Adaptive Filters

V. John Mathews
Scott C. Douglas

Contents

5 Analysis of the LMS Adaptive Filter **3**
5.1 The Analysis Procedure 3
5.2 The Independence Assumption 4
5.3 Mean Analysis of the LMS Adaptive Filter 5
5.3.1 Mean Behavior of the Coefficient Errors in Stationary Environments . 6
5.3.2 Coefficient Dependence on the Input Signal Statistics 6
5.3.3 Summary of the Analysis of the Mean Behavior of the Coefficients . . 7
5.3.4 Conditions for Convergence of the Mean Coefficient Values 8
5.3.5 An Easily-Evaluated Step Size Stability Condition 9
5.3.6 Effects of Correlated Input Signals 9
5.3.7 Mean Convergence Time 11
5.3.8 Effect of Input Signal Spectrum on Coefficient Behavior 13
5.4 Mean-Square Analysis of the LMS Adaptive Filter 14
5.4.1 Key Assumptions 15
5.4.2 Computation of the Mean-Squared Error 16
5.4.3 Misadjustment 17
5.4.4 Evolution of the Coefficient Error Correlation Matrix 18
5.4.5 Misadjustment for Small Step Sizes and Arbitrary Signals 26
5.4.6 Summary of of the Mean-Square Analysis for Stationary Input Signals 29
5.4.7 Conditions for Mean-Square Convergence 30
5.5 Mean-Square Analysis in a Nonstationary Environment 34
5.5.1 Nonstationarity Model 34
5.5.2 Analysis 35
5.6 Other Analyses of the LMS Adaptive Filter 38
5.6.1 An Exact Expectation Analysis 39
5.6.2 Averaging Analysis 47
5.6.3 Robustness Analysis of the LMS Adaptive Filter 49
5.7 Main Points of This Chapter 52
5.8 Bibliographical Notes 53
5.9 Exercises 55

Chapter 5

Analysis of the LMS Adaptive Filter

In the last chapter, we saw how the least-mean-square (LMS) algorithm can be used to adjust the coefficients of an FIR filter for stationary and nonstationary filtering tasks. In this chapter, we analyze the behavior of the LMS adaptive filter using a method of statistical analysis that has proven useful in determining the performance characteristics of the system on average. We will gain insight into the interplay between the LMS adaptive filter's convergence behavior and the input signal statistics of the system using this analysis. The analysis also gives us insight as to how to best choose the step size μ to get desirable adaptation characteristics. Lastly, we compare our analysis method with other types of analyses that have been used to understand the behavior of the LMS adaptive filter.

5.1 The Analysis Procedure

For all the analyses of adaptive algorithms in this and subsequent chapters, we follow a systematic procedure for determining the average behavior of the system under study. This procedure can be described by the following steps:

1. *Make certain assumptions about the signals being processed by the adaptive system.* These assumptions are usually in the form of a statistical description of the signals $x(n)$ and $d(n)$ being processed. For example, we may assume in some cases that the input and desired response signals are jointly Gaussian with zero means and known (finite) covariances and cross-covariances. While not truly accurate in most cases, these assumptions form a starting point from which the analyses can proceed.

2. *Form one or more difference equations relating the quantities of interest at time $n+1$ to quantities at time n.* Depending on the type of study, one or more equations may be necessary to describe the evolutions of the quantities of interest.

3. *Take expectations of both sides of these difference equations by making use of the statistical assumptions.* In this way, we determine a relationship between the averaged

quantities of interest in the form of *evolution equations* over all possible statistical realizations of the signals being processed.

4. *Analyze the resulting equations.* Depending on what is desired, analyses of the resulting equations can yield stability conditions on the step size μ, the final values of the quantities of interest, or evolution curves of the quantities of interest as a function of time.

5.2 The Independence Assumption

We start our analysis by making an assumption about the statistical nature of the input and the desired response signals.

The Independence Assumption: The pair consisting of the input vector $\mathbf{X}(n)$ and the desired response sample $d(n)$ is independent of input vector $\mathbf{X}(m)$ and desired response sample $d(m)$ for $m \neq n$.

This assumption implies that the elemental pairs $\{\mathbf{X}(n), \mathbf{X}(m)\}$ $\{\mathbf{X}(n), d(m)\}$ and $\{d(n), d(m)\}$ are uncorrelated if $m \neq n$. We can immediately see a difficulty with this assumption. By construction, the input data vector $\mathbf{X}(n)$ cannot be independent of $\mathbf{X}(m)$ for a single-channel, finite-impulse response filter if $|n-m| \leq L-1$, because the two vectors share elements. Elements that are identical are *perfectly correlated*, whereas independence implies uncorrelatedness. Moreover, in most practical situations, both the input and desired response signals are correlated over time, thus bringing further question to the validity of the assumption.

Despite these apparent difficulties, we will use this assumption for the following three reasons.

1. *The analyses using the independence assumption are reasonably accurate.* Experience has shown that the analyses using this assumption are quite accurate in predicting the average behavior of the LMS algorithm in situations where the assumption is clearly violated. Such analyses have led to useful design rules for stochastic gradient adaptive filters.

2. *It results in a first-order approximation of a more accurate analysis.* It has been shown [Mazo 1979] that when one considers a more accurate description of the correlations between input vectors and between desired responses, analyses using the independence assumption capture the first order behavior of the algorithm when the step size is small, *i.e.,* when the convergence behavior of the algorithm is suitably slow. We provide results of a more general exact analysis that does not use this assumption later in this chapter, and comparisons of the results of the more accurate analysis with those produced using the independence assumption show no significant difference for small step sizes.

3. *It leads to a simple analysis with results that can be characterized.* The independence assumption allows us to develop a simple description of the mean and mean-square behavior of the filter coefficients. The results can be analyzed further to characterize the performance of the adaptive algorithm in terms of measurable quantities such as the input signal statistics. Our analyses also guide us in the choice of the step size to give accurate adaptation.

5.3 Mean Analysis of the LMS Adaptive Filter

We begin by analyzing the behavior of the mean values of the filter coefficients. For this analysis, we can take the statistical expectations of both sides of the coefficient update equation in (4.21), which gives

$$E\{\mathbf{W}(n+1)\} \;=\; E\{\mathbf{W}(n)\} + \mu E\{e(n)\mathbf{X}(n)\}. \tag{5.1}$$

Now, we can evaluate the second term in (5.1) by substituting for $e(n)$ using (4.22) as

$$\begin{aligned} E\{e(n)\mathbf{X}(n)\} &= E\{d(n)\mathbf{X}(n)\} - E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{W}(n)\} \\ &= \mathbf{P}_{d\mathbf{X}}(n) - E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{W}(n)\}. \end{aligned} \tag{5.2}$$

Recall that the coefficients at time n depend only on the input data at time $n-1$ and before. Therefore, by virtue of the independence assumption, the coefficient vector $\mathbf{W}(n)$ is independent of the input vector $\mathbf{X}(n)$. This result allows us to evaluate the second term in (5.2) as

$$E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{W}(n)\} \;=\; E\{\mathbf{X}(n)\mathbf{X}^T(n)\}E\{\mathbf{W}(n)\}. \tag{5.3}$$

With this simplification, (5.1) becomes a first-order difference equation given by

$$E\{\mathbf{W}(n+1)\} \;=\; E\{\mathbf{W}(n)\} + \mu(\mathbf{P}_{d\mathbf{X}}(n) - \mathbf{R}_{\mathbf{XX}}(n)E\{\mathbf{W}(n)\}). \tag{5.4}$$

Comparing the above equation for $E\{\mathbf{W}(n+1)\}$ with the equation for the coefficient update for the steepest descent algorithm in (4.10), we see that they are identical. This correspondence means that, if the independence assumption is valid, the coefficients of the LMS adaptive filter behaves on average exactly like the coefficients of the steepest descent algorithm with the same initial coefficient values and signal statistics. This result certainly seems reasonable given the results of Example 4.2. The example indicated that the ensemble-averaged behavior of the LMS filter coefficients is quite similar to that of the coefficients for the steepest descent procedure when the statistics of the data are the same as those used in the steepest descent procedure.

5.3.1 Mean Behavior of the Coefficient Errors in Stationary Environments

To continue our analysis, we assume that the input and desired response signal statistics are jointly wide-sense stationary; *i.e.*, $\mathbf{R}_{\mathbf{XX}}(n)$ and $\mathbf{P}_{d\mathbf{X}}(n)$ are constant over time. A consequence of this assumption is that the optimum filter coefficient vector $\mathbf{W}_{opt} = \mathbf{R}_{\mathbf{XX}}^{-1}\mathbf{P}_{d\mathbf{X}}$ is constant over time. It is convenient to define the *coefficient error vector* $\mathbf{V}(n)$ as

$$\mathbf{V}(n) = \mathbf{W}(n) - \mathbf{W}_{opt}. \tag{5.5}$$

This vector is also known as the *coefficient misalignment vector* in the literature.

Substituting (5.5) in (5.4) gives

$$E\{\mathbf{V}(n+1)\} = E\{\mathbf{V}(n)\} + \mu(\mathbf{P}_{d\mathbf{X}} - \mathbf{R}_{\mathbf{XX}}E\{\mathbf{V}(n)\} - \mathbf{R}_{\mathbf{XX}}\mathbf{W}_{opt}). \tag{5.6}$$

Noting that $\mathbf{R}_{\mathbf{XX}}\mathbf{W}_{opt} = \mathbf{P}_{d\mathbf{X}}$ we can simplify this equation further to

$$\begin{aligned} E\{\mathbf{V}(n+1)\} &= E\{\mathbf{V}(n)\} - \mu\mathbf{R}_{\mathbf{XX}}E\{\mathbf{V}(n)\} \\ &= (\mathbf{I} - \mu\mathbf{R}_{\mathbf{XX}})E\{\mathbf{V}(n)\}. \end{aligned} \tag{5.7}$$

The above equation describes the evolution of the mean of the coefficient errors in a compact form. It states that the mean errors in the filter coefficients at time $n+1$ are linear combinations of the mean errors in the filter coefficients at time n. Moreover, the evolutions of these errors depend only on the autocorrelation statistics of the input vector and the step size.

5.3.2 Coefficient Dependence on the Input Signal Statistics

A key observation that we can make from (5.7) is that the behavior of the mean of one filter coefficient is in general tightly coupled with the behaviors of all of the other coefficients. The mean of the ith element of the coefficient error vector at time $n+1$ can be expressed as

$$E\{v_i(n+1)\} = (1 - \mu r_{ii})E\{v_i(n)\} - \mu \sum_{j=1, j\neq i}^{L} r_{ij}E\{v_j(n)\}. \tag{5.8}$$

It is obvious from the above equation that, unless the autocorrelation matrix $\mathbf{R}_{\mathbf{XX}}$ is diagonal such that $r_{ij} = 0$ for $i \neq j$, the value of $E\{v_i(n+1)\}$ depends on other coefficient errors at time n.

We can gain further insight into the behavior of the mean of the adaptive filter coefficients by using a linear transformation of the coefficient error vector that removes the dependence of the elements of the transformed vector from one another. It is shown in Appendix B that we can decompose $\mathbf{R}_{\mathbf{XX}}$ into the following matrix product:

$$\mathbf{R}_{\mathbf{XX}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \tag{5.9}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues [diag(λ_0, λ_1, $\cdots$, λ_{L-1})] and $\mathbf{Q}$ is a Hermitian matrix of eigenvectors such that $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Define a *transformed coefficient error vector* $\widetilde{\mathbf{V}}(n)$ as

$$\widetilde{\mathbf{V}}(n) \quad = \quad \mathbf{Q}^T\mathbf{V}(n). \tag{5.10}$$

Premultiplying both sides of equation (5.7) by $\mathbf{Q}^T$ and recognizing that $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, we can obtain a difference equation that describes the mean behavior of $\widetilde{\mathbf{V}}(n)$ as

$$E\{\widetilde{\mathbf{V}}(n+1)\} \quad = \quad \mathbf{Q}^T(\mathbf{I} - \mu\mathbf{R_{xx}})\mathbf{Q}\mathbf{Q}^T E\{\mathbf{V}(n)\}(\mathbf{I} - \mu\mathbf{\Lambda})E\{\widetilde{\mathbf{V}}(n)\}. \tag{5.11}$$

Because the matrix $(\mathbf{I} - \mu\mathbf{\Lambda})$ is diagonal, we can write the update equation for any one of the elements of $E\{\widetilde{\mathbf{V}}(n)\}$ as

$$E\{\widetilde{v}_i(n+1)\} \quad = \quad (1 - \mu\lambda_i)E\{\widetilde{v}_i(n)\} \tag{5.12}$$

for $0 \leq i \leq L-1$. Thus, the convergence behavior of each of the elements of $E\{\widetilde{\mathbf{V}}(n)\}$ are *decoupled* from the rest of the elements. These scalar equations produce geometric series, whose terms are given by

$$E\{\widetilde{v}_i(n)\} \quad = \quad (1 - \mu\lambda_i)^n E\{\widetilde{v}_i(0)\}. \tag{5.13}$$

5.3.3 Summary of the Analysis of the Mean Behavior of the Coefficients

Let us pause a moment and summarize what we have learned from the mean analysis of the LMS adaptive filter coefficients.

- *The average behavior of the coefficient vector of the LMS adaptive filter is the same as that of the coefficient vector of the steepest descent algorithm.* Because the independence assumption only provides reliable results for small step sizes, we can only draw a correspondence between the mean analysis equation in (5.4) for the LMS adaptive filter and the update in (4.10) for the steepest descent algorithm for small step size values.

