regression (input) data.

Assumption (16.7) allows us to separate the expectation $\mathsf{E}\,\|u_i\|^2|e_a(i)|^2$, which appears in (16.5), into the product of two expectations:

$$\mathsf{E}\left(\|u_i\|^2 \cdot |e_a(i)|^2\right) = \left(\mathsf{E}\,\|u_i\|^2\right) \cdot \left(\mathsf{E}\,|e_a(i)|^2\right) = \mathsf{Tr}(R_u)\zeta^{\mathsf{LMS}}, \quad i \to \infty \qquad (16.9)$$

In order to illustrate this approximation, we show in Fig. 16.1 the result of simulating a 20-tap LMS filter over 1000 experiments. The figure shows the ensemble-averaged curves that correspond to the quantities

$$\mathsf{E}\left(\|u_i\|^2 \cdot |e_a(i)|^2\right) \qquad \text{and} \qquad \left(\mathsf{E}\,\|u_i\|^2\right) \cdot \left(\mathsf{E}\,|e_a(i)|^2\right)$$

SIMULATION

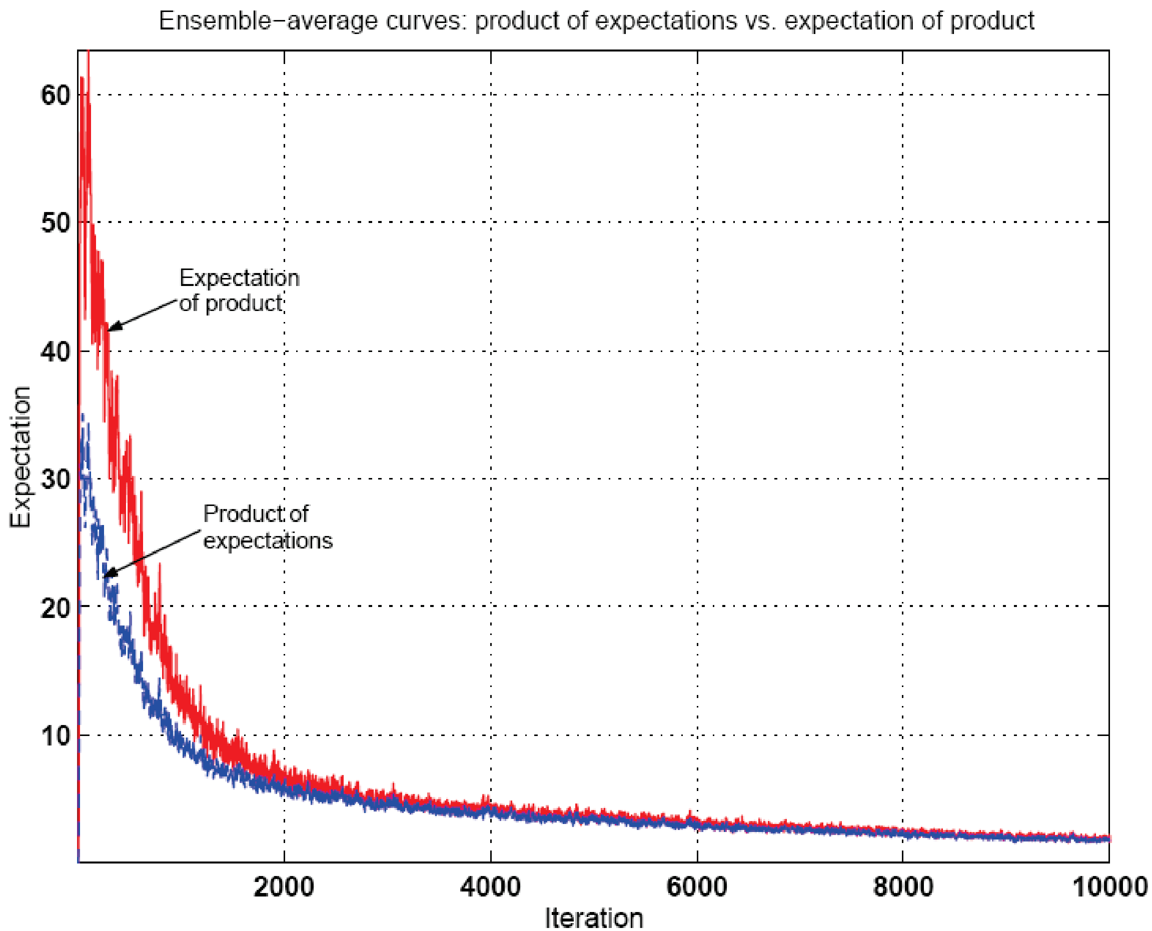

FIGURE 16.1 Ensemble-average curves for the expectation of the product, $\mathsf{E}\left(\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2\right)$ (*upper curve*), and for the product of expectations, $\left(\mathsf{E}\,\|\boldsymbol{u}_i\|^2\right)\cdot\left(\mathsf{E}\,|\boldsymbol{e}_a(i)|^2\right)$ (*lower curve*), for a 20-tap LMS filter with step-size $\mu = 0.001$. The curves are obtained by averaging over 1000 experiments.

SEPARATION PRINCIPLE

Now substituting (16.9) into (16.5) leads to the following expression for the EMSE of LMS:

$$\boxed{\zeta^{\text{LMS}} = \frac{\mu\sigma_v^2 \text{Tr}(R_u)}{2 - \mu \text{Tr}(R_u)}} \qquad \text{(over a wider range of } \mu\text{)} \qquad (16.10)$$

This result will be revisited in Chapter 23; see the discussion following the statement of Thm. 23.3.

16.4 WHITE GAUSSIAN INPUT

One particular case for which the term $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ that appears in (16.5) can be evaluated in closed-form occurs when $\boldsymbol{u}_i$ has a circular Gaussian distribution with a diagonal covariance matrix, say

$$R_u = \sigma_u^2 \mathrm{I}, \qquad \sigma_u^2 > 0 \tag{16.11}$$

That is, when the probability density function of $\boldsymbol{u}_i$ is of the form (cf. Lemma A.1):

$$f_{\boldsymbol{u}}(u) = \frac{1}{\pi^M} \frac{1}{\det R_u} \exp\left\{-uR_u^{-1}u^*\right\} = \frac{1}{(\pi\sigma_u^2)^M} \exp\left\{-\|u\|^2/\sigma_u^2\right\}$$