- *The evolution of the mean value of any one of the filter coefficients to its optimum MMSE value consists of a sum of exponentially-converging and/or diverging terms.* This result follows from Equations (5.5), (5.10), and (5.13), showing that the value of the ith element of $E\{\mathbf{W}(n)\}$ is given by

$$E\{w_i(n)\} \quad = \quad w_{i,opt} + \sum_{j=1}^{L} q_{ij}(1 - \mu\lambda_j)^n E\{\widetilde{v}_j(0)\}, \tag{5.14}$$

  where q_{ij} is the (i,j)th element of the matrix $\mathbf{Q}$.

- *The convergence behavior of the mean value of any one of the filter coefficients depends on the step size and the eigenvalues of the input autocorrelation matrix.* This result follows from (5.13) or (5.14), which shows that the step size and the eigenvalues $\{\lambda_i\}$ determine the magnitude of the exponential terms making up the mean coefficient values. The convergence or divergence of these terms governs the convergence or divergence of the mean of the coefficients of the adaptive filter.

5.3.4 Conditions for Convergence of the Mean Coefficient Values

From the foregoing analysis, we can determine a simple stability condition on the step size μ to guarantee the convergence of the mean values of the coefficients of the LMS adaptive filter. The adaptive algorithm converges in the above sense if the mean values of the coefficient error vector elements attain finite steady-state values in stationary operating environments. Furthermore, if the mean values of the coefficient errors are zero in the steady-state, the mean coefficient vector sequence will converge to $\mathbf{W}_{opt}$. The above statements imply convergence of the filter coefficients *on average*. They do not imply that the filter coefficients $\mathbf{W}(n)$ will converge to $\mathbf{W}_{opt}$ for any particular realization of the input and desired response signals.

We can see from (5.13) that $E\{\tilde{v}_i(n)\}$ will converge to zero for a finite initial error value if and only if

$$-1 \ < \ (1-\mu\lambda_i) \ < \ 1. \tag{5.15}$$

The two inequalities lead to two sets of conditions that must be satisfied for $0 \leq i \leq L-1$ for the mean coefficient values of the LMS adaptive filter to converge:

$$\mu > 0 \quad \text{and} \quad \mu < \frac{2}{\lambda_i} \quad \text{for all } 0 \leq i \leq L-1. \tag{5.16}$$

Since these inequalities must hold for every eigenvalue λ_i, the most stringent inequality is

$$0 < \mu < \frac{2}{\lambda_{max}}, \tag{5.17}$$

where λ_{max} is the maximum eigenvalue of $\mathbf{R_{xx}}$. These step size bounds guarantee convergence of all the modes of the analysis equation in (5.14).

REMARK 5.1: We see from (5.17) that the step size must be positive and must be smaller than a quantity that is proportional to the inverse of the maximum eigenvalue of the input data autocorrelation matrix. One can see an immediate practical difficulty with the LMS algorithm in real-world applications: one must have some knowledge of the input signal statistics in order to choose a "good" step size value. In many situations, this knowledge is unavailable or is known with some uncertainty. In these cases, we must either 1) choose a conservative step size so that the LMS algorithm converges for the range of input statistics for the application or 2) estimate a quantity from the data that gives us the necessary knowledge.

5.3.5 An Easily-Evaluated Step Size Stability Condition

Another difficulty with the step size bound in (5.17) is that it is difficult to evaluate in many situations because procedures to estimate the maximum eigenvalue of $\mathbf{R_{xx}}$ are both computationally-intensive and sensitive to uncertainty in the estimate of the autocorrelation matrix. However, we can get around this difficulty by using bounds on λ_{max} that are easier to calculate. Specifically, we note from Appendix B that $\mathbf{R_{xx}}$ is a symmetric, positive-definite matrix, and thus

$$\mathrm{tr}[\mathbf{R_{xx}}] \;=\; \sum_{i=1}^{L} \lambda_i \;>\; \lambda_{max}. \tag{5.18}$$

We can obtain a more stringent bound on the step size to guarantee convergence of the mean coefficient values from (5.18) as

$$0 < \mu < \frac{2}{\mathrm{tr}[\mathbf{R_{xx}}]}. \tag{5.19}$$

For stationary input signals, this bound becomes

$$0 < \mu < \frac{2}{L\sigma_x^2}, \tag{5.20}$$

where σ_x^2 is the mean-squared value of the input signal, a quantity that can be reliably estimated directly from the input signal.

5.3.6 Effects of Correlated Input Signals

Given the theoretical results derived above for the convergence behavior of the mean values of the coefficients, we now consider the effect that correlation in the input signal has on the convergence behavior of the LMS adaptive filter. For this study, we consider the following examples.

Example 5.1: Mean Behavior of the LMS Adaptive Filter with I.I.D. Input Signals

Let the input signal be a zero-mean, i.i.d. random process such that the autocorrelation of this process is

$$r_{xx}(n) \;=\; \sigma_x^2 \delta(n).$$

In this case, the autocorrelation matrix of the input signal is

$$\mathbf{R_{xx}} \;=\; \sigma_x^2 \mathbf{I}.$$

The eigenvalues of $\mathbf{R_{XX}}$ are all equal and given by $\lambda_i = \sigma_x^2$. Thus, we find from (5.8) that the mean error associated with the ith filter coefficient behaves according to

$$E\{v_i(n+1)\} = \left(1 - \mu\sigma_x^2\right) E\{v_i(n)\}.$$

The above result implies that the mean value of $w_i(n)$ converges to $w_{i,opt}$ at an exponential rate, independent of the behavior of the other coefficients. Moreover, the mean of each of the coefficients converges at the same speed. In this situation, it is easy to choose a step size that gives the desired speed of convergence to all the coefficients at the same time.

Example 5.2: Mean Behavior of the LMS Adaptive Filter with Correlated Input Signals

Consider the case of a two-tap adaptive filter for which $\mathbf{R_{XX}}$ is given by

$$\mathbf{R_{XX}} = \frac{1}{2}\begin{bmatrix} 1+\epsilon & 1-\epsilon \\ 1-\epsilon & 1+\epsilon \end{bmatrix},$$

where ϵ is a positive number smaller than one. It can be shown that $\mathbf{\Lambda}$ and $\mathbf{Q}$ for this autocorrelation matrix are given by

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix}$$

and

$$\mathbf{Q} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

respectively.

We can show using (5.14) that the mean coefficient values evolve in time according to

$$\begin{aligned} E\{w_1(n)\} &= w_{1,opt} + \frac{1}{\sqrt{2}}\left((1-\mu)^n E\{\widetilde{v}_1(0)\} - (1-\mu\epsilon)^n E\{\widetilde{v}_2(0)\}\right) \\ E\{w_2(n)\} &= w_{2,opt} + \frac{1}{\sqrt{2}}\left((1-\mu)^n E\{\widetilde{v}_1(0)\} + (1-\mu\epsilon)^n E\{\widetilde{v}_2(0)\}\right). \end{aligned}$$

For our example, let $\mathbf{W}_{opt} = [1\ \ 1]^T$ and $\mathbf{W}(0) = [3\ \ 2]^T$. The input signal was generated using the model in Example 4.2 with $a = (1-\epsilon)/(1+\epsilon)$ and $b = \sqrt{2\epsilon/(1+\epsilon)}$. The reader can verify that the autocorrelation matrix for this signal is given by that defined in this example. Figure 5.1 and 5.2 show the behavior of $w_1(n)$ and $w_2(n)$, respectively, of the LMS adaptive filter for a step size of $\mu = 0.1$ for three individual experiments, each generated according to the model in Example 4.2 for the the three different values of ϵ given above. The reader can easily see that, although the coefficients of the LMS adaptive filter do not exactly follow the predicted values $E\{w_1(n)\}$ and $E\{w_2(n)\}$, the general trends of the curves are the same in every case. Moreover, as is shown

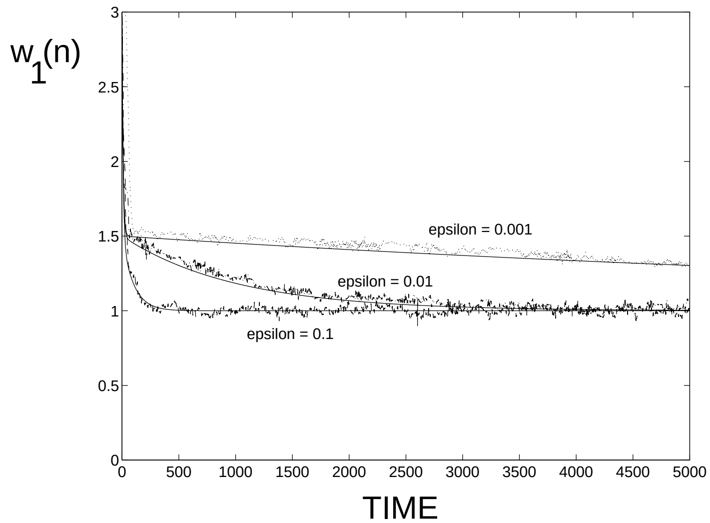

Figure 5.1: Evolution of LMS adaptive filter coefficient $w_1(n)$ in Example 5.2.

throughout this chapter, by averaging the results of several experiments together, the resulting ensemble averages closely follow the expected values of the quantities of interest as determined from our analyses.

From the plots, we find that the convergence of the filter coefficients is slowed for smaller values of ϵ. This behavior can be explained using the equations describing the evolution of $E\{w_1(n)\}$ and $E\{w_2(n)\}$ given above. The error terms associated with $E\{\widetilde{v}_2(0)\}$ each have an exponential factor $(1-\mu\epsilon)$ that is very close to one when the product of μ and ϵ is small. Since the step size bound in (5.17) limits the magnitude of the step size in this example to $0 < \mu < 2$ for stability considerations, it is impossible to have fast convergence for small ϵ.

5.3.7 Mean Convergence Time

The concepts illustrated by Examples 5.1 and 5.2 naturally extend to the case of an L-coefficient LMS adaptive filter. We now quantify the speed of convergence of the mean values of the filter coefficients. Consider the evolution of $E\{w_i(n)\}$ as given by (5.14). We can write this equation as

$$E\{w_i(n)\} \quad = \quad w_{i,opt} + \sum_{j=1}^{L} q_{ji} a_j^n E\{\widetilde{v}_j(0)\}, \tag{5.21}$$

where $a_j = 1 - \mu\lambda_j$. The mean error in the ith coefficient consists of a weighted sum of L exponentially-decaying terms that converge exponentially to zero according to the magnitudes of a_j. We can associate a *time constant* τ_j for the jth mode of convergence as the time

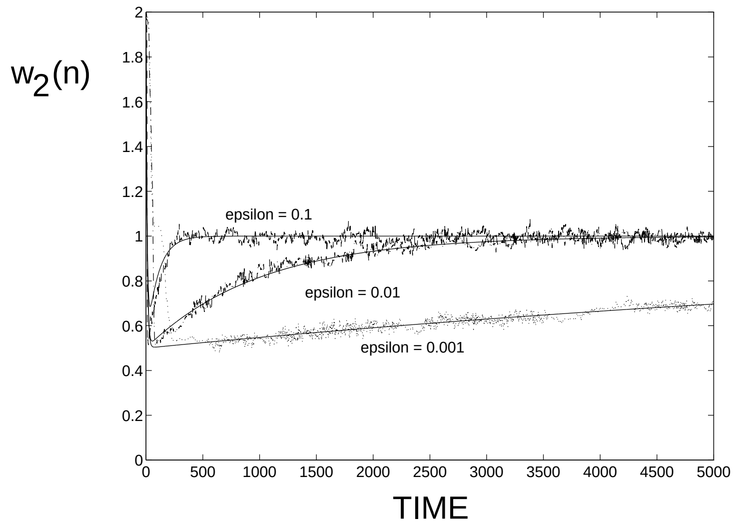

Figure 5.2: Evolution of LMS adaptive filter coefficient $w_2(n)$ in Example 5.2.

it takes to decay from $E\{\widetilde{v}_j(0)\}$ to $(1/e)E\{\widetilde{v}_j(0)\}$. Thus, τ_j is defined by the relation

$$(1-\mu\lambda_j)^{\tau_j} E\{\widetilde{v}_j(0)\} \quad = \quad \frac{1}{e} E\{\widetilde{v}_j(0)\}. \tag{5.22}$$

Assuming that $\mu\lambda_j < 1$, we can solve for τ_j as

$$\tau_j \quad = \quad -\frac{1}{\ln(1-\mu\lambda_j)}. \tag{5.23}$$

Moreover, if $\mu\lambda_j \ll 1$, which is often the case in practical situations, we can use the Taylor series approximation $-\ln(1-\mu\lambda_j) \approx \mu\lambda_j$ to express τ_j as

$$\tau_j \quad \approx \quad \frac{1}{\mu\lambda_j}. \tag{5.24}$$

Thus, the time constant for each mode of convergence is inversely proportional to μ and the corresponding eigenvalue λ_j. Increasing the step size μ for a given set of eigenvalues will lead to smaller time constants and faster adaptation time.

Convergence Speed of the LMS Adaptive Filter

Each time constant can be useful in describing one mode of convergence of the adaptive filter. However, it is not straightforward to combine these time constants in a meaningful way to provide a single measure of the rate of convergence of the filter. In the best scenario, a slowly-converging mode may not be of concern if the initial coefficient error associated with this mode is small. In the worst case situation where the error in the transformed coefficient associated with the slowest mode of convergence is very large, the rates of convergence of the mean values of the filter coefficients depend on the mode associated with the largest time constant. Since we typically have no knowledge of the initial coefficient errors, the slowest converging mode is often associated with the speed of convergence of the adaptive filter. Other measures describing the convergence speed of the LMS adaptive filter are the *average* of the time constants, given by

$$\tau_{ave} \quad = \quad \frac{1}{L}\sum_{i=0}^{L-1}\tau_i, \tag{5.25}$$

and the *harmonic mean* of the time constants, given by

$$\tau_h \quad = \quad \left(\frac{1}{L}\sum_{i=0}^{L-1}\frac{1}{\tau_i}\right)^{-1}. \tag{5.26}$$

Affect of the Condition Number of the Input Autocorrelation Matrix on the Convergence Speed

From our analysis, we see that the difficulties in achieving fast convergence behavior of the means of the LMS adaptive filter coefficients are greater if the ratio of the minimum to the maximum eigenvalues of $\mathbf{R_{xx}}$ is large. The quantity $\lambda_{max}/\lambda_{min}$ is known as the *condition number* of the autocorrelation matrix, and it indicates the severity of the convergence problem that we have outlined. In general, highly-correlated input signals lead to large condition numbers for $\mathbf{R_{xx}}$, and the LMS algorithm is slow to converge for signals with such characteristics.

5.3.8 Effect of Input Signal Spectrum on Coefficient Behavior

We can relate the above concepts regarding input signal correlation to the power spectral density of the input signal. The power spectral density of a random process $x(n)$ is defined as

$$S_{xx}(\omega) \quad = \quad \mathcal{F}\{r_{xx}(n)\}, \tag{5.27}$$

where $\mathcal{F}\{\cdot\}$ denotes the discrete-time Fourier transform operation. It can be shown [Gray 1974] that as the dimension L of the autocorrelation matrix $\mathbf{R_{xx}}$ is increased, the eigenvalues of $\mathbf{R_{xx}}$ tend toward the values of the power spectrum $S_{xx}(\omega)$ evaluated at equally-spaced points across the frequency range $0 \leq \omega \leq \pi$. This result implies that that the limiting value of the condition number of the autocorrelation matrix as L tends to infinity is the ratio of the maximum and minimum values of the power spectral density of the input signal, *i.e.,*

$$\lim_{L\to\infty} \frac{\lambda_{max}}{\lambda_{min}} \quad = \quad \frac{\max\limits_{0\leq\omega\leq\pi} S_{xx}(\omega)}{\min\limits_{0\leq\omega\leq\pi} S_{xx}(\omega)}. \tag{5.28}$$

Thus, the ratio of the largest and smallest values of the power spectral density of the input signal can be used to indicate the severity of the difficulty in achieving fast convergence of the LMS adaptive filter. Moreover, fast convergence becomes harder to attain as the length of the filter L is increased. A consequence of the above result is that the autocorrelation matrices of bandlimited signals tend to have high condition numbers. Consequently, the LMS adaptive filter will converge slowly for such input signals.

A few simple examples illustrate the above concepts.

- The power spectral density of i.i.d. signals is constant across the frequency range $0 \leq \omega \leq \pi$. Consequently, the expression in (5.28) implies that $\lambda_{max}/\lambda_{min} \to 1$ for this input signal. This statement agrees with the direct evaluation of the autocorrelation matrix $\mathbf{R_{xx}} = \sigma_x^2\mathbf{I}$ for any value of L in this case.

- Suppose the input signal consists of N sinusoids. Since a sum-of-sinusoids signal contains no energy at frequencies not represented in the signal, the minimum value of the power spectral density is zero, and thus the expression in (5.28) gives $\lambda_{max}/\lambda_{min} \to \infty$ as L is increased. It can be shown that for $L > 2N$, the matrix $\mathbf{R_{xx}}$ has $L - 2N$ zero eigenvalues, and thus $\lambda_{max}/\lambda_{min}$ is infinite. Consequently, $L-2N$ modes of convergence of the adaptive filter will not be affected by this input signal.

- Consider the input signal for the two-tap LMS adaptive filter in Example 5.2. The power spectra of this signal for the three values of ϵ considered are shown in Figure 5.3. Here, we have normalized the power spectra by their values at zero frequency for the purpose of comparison. Comparing the results in Figures 5.1 and 5.2 with the results in Figure 5.3, we see that the narrower the spectrum of the input signal, the slower the adaptive filter converges.

5.4 Mean-Square Analysis of the LMS Adaptive Filter

We have up to this point considered the behavior of the mean values of the coefficients of the LMS adaptive filter. In this section, we analyze the mean-square performance of the LMS adaptive filter. Mean-square convergence means that the power in the error of the system converges. This criterion is stronger than the convergence of the mean values of the filter coefficients. Consequently, the results we develop in this section provide a more complete

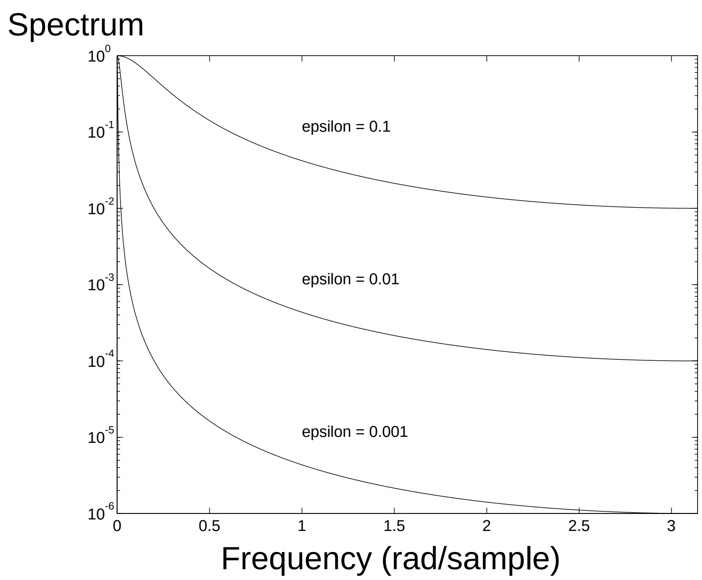

Figure 5.3: Normalized power spectral density $S_{xx}(\omega)/S_{xx}(0)$, of the input processes in Example 5.2.

characterization of the properties of the system than the previous results. This analysis also illustrates an important tradeoff in adaptive systems: convergence speed versus steady-state mean-squared error.

Our technique for analyzing the mean-squared behavior of the LMS adaptive filter is similar to that used for analyzing the mean coefficient behavior. We develop an iterative equation for the *correlation matrix* of the filter coefficient errors, defined as

$$\mathbf{K}(n) \quad = \quad E\{\mathbf{V}(n)\mathbf{V}^T(n)\}. \tag{5.29}$$

The elements of $\mathbf{K}(n)$ can then be used to describe the evolution of the mean-squared error $E\{e^2(n)\}$.

5.4.1 Key Assumptions

For the mean-square analysis, we use the independence and stationarity assumptions as previously described. In addition, we employ two other assumptions in order to simplify the analysis.

System Identification Model. The adaptive filter is configured to identify an unknown system from measurements of its input signal and a noisy version of its output signal. The unknown system is assumed to be a time-invariant FIR filter with coefficient vector $\mathbf{W}_{opt}$. Thus, the desired response signal can be described by

$$d(n) \quad = \quad \mathbf{W}_{opt}^T \mathbf{X}(n) + \eta(n), \tag{5.30}$$

where $\eta(n)$ is a zero mean and i.i.d. additive noise sequence. In addition, we assume that the adaptive filter uses at least the same number of coefficients as the unknown system.

This assumption on the desired response signal allows us to study the mean-square convergence behavior of the coefficient error vector $\mathbf{V}(n)$ directly. It also is a realistic assumption for tasks where the system identification model accurately describes the problem at hand.

Input Statistics Model. The input signal $x(n)$ is generated either from a zero-mean, possibly correlated Gaussian process or from a zero-mean, i.i.d. process with an even-symmetric probability density function.

This assumption on the input signal statistics simplifies the analysis considerably. In particular, it enables us to develop closed-form expressions for several parameters that characterize the steady-state behavior of the adaptive filter. The i.i.d. input signal model is valid in many applications involving system identification. The Gaussian input signal model is reasonable for signals generated as a sum of the outputs of many independent random sources, since the central limit theorem guarantees that the distribution of such data will be close to Gaussian in form as the number of sources becomes large.

5.4.2 Computation of the Mean-Squared Error

We can express the mean-squared error as

$$E\{e^2(n)\} \quad = \quad E\{(d(n) - \mathbf{W}^T(n)\mathbf{X}(n))^2\}. \tag{5.31}$$

Substituting the expression for $d(n)$ from (5.30) in the above equation gives

$$E\{e^2(n)\} \quad = \quad E\{(\mathbf{W}_{opt}^T \mathbf{X}(n) + \eta(n) - \mathbf{W}^T(n)\mathbf{X}(n))^2\}. \tag{5.32}$$

We can simplify this equation further by expressing it in terms of the coefficient error vector $\mathbf{V}(n)$. The mean-squared estimation error is then given by

$$E\{e^2(n)\} = E\{(\eta(n) - \mathbf{V}^T(n)\mathbf{X}(n))^2\}. \tag{5.33}$$

Expanding the right-hand-side of (5.33) and taking the expectations of the individual terms give

$$E\{e^2(n)\} = E\{(\eta(n))^2\} - 2E\{\eta(n)\mathbf{X}^T(n)\mathbf{V}(n)\} + E\{\mathbf{V}^T(n)\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\}. \tag{5.34}$$

Since $\eta(n)$ is uncorrelated with $\mathbf{X}(n)$ and since $\mathbf{V}(n)$ depends only on past values of both $\mathbf{X}(n)$ and $\eta(n)$, the second term on the right-hand-side of the above equation is zero. With this simplification, we arrive at the following expression for the mean-squared error at time n:

$$E\{e^2(n)\} = \sigma_\eta^2(n) + E\{\mathbf{V}^T(n)\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\}. \tag{5.35}$$

We see that the MSE at any time n is the sum of two terms: a noise power term that is independent of the filter coefficients, and a second term that depends upon the filter coefficients and the input data. The second term is called the *excess MSE* because it is the additional mean-square error in the filter output due to deviations in the filter coefficients from their optimum values.

Since $\mathbf{V}(n)$ is uncorrelated with the input vector under the independence assumption, we can express the excess MSE as

$$\begin{aligned}\xi_{ex}(n) &= E\{\mathbf{V}^T(n)\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\} \\ &= E\{\mathbf{V}^T(n)\mathbf{R_{xx}}\mathbf{V}(n)\} \\ &= \mathrm{tr}[\mathbf{R_{xx}}(n)\mathbf{K}(n)]. \end{aligned} \tag{5.36}$$

The last equality in (5.36) arises from the fact that the trace of the product of the two matrices $\mathbf{R_{xx}}$ and $\mathbf{V}(n)\mathbf{V}^T(n)$ is the same as the quantity $\mathbf{V}^T(n)\mathbf{R_{xx}}\mathbf{V}(n)$, as can be verified by direct evaluation.

5.4.3 Misadjustment

The *misadjustment* of the adaptive filter is a normalized measure of the noise in the filter output in the steady-state due to the fluctuations in the filter coefficients. The misadjustment M is defined as the ratio of the excess MSE in the steady-state and the minimum MSE,

$$\begin{aligned} M &= \lim_{n\to\infty} \frac{\xi_{ex}(n)}{\sigma_\eta^2} \\ &= \frac{\xi_{ex,ss}}{\sigma_\eta^2}, \end{aligned} \tag{5.37}$$

where $\xi_{ex,ss} = \lim_{n\to\infty} \xi_{ex}(n)$ is the steady-state value of the excess MSE. With this definition, we see that the total MSE in the filter output in steady-state is

$$\lim_{n\to\infty} E\{e^2(n)\} = (1+M)\sigma_\eta^2. \tag{5.38}$$

Since the minimum MSE obtainable at the filter output is the noise power σ_η^2, it is desirable to keep M as small as possible.