The diagonal structure of R_u amounts to saying that the entries of $\boldsymbol{u}_i$ are uncorrelated among themselves and that each has variance σ_u^2. The analysis can still be carried out in closed form even without this whiteness assumption; it suffices to require the regressors to be Gaussian. Moreover, R_u does not need to be a scaled multiple of the identity. We treat this more general situation in Sec. 23.1; see Prob. IV.19 for further motivation and the discussion following Thm. 23.3.

STEADY-STATE ASSUMPTION

In addition to (16.11), we shall assume in this subsection that

$$\boxed{\text{At steady state, } \tilde{\boldsymbol{w}}_{i-1} \text{ is independent of } \boldsymbol{u}_i} \tag{16.12}$$

Conditions (16.11) and (16.12) enable us to evaluate $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ explicitly. Before doing so, however, it is worth pointing out that to perform this task, it has been common in the literature to rely not on (16.12) but instead on a set of conditions known collectively as the *independence assumptions*.

INDEPENDENCE ASSUMPTIONS

(i) The sequence $\{\boldsymbol{d}(i)\}$ is i.i.d.

(ii) The sequence $\{\boldsymbol{u}_i\}$ is also i.i.d.

(iii) Each $\boldsymbol{u}_i$ is independent of previous $\{\boldsymbol{d}_j, j < i\}$.

(iv) Each $\boldsymbol{d}(i)$ is independent of previous $\{\boldsymbol{u}_j, j < i\}$.

(v) The $\boldsymbol{d}(i)$ and $\boldsymbol{u}_i$ are jointly Gaussian.

(vi) In the case of complex-valued data, the $\boldsymbol{d}(i)$ and $\boldsymbol{u}_i$ are individually and jointly circular random variables, i.e., they satisfy $\mathsf{E}\,\boldsymbol{u}_i^{\mathsf{T}}\boldsymbol{u}_i = 0$, $\mathsf{E}\,\boldsymbol{d}^2(i) = 0$, and $\mathsf{E}\,\boldsymbol{u}_i^{\mathsf{T}}\boldsymbol{d}(i) = 0$.

The independence assumptions (i)–(vi) are in general restrictive since, in practice, the sequence $\{\boldsymbol{u}_i\}$ is rarely i.i.d. Consider, for example, the case in which the regressors $\{\boldsymbol{u}_i\}$ correspond to state vectors of an FIR implementation, as in the channel estimation application of Sec. 10.5. In this case, two successive regressors share common entries and cannot be statistically independent. Still, when the step-size is sufficiently small, the conclusions that are obtained under the independence assumptions (i)–(vi) tend to be realistic — see App. 24.A.