5.4.4 Evolution of the Coefficient Error Correlation Matrix

We begin our analysis by using the coefficient update equation in (4.21), the coefficient error definition in (5.5), and the desired response model in (5.30) to write an equation describing the evolution of the coefficient error vector $\mathbf{V}(n)$ as

$$\begin{aligned}\mathbf{V}(n+1) &= \mathbf{V}(n)+\mu\mathbf{X}(n)(\mathbf{X}^T(n)\mathbf{W}_{opt}+\eta(n)-\mathbf{X}^T(n)\mathbf{W}(n)) \\ &= (\mathbf{I}-\mu\mathbf{X}(n)\mathbf{X}^T(n))\mathbf{V}(n)+\mu\eta(n)\mathbf{X}(n). \end{aligned} \tag{5.39}$$

To develop an expression for $\mathbf{K}(n)$, we post-multiply both sides of (5.39) by their respective transposes, which gives

$$\begin{aligned}\mathbf{V}(n+1)\mathbf{V}^T(n+1) &= (\mathbf{I}-\mu\mathbf{X}(n)\mathbf{X}^T(n))\mathbf{V}(n)\mathbf{V}^T(n)(\mathbf{I}-\mu\mathbf{X}(n)\mathbf{X}^T(n)) \\ &\quad+\mu\eta(n)(\mathbf{I}-\mu\mathbf{X}(n)\mathbf{X}^T(n))\mathbf{V}(n)\mathbf{X}^T(n) \\ &\quad+\mu\eta(n)\mathbf{X}(n)\mathbf{V}^T(n)(\mathbf{I}-\mu\mathbf{X}(n)\mathbf{X}^T(n)) \\ &\quad+\mu^2\eta^2(n)\mathbf{X}(n)\mathbf{X}^T(n). \end{aligned} \tag{5.40}$$

Since $\eta(n)$ is a zero-mean process that is uncorrelated with the elements of the input vectors and coefficient vectors that appear in the second and third terms on the right-hand-side of (5.40), the statistical expectation of both of these terms are zero matrices. Consequently, taking expectations of both sides of (5.40) results in

$$\begin{aligned}\mathbf{K}(n+1) &= E\{(\mathbf{I}-\mu\mathbf{X}(n)\mathbf{X}^T(n))\mathbf{V}(n)\mathbf{V}^T(n)(\mathbf{I}-\mu\mathbf{X}(n)\mathbf{X}^T(n))\} \\ &\quad+\mu^2\sigma_\eta^2(n)\mathbf{R}_{\mathbf{XX}}. \end{aligned} \tag{5.41}$$

We can further simplify the above expression by employing the independence assumption to separate the expectations involving the signal and coefficient vectors, which gives

$$\begin{aligned}\mathbf{K}(n+1) &= \mathbf{K}(n)-\mu\left(\mathbf{R}_{\mathbf{XX}}\mathbf{K}(n)+\mathbf{K}(n)\mathbf{R}_{\mathbf{XX}}\right)+\mu^2\sigma_\eta^2\mathbf{R}_{\mathbf{XX}} \\ &\quad+\mu^2E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\mathbf{V}^T(n)\mathbf{X}(n)\mathbf{X}^T(n)\}. \end{aligned} \tag{5.42}$$

In order to proceed, we must evaluate the last expectation in (5.42). The calculations are simplified considerably if we express this expectation in terms of the transformed coefficient error vector $\widetilde{\mathbf{V}}(n)$ as defined in (5.10). Let us also define a similarly-transformed input vector as

$$\widetilde{\mathbf{X}}(n) = \mathbf{Q}^T\mathbf{X}(n). \tag{5.43}$$

It is easy to show that the covariance matrix of $\widetilde{\mathbf{X}}(n)$ is a diagonal matrix given by

$$\mathbf{R}_{\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}} = \mathbf{\Lambda}. \tag{5.44}$$

The last term of (5.42) can be written in terms of these transformed vectors as

$$\begin{aligned} & E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\mathbf{V}^T(n)\mathbf{X}(n)\mathbf{X}^T(n)\} \\ &= \mathbf{Q}E\{\mathbf{Q}^T\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{Q}\mathbf{Q}^T\mathbf{V}(n)\mathbf{V}^T(n)\mathbf{Q}\mathbf{Q}^T\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{Q}\}\mathbf{Q}^T \\ &= \mathbf{Q}E\{\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\}\mathbf{Q}^T. \end{aligned} \tag{5.45}$$

The (i,j)th element of the matrix $E\{\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\}$ is given by

$$\begin{aligned} & \Big[E\{\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\}\Big]_{i,j} \\ &= E\left\{\sum_{k=0}^{L-1}\sum_{m=0}^{L-1}\widetilde{x}_i(n)\widetilde{x}_j(n)\widetilde{x}_k(n)\widetilde{x}_m(n)\widetilde{v}_k(n)\widetilde{v}_m(n)\right\} \\ &= \sum_{k=0}^{L-1}\sum_{m=0}^{L-1}E\{\widetilde{x}_i(n)\widetilde{x}_j(n)\widetilde{x}_k(n)\widetilde{x}_m(n)\}E\{\widetilde{v}_k(n)\widetilde{v}_m(n)\}, \end{aligned} \tag{5.46}$$

where $\widetilde{x}_p(n)$ is the pth element of $\widetilde{\mathbf{X}}(n)$ and we have used the independence assumption to separate the expected values of the coefficients error elements and the input vector elements.

To evaluate (5.46), we must determine the fourth-order expectations within the summation. Recall from our input signal model that $\{x(n)\}$ is either a zero-mean i.i.d. or a zero-mean Gaussian random process. In addition, since $\mathbf{R}_{\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}}$ is a diagonal matrix, $\widetilde{x}_i(n)$ is uncorrelated with $\widetilde{x}_j(n)$ for $i \neq j$. Taken together, these two facts imply that $\widetilde{x}_i(n)$ and $\widetilde{x}_j(n)$ are statistically independent[1] for $i \neq j$. Consequently, we can evaluate $E\{\widetilde{x}_i(n)\widetilde{x}_j(n)\widetilde{x}_k(n)\widetilde{x}_m(n)\}$ as

$$E\{\widetilde{x}_i(n)\widetilde{x}_j(n)\widetilde{x}_k(n)\widetilde{x}_m(n)\} = \begin{cases} \kappa\lambda_i^2; & \text{if } i=j=k=m \\ \lambda_i\lambda_k; & \text{if } i=j\neq k=m \\ \lambda_i\lambda_j; & \text{if } i=k\neq j=m \text{ or } i=m\neq k=j \\ 0; & \text{otherwise} \end{cases}, \tag{5.47}$$

where κ is defined as

$$\kappa = \frac{E\{x^4(n)\}}{(E\{x^2(n)\})^2}. \tag{5.48}$$

For Gaussian input data, $\kappa = 3$. For other input types, we can evaluate κ directly from the probability density function.

We can evaluate the right-hand-side of (5.46) using (5.47) as

$$\begin{aligned} & \Big[E\{\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\}\Big]_{i,j} \\ &= \begin{cases} 2\lambda_i\lambda_j E\{\widetilde{v}_i(n)\widetilde{v}_j(n)\}; & \text{if } i\neq j \\ \kappa\lambda_i^2 E\{(\widetilde{v}_i(n))^2\} + \displaystyle\sum_{m=1,m\neq i}^{L}\lambda_i\lambda_m E\{(\widetilde{v}_m(n))^2\} & \text{if } i=j. \end{cases} \end{aligned} \tag{5.49}$$

[1] It can be shown that uncorrelated Gaussian random variables are also independent.

We now have the tools necessary to evaluate (5.42) for the input signal models of interest We consider the Gaussian and i.i.d. input cases separately.

The Gaussian Input Case.

For Gaussian input signals, we can use (5.49) to express the expectation on the right-hand-side of (5.45) as

$$\begin{aligned} & E\{\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\} \\ = \;& 2\mathbf{\Lambda} E\{\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\}\mathbf{\Lambda} + \mathbf{\Lambda}\mathrm{tr}[\mathbf{\Lambda} E\{\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\}]. \end{aligned} \tag{5.50}$$

It is left as an exercise for the reader to show by substitution of (5.50) into (5.45) that the evolution equation for $\mathbf{K}(n)$ in (5.42) simplifies to

$$\begin{aligned} \mathbf{K}(n+1) &= \mathbf{K}(n) - \mu\left(\mathbf{R_{xx}}\mathbf{K}(n) + \mathbf{K}(n)\mathbf{R_{xx}}\right) + 2\mu^2\mathbf{R_{xx}}\mathbf{K}(n)\mathbf{R_{xx}} \\ &\quad + \mu^2\mathbf{R_{xx}}\left(\sigma_\eta^2 + \mathrm{tr}[\mathbf{R_{xx}}\mathbf{K}(n)]\right). \end{aligned} \tag{5.51}$$

In a similar fashion as in the analysis of the mean behavior of the coefficients, it is useful to consider the evolution equation for the correlation matrix of the transformed coefficient error vector. This matrix is defined as

$$\begin{aligned} \widetilde{\mathbf{K}}(n) &= E\{\widetilde{\mathbf{V}}(n)\widetilde{\mathbf{V}}^T(n)\} \\ &= \mathbf{Q}^T\mathbf{K}(n)\mathbf{Q}. \end{aligned} \tag{5.52}$$

We can pre- and post-multiply both sides of (5.51) by $\mathbf{Q}^T$ and $\mathbf{Q}$, respectively, to get

$$\begin{aligned} \widetilde{\mathbf{K}}(n+1) &= \widetilde{\mathbf{K}}(n) - \mu\left(\mathbf{\Lambda}\widetilde{\mathbf{K}}(n) + \widetilde{\mathbf{K}}(n)\mathbf{\Lambda}\right) + 2\mu^2\mathbf{\Lambda}\widetilde{\mathbf{K}}(n)\mathbf{\Lambda} \\ &\quad + \mu^2\mathbf{\Lambda}\left(\sigma_\eta^2 + \mathrm{tr}[\mathbf{\Lambda}\widetilde{\mathbf{K}}(n)]\right). \end{aligned} \tag{5.53}$$

The off-diagonal elements of (5.53) evolve according to

$$\widetilde{k}_{i,j}(n+1) = \left(1 - \mu(\lambda_i + \lambda_j) + 2\mu^2\lambda_i\lambda_j\right)\widetilde{k}_{i,j}(n), \quad i \neq j, \tag{5.54}$$

and the diagonal elements of $\widetilde{\mathbf{K}}(n)$ evolve according to

$$\begin{aligned} \widetilde{k}_{i,i}(n+1) &= \left(1 - 2\mu\lambda_i + 2\mu^2\lambda_i^2\right)\widetilde{k}_{i,i}(n) \\ &\quad + \mu^2\lambda_i\left(\sigma_\eta^2 + \sum_{l=1}^{L}\lambda_l\widetilde{k}_{l,l}(n)\right). \end{aligned} \tag{5.55}$$

We defer a discussion of the conditions for the convergence of the sequences in (5.54) and (5.55) to Section 5.4.7. Howver, it is not difficult to see that if the system of equations in (5.54) converges, the off-diagonal elements of $\widetilde{\mathbf{K}}(n)$ converge to zero values.

Misadjustment for Gaussian Input Signals. Although the behavior of the evolution equation in (5.53) for a general set of eigenvalues $\{\lambda_0,\ \lambda_1,\ \cdots,\ \lambda_{L-1}\}$ is difficult to summarize, we can determine a closed-form expression for the misadjustment of the LMS adaptive filter from this equation. Define a vector of eigenvalues as

$$\mathbf{L} \;=\; [\lambda_0\ \lambda_1\ \cdots\ \lambda_{L-1}]^T. \tag{5.56}$$

Let us also define a vector sequence $\{\widetilde{\mathbf{S}}(n)\}$ containing only the diagonal elements of $\widetilde{\mathbf{K}}(n)$ as

$$\widetilde{\mathbf{S}}(n) \;=\; [E\{(\widetilde{v}_0(n))^2\}\ E\{(\widetilde{v}_1(n))^2\}\ \cdots\ E\{(\widetilde{v}_{L-1}(n))^2\}]^T. \tag{5.57}$$

Then, the excess MSE can be expressed as

$$\begin{aligned}\xi_{ex}(n) &= \mathrm{tr}[\mathbf{R_{xx}}\mathbf{K}(n)] \\ &= \mathrm{tr}[\mathbf{\Lambda}\widetilde{\mathbf{K}}(n)] \\ &= \mathbf{L}^T\widetilde{\mathbf{S}}(n).\end{aligned} \tag{5.58}$$

Thus, the diagonal elements of $\widetilde{\mathbf{K}}(n)$ contained in $\widetilde{\mathbf{S}}(n)$ are all that are necessary to determine the excess MSE $\xi_{ex}(n)$.

It can be shown using (5.55) that $\widetilde{\mathbf{S}}(n)$ evolves according to

$$\widetilde{\mathbf{S}}(n+1) \;=\; \mathbf{A}\widetilde{\mathbf{S}}(n) + \mathbf{B}, \tag{5.59}$$

where the matrix $\mathbf{A}$ and the vector $\mathbf{B}$ are given by

$$\mathbf{A} \;=\; \mathbf{I} - 2\mu\mathbf{\Lambda} + \mu^2(2\mathbf{\Lambda}^2 + \mathbf{L}\mathbf{L}^T) \tag{5.60}$$

and

$$\mathbf{B} \;=\; \mu^2\sigma_\eta^2\mathbf{L}, \tag{5.61}$$

respectively. Assuming that the system of equations in (5.59) converges, the steady-state value of $\widetilde{\mathbf{S}}(n)$ can be found by setting $\widetilde{\mathbf{S}}(n+1) = \widetilde{\mathbf{S}}(n) = \widetilde{\mathbf{S}}_{ss}$ in (5.59) and solving for $\widetilde{\mathbf{S}}_{ss}$. The solution is

$$\begin{aligned}\widetilde{\mathbf{S}}_{ss} &= (\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \\ &= \mu\sigma_\eta^2\left(2\mathbf{\Lambda} - \mu(2\mathbf{\Lambda}^2 + \mathbf{L}\mathbf{L}^T)\right)^{-1}\mathbf{L}.\end{aligned} \tag{5.62}$$

It is left as an exercise for the reader to show that

$$[\widetilde{\mathbf{S}}_{ss}]_i \;=\; \sigma_\eta^2\left(\frac{\mu/(1-\mu\lambda_i)}{2 - \displaystyle\sum_{m=1}^{L}\mu\lambda_m/(1-\mu\lambda_m)}\right). \tag{5.63}$$

The derivation of (5.63) requires the use of the matrix inversion lemma[2] and is relatively straightforward. The steady-state excess MSE is given by

$$\begin{aligned}\xi_{ex,ss} &= \mathbf{L}^T\widetilde{\mathbf{S}}_{ss} \\ &= \sigma_\eta^2 \frac{\sum_{i=1}^{L} \mu\lambda_i/(1-\mu\lambda_i)}{2-\sum_{m=1}^{L} \mu\lambda_m/(1-\mu\lambda_m)}. \end{aligned} \tag{5.64}$$

Dividing the expression for the steady-state excess MSE by the noise variance gives the misadjustment for Gaussian input signals as

$$M = \frac{\sum_{i=1}^{L} \mu\lambda_i/(1-\mu\lambda_i)}{2-\sum_{m=1}^{L} \mu\lambda_m/(1-\mu\lambda_m)}. \tag{5.65}$$

In addition, we can use (5.52), (5.63) and the fact that the off-diagonal elements of $\widetilde{\mathbf{K}}_{ss}$ are zero to get an expression for the entries of $\mathbf{K}_{ss} = \lim_{n\to\infty}\mathbf{K}(n)$ as

$$k_{ij,ss} = \frac{\mu\sigma_\eta^2}{2-\sum_{m=1}^{L} \mu\lambda_m/(1-\mu\lambda_m)} \sum_{p=1}^{L} q_{ip}q_{jp}\frac{1}{1-\mu\lambda_p}. \tag{5.66}$$

Since the summation in (5.66) is non-zero in general for all $1 \leq (i,j) \leq L$, the two quantities $v_i(n)$ and $v_j(n)$ are correlated in steady state. However, for small step sizes, the value of the summation is close to zero if $i \neq j$ because the columns of $\mathbf{Q}$ are orthogonal and $1/(1-\mu\lambda_p) \approx 1$ for small values of μ.

Example 5.3: Comparison of Simulated and Analytical Behavior of the LMS Adaptive Filter with Gaussian Signals.

Consider the system described in Example 4.2 driven by correlated Gaussian input and desired response signals. The elements of the matrix $\mathbf{K}(n)$ were estimated by averaging the results of one hundred independent experiments performed using signal sets with the same statistics as those in Example 4.2. The coefficients were initialized using $\mathbf{W}(0) = [3\ 2]^T$ in each experiment.

Figure 5.4 shows the ensemble averages of $v_0^2(n)$ and $v_1^2(n)$ as computed from the simulations as well as $E\{v_0^2(n)\}$ and $E\{v_1^2(n)\}$ from (5.51) for the step size $\mu = 0.04$. Even though the initial mean-squared coefficient errors $E\{(v_0(0))^2\} = 4$ and $E\{(v_1(0))^2\} = 1$ are different, the steady-state values of these terms are the same, as predicted by our analysis. The predicted behavior of the system is quite close to the actual behavior as can be observed from the simulation results.

[2]The matrix inversion lemma is stated as follows: for any matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ such that $\mathbf{A}^{-1}$ and $\mathbf{C}^{-1}$ both exist and $\mathbf{A}+\mathbf{BCD}$ is a valid expression, $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1}$.

Table 5.1: Comparison of excess MSEs for different filter lengths from simulation and as predicted from (5.64) in Example 5.3.

L	*Sim.*	*Eqn. (5.64)*
2	4.74×10^{-5}	5.06×10^{-5}
4	8.62×10^{-5}	1.02×10^{-4}
6	1.51×10^{-4}	1.54×10^{-4}
10	2.54×10^{-4}	2.59×10^{-4}
16	4.35×10^{-4}	4.20×10^{-4}
25	6.78×10^{-4}	6.73×10^{-4}
63	1.85×10^{-3}	1.89×10^{-3}
100	3.37×10^{-3}	3.37×10^{-3}

Figure 5.5 displays the excess MSE from simulations and as predicted by our analysis. Again, the predictions accurately follow the actual behavior in simulation, suggesting that for this step size, the independence assumption is reasonable. We can predict the steady-state excess MSE using (5.64) for this case, where $L = 2$, $\lambda_0 = 1.5$, and $\lambda_1 = 0.5$. The value is $\xi_{ex,ss} = 4.40 \times 10^{-4}$. This result closely matches the time average of the ensemble average of the excess MSE over the range $400 \leq n \leq 500$, which is $\widehat{\xi}_{ex,ss} = 4.43 \times 10^{-4}$.

Figure 5.6(a) and (b) show the excess MSE as obtained from simulations and as predicted by our analysis of the system, respectively, for a step size of $\mu = 0.004$. As would be expected, the system converges more slowly in this case as compared to the previous case. Our analysis is clearly accurate in predicting the behavior of the system. From the simulations, the steady-state excess MSE averaged over the range $4000 \leq n \leq 5000$ is $\widehat{\xi}_{ex,ss} = 3.74 \times 10^{-5}$, which is close to the predicted value of $\xi_{ex,ss} = 4.04 \times 10^{-5}$ from the analysis.

We now explore the effect of the filter length on the steady-state excess MSE for the input and desired response signal statistics in this example. For these simulations, we have chosen several filter lengths in the range $2 \leq L \leq 100$ for the LMS adaptive filter, where we have chosen a step size of $\mu = 0.005$ in each case. We initialize the coefficients of the system as $\mathbf{W}(0) = [\mathbf{W}_{opt}^T \mathbf{0}^T]^T$, where $\mathbf{0}$ is an appropriately-long vector of zeros so that $\mathbf{W}(n)$ is of length L. We then measure the excess MSE from an average of one hundred independent experiments over the range $700 \leq n \leq 800$ in each case. Table 5.1 shows the resulting excess MSEs as well as the predicted values from (5.64). As can be seen from the table entries, the excess MSE increases as the length of the adaptive filter increases. Moreover, the analysis is reasonably accurate in predicting the excess MSEs in these situations.

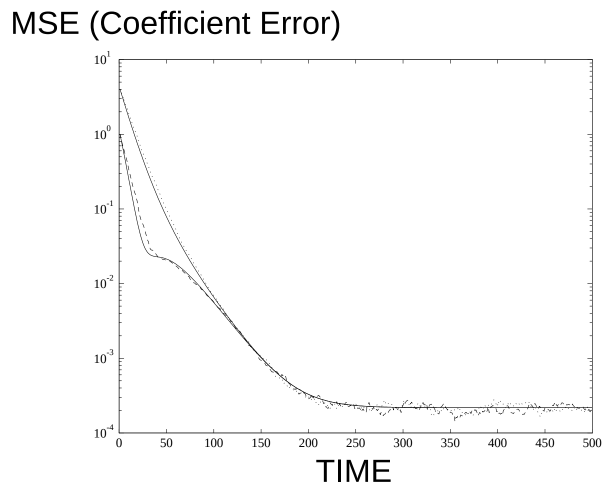

Figure 5.4: Evolution of $E\{(v_1(n))^2\}$ (dotted line) and $E\{(v_2(n))^2\}$ (dashed line) from simulation and theoretical predictions (solid lines) from theory in Example 5.3.

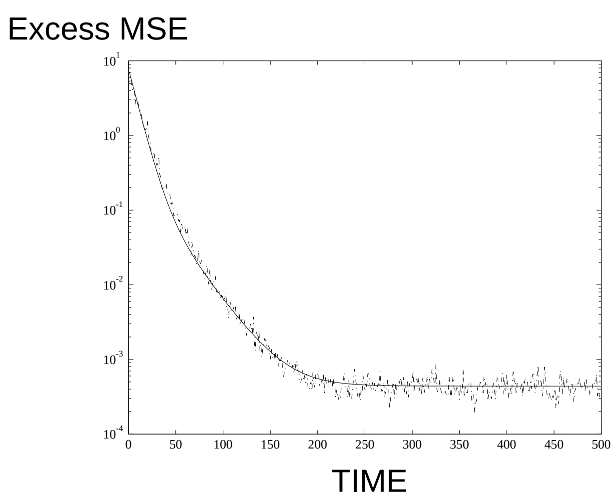

Figure 5.5: Evolution of excess MSE from simulation (dot-dash line) and theory (solid line) for $\mu = 0.04$ in Example 5.3.

The I.I.D. Input Case

For i.i.d. input signals, we find that $\mathbf{R_{xx}} = \sigma_x^2\mathbf{I}$, and thus $\lambda_i = \sigma_x^2$ for all $0 \le i \le L-1$. Moreover, the vectors $\mathbf{V}(n)$ and $\widetilde{\mathbf{V}}(n)$ are the same. We get the following evolution equation for the off-diagonal elements of $\mathbf{K}(n)$ by combining the results of (5.49) and (5.42):

$$k_{i,j}(n+1) = (1 - 2\mu\sigma_x^2 + 2\mu^2\sigma_x^4)k_{i,j}(n), \quad i \neq j. \tag{5.67}$$

The diagonal elements of $\mathbf{K}(n)$ evolve in time as

$$\begin{aligned} k_{i,i}(n+1) &= (1 - 2\mu\sigma_x^2 + \mu^2\kappa\sigma_x^4)k_{i,i}(n) \\ &\quad + \mu^2\sigma_x^2\left(\sigma_\eta^2 + \sigma_x^2\sum_{m=1,m\neq i}^{L} k_{m,m}(n)\right). \end{aligned} \tag{5.68}$$

Note that the updates for the diagonal elements of $\mathbf{K}(n)$ are *uncoupled* from the off-diagonal elements for i.i.d. input signals. For this reason, we can derive an evolution equation for $\mathrm{tr}[\mathbf{K}(n)]$. By summing both sides of (5.68) over all values of i, we can show that

$$\mathrm{tr}[\mathbf{K}(n+1)] = \Big(1 - 2\mu\sigma_x^2 + \mu^2\sigma_x^4(L-1+\kappa)\Big)\,\mathrm{tr}[\mathbf{K}(n)] + \mu^2\sigma_x^2\sigma_\eta^2 L. \tag{5.69}$$

Misadjustment for I.I.D. Input Signals. The excess MSE for i.i.d. input signals is given by

$$\xi_{ex}(n) = \sigma_x^2\mathrm{tr}[\mathbf{K}(n)]. \tag{5.70}$$

Since the excess MSE depends only on $\mathrm{tr}[\mathbf{K}(n)]$, we can write a scalar equation that describes the behavior of $\xi_{ex}(n)$ directly using (5.69) as

$$\begin{aligned} \xi_{ex}(n+1) &= \sigma_x^2\mathrm{tr}[\mathbf{K}(n)] \\ &= \Big(1 - 2\mu\sigma_x^2 + \mu^2\sigma_x^4(L-1+\kappa)\Big)\,\sigma_x^2\mathrm{tr}[\mathbf{K}(n)] + \mu^2\sigma_x^4\sigma_\eta^2 L \\ &= \Big(1 - 2\mu\sigma_x^2 + \mu^2\sigma_x^4(L-1+\kappa)\Big)\,\xi_{ex}(n) + \mu^2\sigma_x^4\sigma_\eta^2 L. \end{aligned} \tag{5.71}$$

We can find the steady-state value of the excess MSE from (5.71) by setting $\xi_{ex}(n+1) = \xi_{ex}(n) = \xi_{ex,ss}$ and solving. This operation assumes that $\xi_{ex}(n)$, as described by the above evolution equation, converges. The steady-state value of the excess MSE is given by

$$\xi_{ex,ss} = \frac{\mu\sigma_x^2\sigma_\eta^2 L}{2 - \mu(L-1+\kappa)\sigma_x^2}. \tag{5.72}$$

The misadjustment can now be evaluated as

$$M = \frac{\mu\sigma_x^2 L}{2 - \mu(L-1+\kappa)\sigma_x^2}. \tag{5.73}$$

Example 5.4: Mean-Square Behavior of the LMS Adaptive Filter with I.I.D. Input Signals

We examine the mean-square behavior of the LMS algorithm for i.i.d. inputs in this example. We choose a binary input signal with probability density function given by

$$f_X(x(n)) = \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5. \end{cases}$$

For this input, $E\{x^M(n)] = 1$ if M is even and is zero otherwise. Hence, for the mean-square analysis equation in (5.71), we see that $\sigma_x^2 = \kappa = 1$ for this distribution.

The desired response signal is the corrupted output of a nine-tap FIR system whose input is the same as that of the adaptive filter and is given by

$$\begin{aligned} d(n) &= 0.2(x(n) + x(n-8)) + 0.4(x(n-1) + x(n-7) + 0.6(x(n-2) + x(n-6)) \\ &\quad + 0.8(x(n-3) + x(n-5)) + x(n-4) + \eta(n), \end{aligned}$$

where $\{\eta(n)\}$ is a white Gaussian noise sequence with variance $\sigma_\eta^2 = 0.01$. These coefficients were chosen arbitrarily, and they only enter into the mean square analysis in the calculation of the initial coefficient error vector. For illustrative purposes, we choose $\mathbf{V}(0) = [1\ 1\ \cdots\ 1]^T$ and $\mu = 0.1$ for each run.

Figure 5.7 compares the behavior of the trace of the coefficient correlation matrix $\mathrm{tr}[\mathbf{K}(n)]$, and the excess MSE as found from (5.69) and (5.71), respectively, and as obtained from ensemble averages of one hundred independent experiments of the system. Because $\sigma_x^2 = 1$ for this input signal, the two curves closely follow one another, and both closely follow the theoretical predictions of their behavior as determined from equation (5.71) with $\xi_{ex}(0) = 9$. In steady-state, the excess MSE and trace of $\mathbf{K}(n)$ are estimated to be 0.00823 and 0.00824, respectively, close to the predicted value of 0.00818 as determined from (5.72). These estimates have been computed by averaging the last one hundred values of the ensemble average curves shown in the figure.