INDEPENDENCE ASSUMPTIONS

Condition (16.12) is less restrictive than the independence assumptions (i)–(vi). Actually, assumption (16.12) is implied by the independence conditions. To see this, recall from the discussion in Sec. 15.2 that $\tilde{\boldsymbol{w}}_{i-1}$ is a function of the variables $\{\boldsymbol{w}_{-1}; \boldsymbol{d}(i-1), \ldots, \boldsymbol{d}(0); \boldsymbol{u}_{i-1}, \ldots, \boldsymbol{u}_0\}$. Therefore, if the sequence $\boldsymbol{u}_i$ is assumed i.i.d., and if $\boldsymbol{u}_i$ is independent of all previous $\{\boldsymbol{d}(j)\}$ and of $\boldsymbol{w}_{-1}$, then $\boldsymbol{u}_i$ will be independent of $\tilde{\boldsymbol{w}}_{i-1}$ *for all* i. Note, in addition, that condition (16.12) is only requiring the independence of $\{\tilde{\boldsymbol{w}}_{i-1}, \boldsymbol{u}_i\}$ to hold in *steady-state*; which is a considerably weaker assumption than what is implied by the full blown independence assumptions (i)–(vi). Moreover, assumption (16.12) is reasonable for small step-sizes μ. Intuitively, this is because the update term in (15.24) is relatively small for small μ and the statistical dependence of $\tilde{\boldsymbol{w}}_{i-1}$ on $\boldsymbol{u}_i$ becomes weak. Furthermore, in steady-state, the error $\boldsymbol{e}(i)$ is also small, which makes the update term in (15.24) even smaller.

MEAN-SQUARE PERFORMANCE

$$\boxed{\zeta^{\mathsf{LMS}} = \frac{\mu}{2}\left[\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \;+\; \sigma_v^2\mathsf{Tr}(R_u)\right], \quad \text{as } i\to\infty} \qquad (16.5)$$

So let us return to the term $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ in (16.5) and show how it can be evaluated under (16.12), and under the assumption of circular Gaussian regressors. First we show how to express $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ in terms of $\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$ (see (16.17) further ahead). Thus note the following sequence of identities:

$$\begin{aligned}
\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 &= \mathsf{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i^*\left(\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\right)\right)\\
&= \mathsf{E}\;\mathsf{Tr}\left(\boldsymbol{u}_i\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\right)\\
&= \mathsf{E}\;\mathsf{Tr}\left(\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\boldsymbol{u}_i\right)\\
&= \mathsf{Tr}\;\mathsf{E}\left(\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\boldsymbol{u}_i\right) \qquad (16.13)
\end{aligned}$$

where in the second equality we used the fact that the trace of a scalar is equal to the scalar itself, and in the third equality we used the property that $\mathsf{Tr}(AB)=\mathsf{Tr}(BA)$ for any matrices A and B of compatible dimensions.

MEAN-SQUARE PERFORMANCE

We now evaluate the term $\mathsf{E}\left(\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\boldsymbol{u}_i\right)$, which is a covariance matrix. To do so, we recall the following property of conditional expectations, namely, that for any two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$, it holds that $\mathsf{E}\,\boldsymbol{x} = \mathsf{E}\left(\mathsf{E}\,[\boldsymbol{x}|\boldsymbol{y}]\right)$ — see (1.4). Therefore, in steady-state,

$$\begin{aligned}
\mathsf{E}\left(\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\boldsymbol{u}_i\right) &= \mathsf{E}\left[\mathsf{E}\left(\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\boldsymbol{u}_i \mid \boldsymbol{u}_i\right)\right] \\
&= \mathsf{E}\left[\boldsymbol{u}_i^*\boldsymbol{u}_i\,\mathsf{E}\left(\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^* \mid \boldsymbol{u}_i\right)\boldsymbol{u}_i^*\boldsymbol{u}_i\right] \\
&= \mathsf{E}\left(\boldsymbol{u}_i^*\boldsymbol{u}_i C_{i-1}\boldsymbol{u}_i^*\boldsymbol{u}_i\right) \qquad (16.14)
\end{aligned}$$

where in the last step we used assumption (16.12), namely, that $\tilde{\boldsymbol{w}}_{i-1}$ and $\boldsymbol{u}_i$ are independent so that

$$\mathsf{E}\left(\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^* \mid \boldsymbol{u}_i\right) = \mathsf{E}\,\tilde{\boldsymbol{w}}_{i-1}\tilde{\boldsymbol{w}}_{i-1}^* \stackrel{\Delta}{=} C_{i-1}$$