5.4.5 Misadjustment for Small Step Sizes and Arbitrary Signals

The assumption that the input signal is either a Gaussian or an i.i.d. random process allows us to develop expressions for the evolution of the elements of $\mathbf{K}(n)$ and the misadjustment of the algorithm. In most practical uses of the LMS adaptive filter, the step size employed is small such that $0 < \mu\lambda_{max} \ll 1$. We now develop an expression for the misadjustment when the step size satisfies the above constraint. This result is valid for input signals with arbitrary distributions and is also very simple to characterize.

Consider the evolution equation for $\mathbf{K}(n)$ in (5.42). Setting $\mathbf{K}(n+1) = \mathbf{K}(n) = \mathbf{K}_{ss}$ yields the following equation that defines the steady-state value of the matrix $\mathbf{K}(n)$:

$$\mathbf{R_{XX}}\mathbf{K}_{ss} + \mathbf{K}_{ss}\mathbf{R_{XX}} - \mu E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}_{ss}\mathbf{X}(n)\mathbf{X}^T(n)\} = \mu\sigma_\eta^2\mathbf{R_{XX}}. \tag{5.74}$$

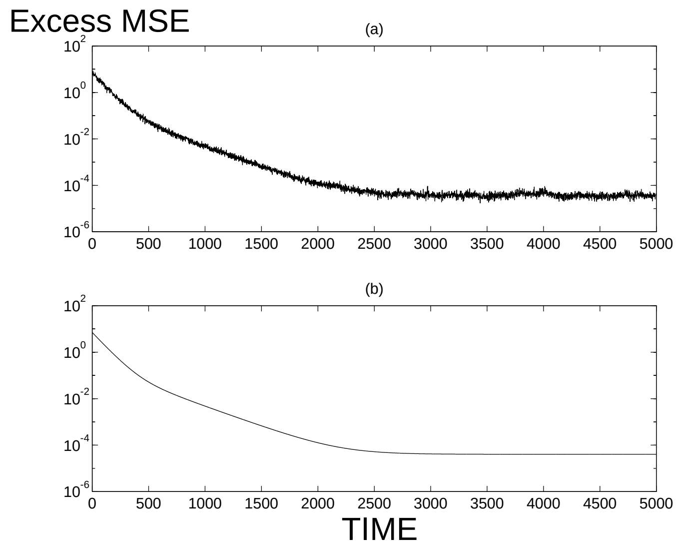

Figure 5.6: Evolution of excess MSE from (a) simulation and (b) theory for $\mu = 0.004$ in Example 5.3.

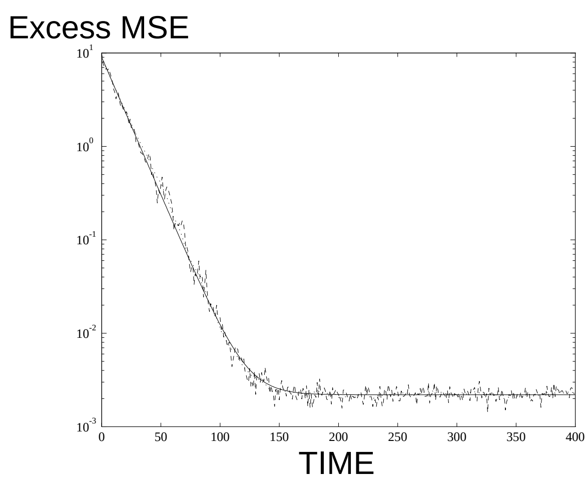

Figure 5.7: Evolution of tr$[\mathbf{K}(n)]$ (dotted line) and excess MSE (dashed line) from simulation and their theoretically predicted values (solid line) for Example 5.4.

In the case where the step size μ is a small value, the first two terms on the left-hand-side of (5.74) are much larger than the third term. As a first-order approximation, we neglect the third term. This approximation results in

$$\mathbf{R_{XX}}\mathbf{K}_{ss} + \mathbf{K}_{ss}\mathbf{R_{XX}} = \mu\sigma_\eta^2\mathbf{R_{XX}}. \tag{5.75}$$

The above equation is linear in the elements of $\mathbf{K}_{ss}$. Moreover, since $\mathbf{R_{XX}}$ is symmetric and positive definite, it can be shown that the solution for $\mathbf{K}_{ss}$ is unique. The value of $\mathbf{K}_{ss}$ that satisfies the relation in (5.75) is

$$\mathbf{K}_{ss} = \frac{\mu\sigma_\eta^2}{2}\mathbf{I}. \tag{5.76}$$

In other words, the variances of the coefficient vector elements are all approximately the same and are given by $\mu\sigma_\eta^2/2$. Moreover, no matter what the statistics of the input signal are, the elements of the coefficient error vector $\mathbf{V}(n)$ are approximately uncorrelated in the steady-state.

Using (5.36), (5.37), and (5.76), we can derive an expression for the misadjustment for small step sizes as

$$\begin{aligned} M &= \frac{\mu}{2}\mathrm{tr}[\mathbf{R_{XX}}] \\ &= \frac{\mu\sigma_x^2 L}{2}. \end{aligned} \tag{5.77}$$

REMARK 5.2: The earlier derivations that resulted in the misadjustment expressions in (5.65) and (5.73) used the Gaussian or i.i.d. input signals assumptions, respectively, to evaluate the quantity $E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$. This term was dropped from (5.74) to derive the misadjustment expression for small step sizes. Consequently, the result in (5.77) is valid for all signal distributions. The reader should also verify that (5.77) can be derived from (5.65) as well as from (5.73) if the step size is sufficiently small.

Example 5.5: Mean-Square Behavior of the LMS Adaptive Filter for Small Step Sizes

In this example, we compare the steady-state excess MSEs for three LMS adaptive filters adapted using i.i.d. input signals with three different probability distributions. For the first LMS adaptive filter, we choose a binary input signal as described in Example 5.4. For the second adaptive filter, we choose a zero-mean white Gaussian input signal with unity variance. For the third LMS adaptive filter, we choose an i.i.d. zero-mean input signal with unit variance and with probability density function given by

$$f_X(x(n)) = \frac{1}{\sqrt{2}}e^{\sqrt{2}|x(n)|}.$$

The above distribution is known as the *Laplacian* distribution and is often used as a model for the amplitude fluctuations in speech signals. In each case, we generate the desired signal from

Table 5.2: Comparison of misadjustments as predicted from (5.77) for small step sizes, as predicted from (5.73), and as determined via simulation for Example 5.5.

Step Size	*Eqn. (5.77)*	*Binary Input* Eqn. (5.73)	*Binary Input* Sim.	*Gaussian Input* Eqn. (5.73)	*Gaussian Input* Sim.	*Laplacian Input* Eqn. (5.73)	*Laplacian Input* Sim.
0.03	1.35×10^{-1}	1.59×10^{-1}	1.56×10^{-1}	1.65×10^{-1}	1.66×10^{-1}	1.74×10^{-1}	1.64×10^{-1}
0.01	4.50×10^{-2}	4.74×10^{-2}	4.56×10^{-2}	4.79×10^{-2}	4.74×10^{-2}	4.87×10^{-2}	4.89×10^{-2}
0.003	1.35×10^{-2}	1.37×10^{-2}	1.43×10^{-2}	1.38×10^{-2}	1.24×10^{-2}	1.38×10^{-2}	1.39×10^{-2}
0.001	4.50×10^{-3}	4.52×10^{-3}	4.52×10^{-3}	4.53×10^{-3}	4.24×10^{-3}	4.53×10^{-3}	4.42×10^{-3}
0.0003	1.35×10^{-3}	1.35×10^{-3}	1.29×10^{-3}	1.35×10^{-3}	1.32×10^{-3}	1.35×10^{-3}	1.45×10^{-3}
0.0001	4.50×10^{-4}	4.50×10^{-4}	4.33×10^{-4}	4.50×10^{-4}	4.38×10^{-4}	4.50×10^{-4}	4.09×10^{-4}

the input signal using the nine-coefficient FIR model in Example 5.4, where $\sigma_\eta^2 = 0.01$. We then compute the steady-state excess MSEs by averaging one hundred iterations of ensemble averages of one hundred independent simulation runs, where we allow enough time for each system to reach its steady state value from an initial coefficient vector of $\mathbf{W}(0) = \mathbf{W}_{opt}$.

Table 5.2 lists the misadjustments for each of the systems when adapted using six different step sizes in the range $0.0001 \leq \mu \leq 0.01$. Also shown in the table are the approximate misadjustments as computed from (5.77) as well as the misadjustments calculated from (5.73), where $\kappa = 1$, 3, and 6 for binary, Gaussian, and Laplacian signals, respectively. We can see that the approximate misadjustment values are reasonably close to both the more-accurate calculated values and the values obtained from simulation for all of the step sizes chosen. Moreover, all of the values are closer for smaller step sizes, indicating that the approximate analysis is indeed more accurate as the step size is reduced.

5.4.6 Summary of of the Mean-Square Analysis for Stationary Input Signals

At this point, we summarize what we have learned about the mean-square convergence behavior of the LMS algorithm for stationary input signals:

- *The mean-square evolution of any one of the coefficient error values $v_i(n)$ consists of a sum of exponentially-converging and/or diverging terms.* This result follows from the facts that both the off-diagonal and diagonal elements of $\widetilde{\mathbf{K}}(n)$ evolve according to first-order difference equations in the elements of $\widetilde{\mathbf{K}}(n)$, from (5.54) and (5.55) or (5.67) and (5.68), and the diagonal elements of $\mathbf{K}(n)$ are simply linear combinations of the elements of $\widetilde{\mathbf{K}}(n)$.

- *If all the filter coefficients $\mathbf{W}(n)$ converge in mean-square, the coefficients continue to fluctuate around their optimum values $\mathbf{W}_{opt}$ in the steady-state if the interfering noise $\{\eta(n)\}$ has non-zero variance.* This result follows from (5.63) and (5.76) that show for different cases that the steady-state values of the diagonal elements of $\mathbf{K}(n)$ (or $\widetilde{\mathbf{K}}(n)$) are nonzero if $\sigma_\eta^2 > 0$.

- *If the filter coefficients converge in the mean-square sense, the steady-state variances of the filter coefficients are all the same for small step sizes, irrespective of the statistics of the input signal. Moreover, the coefficient error variances are all proportional to both the step size and the interfering noise variance, and the variations of the filter coefficients are approximately uncorrelated.* These results follows directly from (5.76).
- *The steady-state misadjustment is an increasing function both of the step size and of the filter length.* From (5.77), we see that the dependence of M on both μ and L is linear for small step sizes, and it can be shown from (5.65) and (5.73) that M increases as either μ or L is increased for both Gaussian and i.i.d. input signals, respectively.

5.4.7 Conditions for Mean-Square Convergence

In Section 5.3.3, we derived bounds on the step size that guarantee convergence of the mean coefficient values. We now derive conditions for the convergence of the coefficient error correlation matrix $\mathbf{K}(n)$. Convergence of this matrix is equivalent to convergence of each element of the matrix to a steady-state value. Since any element of this matrix can be expressed as a linear combination of the elements of the transformed coefficient error correlation matrix $\widetilde{\mathbf{K}}(n)$, it is sufficient to consider the convergence of the elements of $\widetilde{\mathbf{K}}(n)$. Convergence of $\mathbf{K}(n)$ also ensures the convergence of the mean-squared estimation error sequence under the independence assumptions.

It is easy to see from (5.54) and (5.67) that if all the terms $\{\widetilde{k}_{i,j}(n),\ i \neq j\}$ converge, then the steady-state value of these terms is zero. Furthermore, since

$$|E\{\widetilde{v}_i(n)\widetilde{v}_j(n)\}| \quad \leq \quad \sqrt{E\{(\widetilde{v}_i(n))^2\}E\{(\widetilde{v}_j(n))^2\}} \tag{5.78}$$

by the Schwartz inequality, the off-diagonal elements of $\widetilde{\mathbf{K}}(n)$ converge if the diagonal elements of $\widetilde{\mathbf{K}}(n)$ converge. Consequently, we only need to consider the conditions for convergence of the diagonal elements of the transformed coefficient error correlation matrix. We consider the Gaussian and i.i.d. input cases separately.

Convergence Conditions for Gaussian Inputs

Recall from (5.59) that the vector $\widetilde{\mathbf{S}}(n)$ evolves in time according to

$$\widetilde{\mathbf{S}}(n+1) \quad = \quad \mathbf{A}\widetilde{\mathbf{S}}(n) + \mathbf{B}, \tag{5.79}$$

where the matrix $\mathbf{A}$ and vector $\mathbf{B}$ are as defined in (5.60) and (5.61), respectively. The vector sequence $\widetilde{\mathbf{S}}(n)$ converges to the steady-state value $\widetilde{\mathbf{S}}_{ss}$ given by (5.62) or (5.63) if and only if all the eigenvalues of the matrix $\mathbf{A}$ are bounded in magnitude by one. Some relatively straightforward but somewhat lengthy calculations will show that the above conditions correspond to the following conditions on the step size given by [Feuer 1985]

$$0 \;<\; \mu \;<\; \frac{1}{\lambda_{max}} \tag{5.80}$$

and

$$\sum_{i=0}^{L-1} \frac{\mu\lambda_i}{1-\mu\lambda_i} < 2. \tag{5.81}$$

The conditions are both necessary and sufficient to guarantee mean-square convergence of the adaptive filter under the independence assumption. Unfortunately, the bounds are not easily evaluated since the eigenvalues of the autocorrelation matrix are unknown in most situations. Moreover, the conditions above do not lend themselves to an intuitive understanding of the convergence properties of the LMS adaptive filter. Consequently, we will now derive an easily-evaluated sufficient (but not necessary) condition for the mean-square convergence of the LMS adaptive filter.