We are also denoting the covariance matrix of $\tilde{\boldsymbol{w}}_{i-1}$ by C_{i-1}. We do not need to know the value of C_{i-1}, as the argument will demonstrate — see the remark following (16.17). We are then reduced to evaluating the expression $\mathsf{E}\,\boldsymbol{u}_i^*\boldsymbol{u}_i C_{i-1}\boldsymbol{u}_i^*\boldsymbol{u}_i$.

Due to the circular Gaussian assumption on $\boldsymbol{u}_i$, this term has the same form as the general term that we evaluated earlier in Lemma A.3 for Gaussian variables, with the identifications

$$z \leftarrow \boldsymbol{u}_i^*, \quad W \leftarrow C_{i-1}, \quad \Lambda \leftarrow \sigma_u^2 \mathrm{I}$$

so that we can use the result of that lemma to write

$$\mathsf{E}\,\boldsymbol{u}_i^* \boldsymbol{u}_i C_{i-1} \boldsymbol{u}_i^* \boldsymbol{u}_i \;=\; \sigma_u^4 \left[\mathsf{Tr}(C_{i-1})\mathrm{I} + C_{i-1}\right] \tag{16.15}$$

Substituting this equality into (16.13) we obtain

$$\mathsf{E}\,\|\boldsymbol{u}_i\|^2 |\boldsymbol{e}_a(i)|^2 \;=\; \mathsf{Tr}\left[\sigma_u^4 \mathsf{Tr}(C_{i-1})\mathrm{I} + \sigma_u^4 C_{i-1}\right] \;=\; (M+1)\sigma_u^4 \mathsf{Tr}(C_{i-1}) \tag{16.16}$$

Now repeating the same argument that led to (16.14), we also find that

$$\begin{aligned}
\mathsf{E}\,|\boldsymbol{e}_a(i)|^2 = \mathsf{E}\left(\boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1} \tilde{\boldsymbol{w}}_{i-1}^* \boldsymbol{u}_i^*\right) \;&=\; \mathsf{E}\left(\boldsymbol{u}_i C_{i-1} \boldsymbol{u}_i^*\right) \\
&=\; \mathsf{E}\left(\boldsymbol{u}_i C_{i-1} \boldsymbol{u}_i^*\right) \\
&=\; \mathsf{Tr}\;\mathsf{E}\left(\boldsymbol{u}_i^* \boldsymbol{u}_i C_{i-1}\right) \\
&=\; \mathsf{Tr}\;\mathsf{E}\left(R_u C_{i-1}\right) \\
&=\; \sigma_u^2 \mathsf{Tr}\;\mathsf{E}\left(C_{i-1}\right)
\end{aligned}$$

COMPLEX DATA

This expression relates $\text{Tr}(C_{i-1})$ to $\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$. Substituting into (16.16) we obtain

$$\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \;=\; (M+1)\sigma_u^2\mathsf{E}\,|\boldsymbol{e}_a(i)|^2 \qquad (16.17)$$

This relation expresses the desired term $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ as a scaled multiple of $\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$ alone — observe that C_{i-1} is cancelled out. Using this result in (16.5), we get

$$\boxed{\zeta^{\text{LMS}} = \frac{\mu M\sigma_v^2\sigma_u^2}{2-\mu(M+1)\sigma_u^2} \quad \text{(for complex-valued data)}} \qquad (16.18)$$

REAL DATA

The above derivation assumes complex-valued data. If the data were real-valued, then the same arguments would still apply with the only exception of Lemma A.3. Instead, we would employ the result of Lemma A.2 and replace (16.15) by

$$\mathsf{E}\, \boldsymbol{u}_i^{\mathsf{T}} \boldsymbol{u}_i C_{i-1} \boldsymbol{u}_i^{\mathsf{T}} \boldsymbol{u}_i \;=\; \sigma_u^4 \left[\mathsf{Tr}(C_{i-1}) \mathrm{I} + 2C_{i-1} \right]$$

with an additional scaling factor of 2 (now $C_{i-1} = \mathsf{E}\, \tilde{\boldsymbol{w}}_{i-1} \tilde{\boldsymbol{w}}_{i-1}^{\mathsf{T}}$). Then (16.16) and (16.17) would become

$$\mathsf{E}\, \|\boldsymbol{u}_i\|^2 \boldsymbol{e}_a^2(i) = (M+2)\sigma_u^4 \mathsf{Tr}(C_{i-1}) = (M+2)\sigma_u^2 \mathsf{E}\, \boldsymbol{e}_a^2(i) \qquad (16.19)$$

and the resulting expression for the EMSE is

$$\boxed{\zeta^{\mathsf{LMS}} = \frac{\mu M \sigma_v^2 \sigma_u^2}{2 - \mu(M+2)\sigma_u^2} \quad \text{(for real-valued data)}} \qquad (16.20)$$

RESULTS

Lemma 16.1 (EMSE of LMS) Consider the LMS recursion (16.1) and assume the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy model (15.16). Then its EMSE can be approximated by the following expressions:

1. For sufficiently small step-sizes, it holds that $\zeta^{\text{LMS}} = \mu\sigma_v^2 \text{Tr}(R_u)/2$.
2. Under the separation assumption (16.7), it holds that

$$\zeta^{\text{LMS}} = \mu\sigma_v^2 \text{Tr}(R_u)/[2 - \mu \text{Tr}(R_u)]$$

3. If $\boldsymbol{u}_i$ is Gaussian with $R_u = \sigma_u^2 \text{I}$, and under the steady-state assumption (16.12), it holds that

$$\zeta^{\text{LMS}} = \mu M \sigma_v^2 \sigma_u^2 / [2 - \mu(M + \gamma)\sigma_u^2]$$

   where $\gamma = 2$ if the data is real-valued and $\gamma = 1$ if the data is complex-valued and $\boldsymbol{u}_i$ circular. Here M is the dimension of $\boldsymbol{u}_i$.

In all cases, the misadjustment is obtained by dividing the EMSE by σ_v^2.

16.6 SIMULATION RESULTS

Figures 16.2–16.4 show the values of the steady-state MSE of a 10-tap LMS filter for different choices of the step-size and for different signal conditions. The theoretical values are obtained by using the expressions from Lemma 16.1. For each step-size, the experimental value is obtained by running LMS for 4×10^5 iterations and averaging the squared-error curve $\{|e(i)|^2\}$ over 100 experiments in order to generate the ensemble-average curve. The average of the last 5000 entries of the ensemble-average curve is then used as the experimental value for the MSE. The data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ are generated according to model (15.16) using Gaussian noise with variance $\sigma_v^2 = 0.001$.

In Fig. 16.2, the regressors $\{\boldsymbol{u}_i\}$ do not have shift structure (i.e., they do not correspond to regressors that arise from a tapped-delay-line implementation). The regressors are generated as independent realizations of a Gaussian distribution with a covariance matrix R_u whose eigenvalue spread is $\rho = 5$. Observe from the leftmost plot how expression (16.6) leads to a good fit between theory and practice for small step-sizes. On the other hand, as can be seen from the rightmost plot, expression (16.10) provides a better fit over a wider range of step-sizes.