A Sufficient Condition for Mean-Square Convergence for Gaussian Inputs

We can express $\mathbf{A}$ in (5.60) as

$$\mathbf{A} \;\;=\;\; (\mathbf{I}-\mu\mathbf{\Lambda})^2 + \mu^2(\mathbf{\Lambda}^2 + \mathbf{L}\mathbf{L}^T). \tag{5.82}$$

Since all the eigenvalues of $\mathbf{\Lambda}$ and all the elements of $\mathbf{L}$ are positive, all of the elements of $\mathbf{A}$ are positive, irrespective of the choice of μ. We can now use two results that are valid for symmetric matrices $\mathbf{A}$ where $a_{ij} > 0$ for all values of i and j [Ortega 1991]

- The matrix $\mathbf{A}$ has only non-negative eigenvalues.
- All of the eigenvalues of $\mathbf{A}$ are guaranteed to be less than one if, for all $0 \leq i \leq L-1$,

$$\sum_{j=0}^{L-1} a_{ij} \;\;<\;\; 1. \tag{5.83}$$

We can use the condition in (5.83) to determine a sufficient bound on μ to guarantee stability of the evolution equation in (5.79). Substituting $a_{ij} = 1 - 2\mu\lambda_i\delta_{i-j} + \mu^2(2\lambda_i^2\delta_{i-j} + \lambda_i\lambda_j)$ into (5.83) gives the condition as

$$1 - 2\mu\lambda_i + \mu^2\lambda_i\left(2\lambda_i + \sum_{j=0}^{L-1}\lambda_j\right) \;\;<\;\; 1; \qquad 0 \leq i \leq L-1. \tag{5.84}$$

Subtracting the left-hand-side of the above inequality from both sides of (5.84) gives the sufficient condition as

$$0 \;<\; \mu\lambda_i\left(2-\mu\left(2\lambda_i+\sum_{j=0}^{L-1}\lambda_j\right)\right); \qquad 0\le i\le L-1. \tag{5.85}$$

Since $\lambda_i > 0$ for all i, this condition is equivalent to

$$0 \;<\; \mu \;<\; \frac{2}{2\lambda_i+\sum_{j=0}^{L-1}\lambda_j}; \qquad 0\le i\le L-1. \tag{5.86}$$

Since (5.86) must hold for all $0 \le i \le L-1$, the most stringent bounds are given by

$$0 \;<\; \mu \;<\; \frac{2}{2\lambda_{max}+\mathrm{tr}[\mathbf{R}_{\mathbf{xx}}]}, \tag{5.87}$$

where we have used the result in (5.18).

The condition in (5.87) depends on λ_{max}, a quantity that is difficult to estimate in practice. However, we can use the upper bound for λ_{max} in (5.18) to develop sufficient bounds on μ to guarantee stability of (5.79). These bounds are

$$0 < \mu < \frac{2}{3\mathrm{tr}[\mathbf{R}_{\mathbf{xx}}]}. \tag{5.88}$$

For stationary input data, these stability bounds can be expressed as

$$0 < \mu < \frac{2}{3L\sigma_x^2}. \tag{5.89}$$

Comparing the sufficient conditions for mean-square convergence in (5.89) with the sufficient conditions for convergence of the mean coefficient values in (5.20), we find that the bounds for mean-square convergence are tighter by a factor of three. This result implies that it is possible for the mean coefficient values of the LMS adaptive filter to converge even though the adaptive filter diverges in the mean-square sense. It is important to choose a step size that assures mean-square convergence, however, because the values of the filter coefficients will not be near their optimal values if the variances of the filter coefficients grow with time.

The upper bounds in both (5.89) and (5.20) decrease as the number of filter coefficients L is increased. While this fact does not imply that the necessary and sufficient conditions for mean-square convergence in (5.80)–(5.81) share this dependence on filter length, we will show that, in the i.i.d. case, the upper bound also decreases with increasing filter length.

Convergence Conditions for I.I.D. Inputs

We can determine conditions to guarantee mean-square convergence for i.i.d. inputs from (5.71). The equation for the excess MSE in (5.71) can be written as

$$\xi_{ex}(n+1) \quad = \quad a\xi_{ex}(n) + b, \tag{5.90}$$

where

$$a \quad = \quad 1 - 2\mu\sigma_x^2 + \mu^2(L-1+\kappa)\sigma_x^4 \tag{5.91}$$

and

$$b \quad = \quad \mu^2\sigma_x^4\sigma_\eta^2 L. \tag{5.92}$$

Thus, we can express $\xi_{ex}(n)$ at any iteration as

$$\xi_{ex}(n) \quad = \quad a^n\xi_{ex}(0) + \frac{1-a^n}{1-a}b. \tag{5.93}$$

The excess MSE $\xi_{ex}(n)$ converges to a finite value if and only if $|a| < 1$. It follows immediately that

$$0 < \mu\sigma_x^2(2 - \mu(L-1+\kappa)\sigma_x^2) < 2 \tag{5.94}$$

is a necessary and sufficient condition for convergence of the adaptive filter coefficients in the mean-square sense. It is left as an exercise for the reader to show that the upper bound in (5.94) is always satisfied. Consequently, we obtain necessary and sufficient conditions for convergence from the lower bound in (5.94) as

$$0 < \mu < \frac{2}{(L-1+\kappa)\sigma_x^2}. \tag{5.95}$$

The upper bound in (5.95) indicates that input signals with a high value of κ require a smaller step size for convergence. Thus, Gaussian input signals ($\kappa = 3$) lead to a smaller step size range for convergence as compared to that for binary input signals ($\kappa = 1$).

Dependence on Input Signal Power

One interesting property of the LMS adaptive filter that arises from all of our analyses is that the upper bounds on μ for convergence in (5.20), (5.89) and (5.95) are inversely proportional to the power of the input signal. In situations where the input signal power is unknown *a priori* or is changing over time, the step size for the adaptive filter should be adjusted according to the power of the input signal to ensure that the algorithm is convergent.

The normalized LMS adaptive filter, discussed in Chapter 6, provides a mechanism for continuously scaling the step size using an estimate of the input signal power.

REMARK 5.3: The conditions for convergence given in (5.89) and (5.95) are derived using the independence assumption. Since this assumption is not accurate for large step sizes, the upper bounds should be viewed as *approximate.* In practice, the actual stability of the LMS adaptive filter is dependent on the probability density function of the input signal samples. True upper bounds on μ to guarantee mean-square stability are generally hard to compute, even with complete knowledge of the input probability density function. Later in this chapter, we discuss this issue further and present one method for deriving exact step size bounds.

Typically, the upper bounds in (5.89) and (5.95) are used as rough initial guidelines as to the choice of a good step size for a particular application. Often, a step size less than 1/10th of the values of these upper limits is chosen. More shall be said about the best step size choice for the LMS adaptive filter in Chapter 7.

5.5 Mean-Square Analysis in a Nonstationary Environment

Adaptive filters are ideally suited to signals whose underlying characteristics change over time. We now extend the analyses presented in the previous sections to one particular nonstationary model. For the analysis, we will use all the assumptions that we employed for the mean-square convergence analyses of Section 5.2. In order to make the analysis tractable, we make an additional assumption on the nonstationarity of the operating environment.

5.5.1 Nonstationarity Model

We assume that the input signal of the adaptive filter is stationary and that the desired response signal is generated as the noisy output of a time-varying system described by

$$d(n) \quad = \quad \mathbf{W}_{opt}^T(n)\mathbf{X}(n) + \eta(n). \tag{5.96}$$

In the above description, $\eta(n)$ is an i.i.d. noise sequence and $\mathbf{W}_{opt}(n)$ is a sequence of optimal coefficient vectors generated from the model

$$\mathbf{W}_{opt}(n+1) \quad = \quad \mathbf{W}_{opt}(n) + \mathbf{M}(n), \tag{5.97}$$

where $\mathbf{M}(n)$ is a sequence of vectors such that

$$E\{\mathbf{M}(n)\} \quad = \quad \mathbf{0} \tag{5.98}$$

and

$$E\{\mathbf{M}(k)\mathbf{M}^T(n)\} \quad = \quad \sigma_m^2\mathbf{I}\delta_{k-n}. \tag{5.99}$$

In other words, the elements of each of the vectors $\mathbf{M}(n)$ have zero mean value and are uncorrelated both within each vector and with elements of other vectors. Furthermore, we assume that $\mathbf{M}(n)$ is uncorrelated with the input vector $\mathbf{X}(n)$ and the noise sequence $\eta(n)$. Note that the nonstationarity in the operating environment is due solely to the time-varying system described by (5.97). The most important property of the above model of nonstationarity is that it characterizes the changes in the optimal coefficients from one instant to the next through the mean and mean-square values of $\mathbf{M}(n)$. The model itself generates a coefficient set $\{\mathbf{W}_{opt}(n)\}$ with unbounded variance as n tends to infinity[3] and consequently is not very realistic. However, it enables a simple statistical analysis and provides reasonable design rules for the LMS adaptive filter operating in nonstationary environments in which the rate of change of the optimal coefficients is bounded in the mean-square sense.

5.5.2 Analysis

The coefficient update equation (4.21) can be expressed using the desired response model in (5.96) as

$$\mathbf{W}(n+1) \;=\; \mathbf{W}(n) + \mu\mathbf{X}(n)(\mathbf{X}^T(n)\mathbf{W}_{opt}(n) + \eta(n) - \mathbf{X}^T(n)\mathbf{W}(n)). \qquad (5.100)$$

As in the previous analyses, define the coefficient error vector $\mathbf{V}(n)$ as

$$\mathbf{V}(n) \;=\; \mathbf{W}(n) - \mathbf{W}_{opt}(n). \qquad (5.101)$$

With this definition, we can subtract $\mathbf{W}_{opt}(n+1)$ from the left-hand-side and $\mathbf{W}_{opt}(n)+\mathbf{M}(n)$ from the right-hand-side of (5.100), respectively, to get

$$\mathbf{V}(n+1) \;=\; (\mathbf{I} - \mu\mathbf{X}(n)\mathbf{X}^T(n))\mathbf{V}(n) + \mu\eta(n)\mathbf{X}(n) - \mathbf{M}(n). \qquad (5.102)$$

The derivations are similar to those in the previous mean-square analysis, and therefore we omit the details here. The evolution equation for the coefficient error correlation matrix is given by

$$\begin{aligned}\mathbf{K}(n+1) \;=\; & \mathbf{K}(n) - \mu\left(\mathbf{R}_{\mathbf{XX}}\mathbf{K}(n) + \mathbf{K}(n)\mathbf{R}_{\mathbf{XX}}\right) + \mu^2\sigma_\eta^2\mathbf{R}_{\mathbf{XX}} + \sigma_m^2\mathbf{I} \\ & + \mu^2 E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}. \qquad (5.103)\end{aligned}$$

As before, we transform the above equation by pre- and post-multiplying both sides of this equation by $\mathbf{Q}^T$ and $\mathbf{Q}$, respectively. The resulting expression for $\widetilde{\mathbf{K}}(n)$ is

$$\begin{aligned}\widetilde{\mathbf{K}}(n+1) \;=\; & \widetilde{\mathbf{K}}(n) - \mu\left(\mathbf{\Lambda}\widetilde{\mathbf{K}}(n) + \widetilde{\mathbf{K}}(n)\mathbf{\Lambda}\right) + \mu^2\sigma_\eta^2\mathbf{\Lambda} + \sigma_m^2\mathbf{I} \\ & + \mu^2 E\{\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\widetilde{\mathbf{K}}(n)\widetilde{\mathbf{X}}(n)\widetilde{\mathbf{X}}^T(n)\}. \qquad (5.104)\end{aligned}$$

[3]Several researchers have overcome this difficulty by using

$$\mathbf{W}_{opt}(n+1) = \alpha\mathbf{W}_{opt}(n) + \mathbf{M}(n), \qquad |\alpha| < 1,$$

as the model for the evolution of the optimal coefficients. The results obtained from this model are similar in character to the results derived using the model in (5.97).

The Gaussian Input Signal Case

When the input signal is Gaussian, we can use (5.50) to evaluate the last term in (5.104). The diagonal elements of (5.104) can be shown to evolve according to

$$\widetilde{\mathbf{S}}(n+1) \quad = \quad \mathbf{A}\widetilde{\mathbf{S}}(n) + \mathbf{B} + \sigma_m^2 \mathbf{1}, \tag{5.105}$$

where $\mathbf{A}$ and $\mathbf{B}$ are as defined in (5.60) and (5.61), respectively, and $\mathbf{1}$ is an L-dimensional vector of ones. The steady-state value of the ith component of $\widetilde{\mathbf{S}}_{ss}$ can be evaluated from (5.105) as

$$[\widetilde{\mathbf{S}}_{ss}]_i \quad = \quad \frac{\mu\sigma_\eta^2 + \sigma_m^2/(\mu\lambda_i)}{1-\mu\lambda_i}\left(\frac{1}{2 - \sum_{m=1}^{L}\mu\lambda_m/(1-\mu\lambda_m)}\right). \tag{5.106}$$

We calculate the steady-state excess MSE for the adaptive filter operating in the nonstationary environment using (5.58) as

$$\begin{aligned}\xi_{ex,ss} \quad &= \quad \mathbf{L}^T\widetilde{\mathbf{S}}_{ss} \\ &= \quad \sum_{i=1}^{L}\frac{\mu\sigma_\eta^2\lambda_i + \sigma_m^2/\mu}{1-\mu\lambda_i}\left(\frac{1}{2 - \sum_{m=1}^{L}\mu\lambda_m/(1-\mu\lambda_m)}\right).\end{aligned} \tag{5.107}$$

Comparing (5.64) with (5.107), we see that the nonstationarity in the operating environment introduces an additional component in the expression for the steady-state excess MSE. This component is known as the *tracking error* or the *lag error*. It is proportional to the power of the nonstationarity σ_m^2. For small values of the step size, the lag error is inversely proportional to the step size μ. The component of the excess MSE that is proportional to the noise power σ_η^2 is known as the *adaptation error*, and it quantifies the excess error due to the stochastic nature of the LMS adaptive filter update strategy.