SIMULATION RESULTS

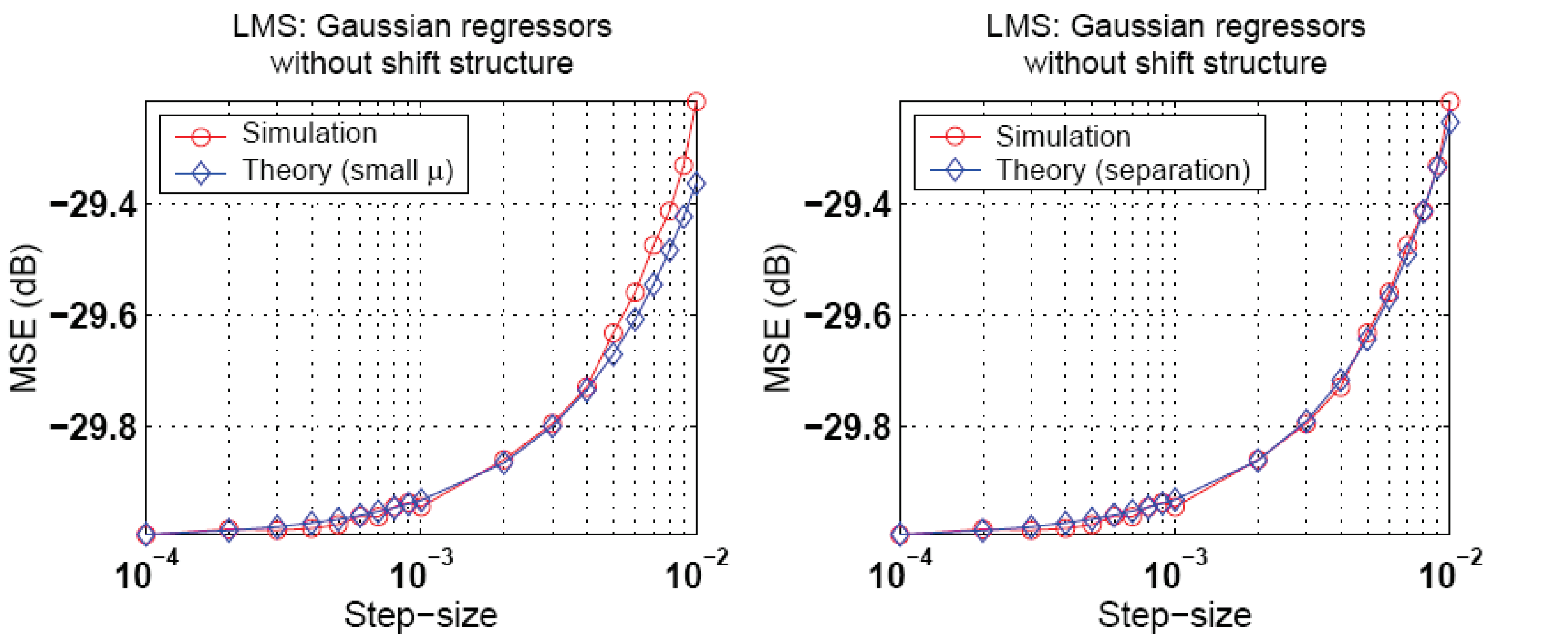

FIGURE 16.2 Theoretical and simulated MSE for a 10-tap LMS filter with $\sigma_v^2 = 0.001$ and Gaussian regressors *without* shift structure. The leftmost plot compares the simulated MSE with expression (16.6), which was derived under the assumption of small step-sizes. The rightmost plot uses expression (16.10), which was derived using the separation assumption (16.7).

SIMULATION RESULTS

In Fig. 16.3, the regressors $\{\boldsymbol{u}_i\}$ have shift structure and they are generated by feeding correlated data $\{\boldsymbol{u}(i)\}$ into a tapped delay line. The correlated data are obtained by filtering a unit-variance i.i.d. Gaussian random process $\{\boldsymbol{s}(i)\}$ through a first-order auto-regressive model with transfer function $\sqrt{1-a^2}/(1-az^{-1})$ and $a = 0.8$. It is shown in Prob. IV.1 that the auto-correlation sequence of the resulting process $\{\boldsymbol{u}(i)\}$ is $r(k) = \mathsf{E}\,\boldsymbol{u}(i)\boldsymbol{u}(i-k) = a^{|k|}$, for all integer values k. In this way, the covariance matrix R_u of the regressor $\boldsymbol{u}_i$ is a 10×10 Toeplitz matrix with entries $\{a^{|i-j|}, 0 \leq i, j \leq M-1\}$.

In Fig. 16.4, regressors with shift structure are again used but they are now generated by feeding into the tapped delay line a unit-variance *white* (as opposed to correlated) process so that $R_u = \sigma_u^2 \mathrm{I}$ with $\sigma_u^2 = 1$. This situation allows us to verify the third result in Lemma 16.1. It is seen from all these simulations that the expressions of Lemma 16.1 provide reasonable approximations for the EMSE of the LMS filter. In particular, expression (16.10), which was derived under the separation assumption (16.7), provides a good match between theory and practice.

SIMULATION RESULTS

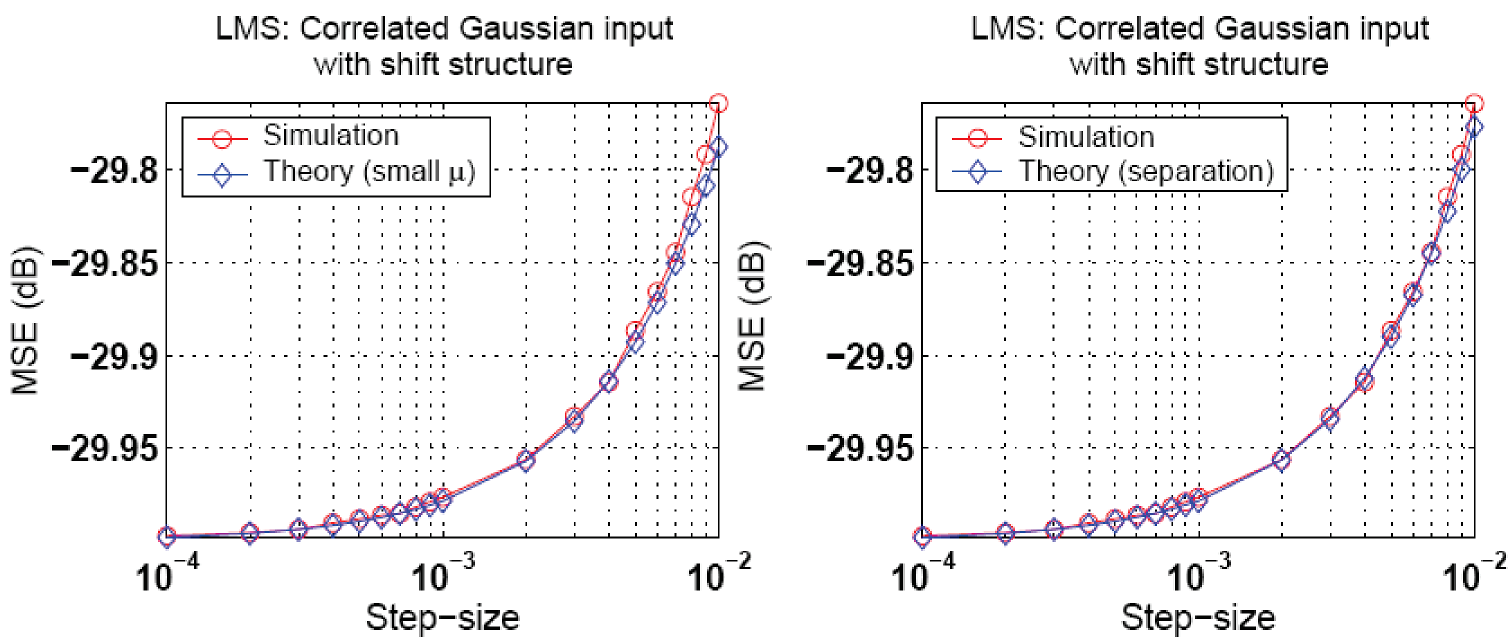

FIGURE 16.3 Theoretical and simulated MSE for a 10-tap LMS filter with $\sigma_v^2 = 0.001$ and regressors *with* shift structure. The regressors are generated by feeding *correlated* data into a tapped delay line. The leftmost plot compares the simulated MSE with expression (16.6), which was derived under the assumption of small step-sizes. The rightmost plot uses expression (16.10), which was derived using the separation assumption (16.7).

SIMULATION RESULTS

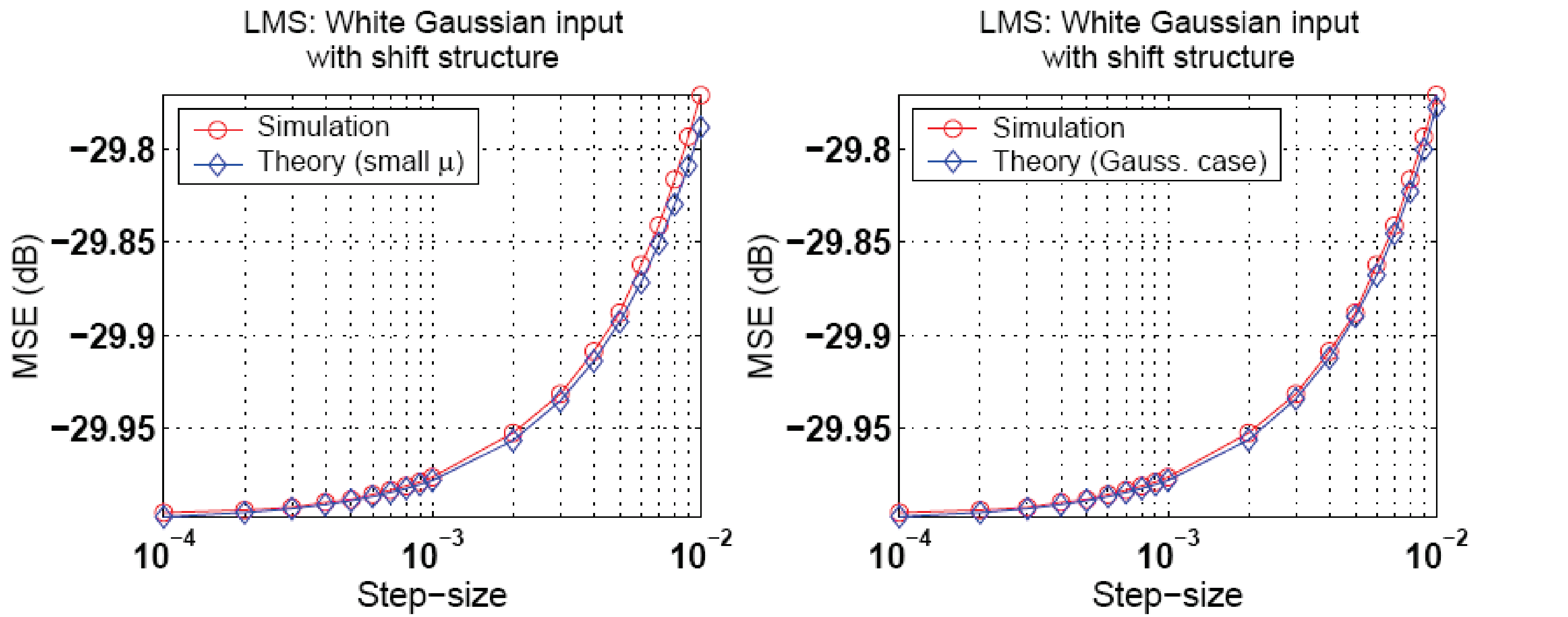

FIGURE 16.4 Theoretical (using (16.20)) and simulated MSE for a 10-tap LMS filter with $\sigma_v^2 = 0.001$ and regressors with shift structure. The regressors are generated by feeding *white* Gaussian input data with unit variance into a tapped delay line.

OTHER FILTERS

TABLE 19.1 Approximate expressions for the excess mean-square performance of several adaptive filters for sufficiently small step-sizes.

Algorithm	EMSE	Reference
LMS	$\mu\sigma_v^2 \text{Tr}(R_u)/2$	Lemma 16.1
ϵ-NLMS	$\frac{\mu\sigma_v^2}{2-\mu}\text{Tr}(R_u)\text{E}\left(\frac{1}{\lVert \boldsymbol{u}_i \rVert^2}\right)$	Lemma 17.1
ϵ-NLMS with power normalization	$\frac{\mu(1+\beta)M\sigma_v^2}{2\gamma(1-\beta)-\mu M(1+\beta)}$	Problem IV.5
sign-error LMS	$\frac{\alpha}{2}\left(\alpha+\sqrt{\alpha^2+4\sigma_v^2}\right), \ \alpha=\sqrt{\frac{\gamma\pi}{8}}\,\mu\text{Tr}(R_u)$	Lemma 18.1
LMF	$\mu\xi_v^6\text{Tr}(R_u)/2\sigma_v^2$	Problem IV.6
LMMN	$\mu a\text{Tr}(R_u)/2b$	Problem IV.6
leaky-LMS	$\frac{\mu\sigma_v^2}{2}\text{Tr}[R_u^2(R_u+\alpha\text{I})^{-1}]$	Problem V.32
sign-regressor LMS	$\mu\sigma_v^2 M/\left(\sqrt{\frac{8}{\pi\sigma_u^2}}-\mu M\right)$	Problem V.25
ϵ-APA	$\frac{\mu\sigma_v^2}{2-\mu}\text{Tr}(R_u)\text{E}\left(\frac{K}{\lVert \boldsymbol{u}_i \rVert^2}\right)$	Problem IV.7
RLS	$\sigma_v^2(1-\lambda)M/2$	Lemma 19.1
CMA2-2	$\mu\,\frac{\text{E}\left(\gamma^2\lvert s\rvert^2-2\gamma\lvert s\rvert^4+\lvert s\rvert^6\right)}{2\text{E}\left(2\lvert s\rvert^2-\gamma\right)}\,\text{Tr}(R_u)$	Problem IV.17
CMA1-2	$\mu\left(\gamma^2+\text{E}\lvert s\rvert^2-2\gamma\text{E}\lvert s\rvert\right)\text{Tr}(R_u)/2$	Problem IV.18