Dependence of Adaptation and Lag Errors on the Step Size

The dependence of the adaptation and tracking errors on the step size μ differ greatly. Notice that the tracking error increases with decreasing values of the step size. This result is intuitively pleasing for the following reason. Small step sizes lead to small corrections to the coefficient values at each iteration of the adaptive filter. If the step size is so small that the corrections are smaller in magnitude than the changes in the optimal coefficients values, the adaptive filter will not be able to adequately track the changes in the operating environment, and the tracking error will be large. However, since the adaptation error increases with increasing values of the step size, too large of a step size will cause excessive fluctuations in the filter coefficient values around their optimum values. Consequently, the selection of the step size involves a compromise between these two errors. We discuss the selection of the optimum value of μ in nonstationary operating environments in the next section.

The I.I.D. Input Signal Case

For i.i.d. input signals, we can evaluate the expression for the excess MSE in (5.70) using the evolution equation in (5.104) as

$$\begin{aligned} \xi_{ex}(n+1) &= \sigma_x^2 \mathrm{tr}[\widetilde{\mathbf{K}}(n+1)] \\ &= \left(1 - 2\mu\sigma_x^2 + \mu^2\sigma_x^4(L-1+\kappa)\right)\xi_{ex}(n) + (\sigma_m^2 + \mu^2\sigma_x^2\sigma_\eta^2)\sigma_x^2 L. \end{aligned} \tag{5.108}$$

Thus, the steady-state excess MSE in this case is

$$\xi_{ex,ss} = \frac{(\mu\sigma_\eta^2\sigma_x^2 + \sigma_m^2/\mu)L}{2 - \mu\sigma_x^2(L-1+\kappa)}. \tag{5.109}$$

The steady-state characteristics of the excess MSE for this case are similar to that described for the Gaussian case.

The Small Step Size Case for Correlated Inputs

In the case where the step size is small, we can neglect the last term in (5.104) while solving for the steady state value of $\widetilde{\mathbf{K}}(n)$. The resulting steady-state value is seen to be

$$\widetilde{\mathbf{K}}_{ss} = \frac{1}{2}\left(\mu\sigma_\eta^2\mathbf{I} + \frac{\sigma_m^2}{\mu}\mathbf{\Lambda}\right). \tag{5.110}$$

Therefore, the steady state excess MSE is given by

$$\begin{aligned} \xi_{ex,ss} &= \mathrm{tr}[\mathbf{\Lambda}\widetilde{\mathbf{K}}_{ss}] \\ &= \frac{1}{2}\left(\mu\sigma_\eta^2\mathrm{tr}[\mathbf{R_{xx}}] + \frac{\sigma_m^2}{\mu}L\right) \\ &= \frac{1}{2}\left(\mu\sigma_\eta^2\sigma_x^2 + \frac{\sigma_m^2}{\mu}\right)L, \end{aligned} \tag{5.111}$$

where the last result follows from the stationarity of the input signal $x(n)$. In this case, the tracking error, given by $\sigma_m^2 L/(2\mu)$, is exactly inversely proportional to the step size, and the adaptation error, given by $\mu\sigma_\eta^2\sigma_x^2 L/2$, is proportional to the step size. Therefore, a tradeoff similar to that in the previous cases is necessary to ensure a low steady-state excess MSE in this case also.

Example 5.6: Mean-Square Behavior of the LMS Adaptive Filter in a Nonstationary Environment

We examine the mean-square behavior of the LMS adaptive filter for correlated Gaussian inputs in this example. The input signal was generated as a zero-mean stationary Gaussian process using the same IIR filter that was used in Example 4.2. The autocorrelation sequence for this model is

$$r_{xx}(n) \;\; = \;\; 0.5^{|n|}, \qquad -\infty < n < \infty.$$

The desired response signal was generated using a nine-coefficient time-varying FIR filter whose initial coefficient values are given by

$$\mathbf{W}_{opt}(0) = [0.2\ 0.4\ 0.6\ 0.8\ 1.0\ 0.8\ 0.6\ 0.4\ 0.2]^T.$$

The nonstationary model for the desired response filter was chosen as (5.96)–(5.99), where $\sigma_m^2 = 0.001$, and the interfering noise was a zero-mean and white Gaussian process with $\sigma_\eta^2 = 0.01$. For illustrative purposes, we chose $\mathbf{V}(0) = [1\ 1\ \cdots\ 1]^T$ and $\mu = 0.01$ for each experiment.

Figure 5.8 compares the behavior of the excess MSE for step sizes of 0.04, 0.01, and 0.0025, respectively. The simulated curves were calculated from ensemble averages of one hundred independent experiments. The theoretical predictions of performance were calculated using (5.105), where $\xi_{ex}(n) = \mathbf{L}^T\widetilde{\mathbf{S}}(n)$. As can be seen, the theoretical predictions of performance agree quite closely with the simulated behavior of the system for all chosen step sizes.

Interestingly, we see that, unlike the previous stationary case, decreasing the step size value can lead to an *increase* in the excess MSE at steady-state. This effect is due to a loss in tracking capability of the adaptive system due to a small chosen step size. This issue is discussed more fully in Chapter 7.

5.6 Other Analyses of the LMS Adaptive Filter

The analyses of the LMS adaptive filter in the previous sections provide useful insights into the average behavior of this system for stochastic models of the input and desired response signals. These analyses do not accurately predict the behavior of the LMS adaptive filter in all situations, however. For example, the analyses are inaccurate when the step size μ is in the vicinity of the stability bounds as given in (5.20), (5.89), and (5.95) for the various input signal cases. In fact, the LMS adaptive filter can diverge for step sizes within the ranges of these bounds for certain signals. These inaccuracies are due to the independence assumptions made in the analyses as described in Section 5.2

In this section, we describe alternatives to the previous analyses. In some cases, these analyses provide different insights into the behavior of the LMS adaptive filter. It should be

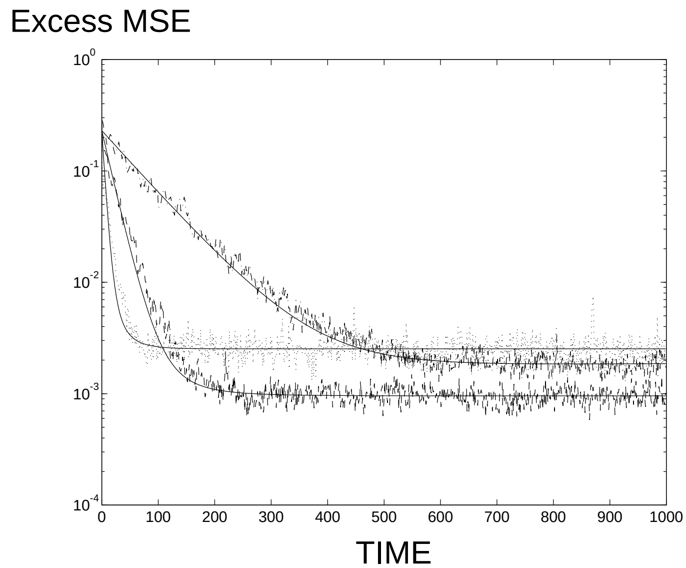

Figure 5.8: Evolution of $\xi_{ex}(n)$ as determined from theory (solid lines) and simulation for step sizes of 0.04 (dotted), 0.01 (dashed), and 0.0025 (dot-dashed) for Example 5.6.

realized, however, that no one analysis provides a complete and accurate description of the LMS adaptive filter in every situation.

5.6.1 An Exact Expectation Analysis

We now describe an extension of our previous analysis that yields an exact statistical description of the LMS adaptive filter for particular types of input and desired response signals. This description can be used to determine exact bounds on the step size to guarantee convergence of the mean-squared error, which for most signals implies that convergence of the filter coefficients to the vicinity of their optimum solutions is guaranteed. The analysis can also be used to provide exact convergence curves of the filter coefficients, the total coefficient error power, and the excess MSE from knowledge of their initial values. The main drawback of the analysis technique is the computational effort needed to derive the description of the system. In general, a computer program that is capable of symbolic manipulation of mathematical expressions is needed to derive the equations for even moderate values of the filter length L.

Assumptions

For this analysis, we assume that $d(n)$ is generated according to the system identification model in (5.30), where $\eta(n)$ is a zero-mean i.i.d. signal. We also assume that $x(n)$ is generated as the output of a linear, time-invariant FIR filter driven by a zero-mean i.i.d. input signal $u(n)$ with unit variance. Figure 5.9 shows the block diagram of the system for the analysis, where $u(n)$ is an i.i.d. signal with a known probability density function $f_u(u)$. Letting $\mathbf{A} = [a_0\ a_1\ \cdots\ a_{M-1}]$ denote the coefficient vector of the filter $A(z)$, we see that

$$x(n) = \sum_{m=0}^{M-1} a_m u(n-m). \tag{5.112}$$

A consequence of this model is that the input signal samples are independent of each other if they are separated by more than M time instants.

The filter $A(z)$ provides a model for the correlation statistics of the input signal $x(n)$. To choose the coefficients of $A(z)$, we note that the autocorrelation of the input signal $r_{xx}(k)$ for $0 \leq k \leq M-1$ can be evaluated as

$$\begin{aligned} r_{xx}(k) &= E\{x(n)x(n-k)\} \\ &= E\left\{\sum_{j=0}^{M-1}\sum_{m=0}^{M-1} a_j a_m u(n-j)u(n-m-k)\right\} \\ &= \sum_{j=0}^{M-1}\sum_{m=0}^{M-1} a_j a_m E\{u(n-j)u(n-m-k)\} \\ &= \sum_{m=0}^{M-k-1} a_m a_{m+k}, \end{aligned} \tag{5.113}$$

where we have used the fact that $E\{u(n-j)u(n-m-k)\} = \delta(j-m-k)$. Equation (5.113) describes the autocorrelation of the input signal in terms of the coefficients of the filter $A(z)$, and these coefficients can be determined from the values of $r_{xx}(k)$ for $0 \leq k \leq M-1$, which are known. In general, finding these coefficients requires the solution of M nonlinear equations in $\{a_0,\ a_1,\ \ldots,\ a_{M-1}\}$. In general, more than one set of coefficients satisfies (5.113).

For the analysis using the above input signal model, complete knowledge of the probability density function of $u(n)$ is not necessary. Define the ith *moment* of the signal $u(n)$ as

$$\gamma_i = E\{x^i(n)\}. \tag{5.114}$$

For the assumptions above, we see that $\gamma_1 = E\{u(n)\} = 0$ and $\gamma_2 = E\{u^2(n)\} = 1$.

Analysis

The exact analysis technique is unlike the previous analyses in that general expressions relating the input and desired response signal statistics to the behavior of an L-coefficient LMS adaptive filter cannot easily be found. Rather, the exact analysis technique is a *systematic procedure* for determining a set of equations that describe the mean or mean-square evolution

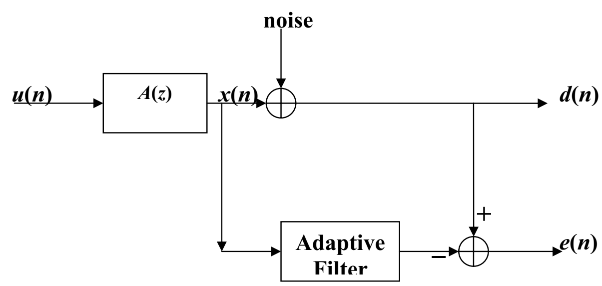

Figure 5.9: Input signal model for the exact expectation analysis.

of the filter coefficients. In general, this procedure must be repeated for both the mean and mean-square analyses for different values of the filter length L and the correlation length M.

We illustrate the procedure for deriving the exact analysis through an example. Consider a single-coefficient filter adapted using the LMS algorithm, in which the input signal is generated from a two-tap correlation filter model, such that $L = 1$ and $M = 2$. While this example is quite simple, it illustrates the derivation procedure. In this case, the LMS adaptive filter is described by

$$w(n+1) = w(n) + \mu e(n)x(n) \quad (5.115)$$
$$e(n) = d(n) - w(n)x(n). \quad (5.116)$$

Substituting the relationship for $d(n)$ in (5.30), we can write the coefficient update equation as

$$w(n+1) = w(n) + \mu(w_{opt} - w(n))x^2(n) + \mu\eta(n)x(n), \quad (5.117)$$

where w_{opt} is the optimum coefficient for the system. Subtracting w_{opt} from both sides of (5.117) gives

$$v(n+1) = (1 - \mu x^2(n))v(n) + \mu\eta(n)x(n), \quad (5.118)$$

where we have defined $v(n) = w(n) - w_{opt}$.

In this example, we wish to find an equation for the exact evolution of the excess MSE $\xi(n)$. The excess MSE in this case is given by

$$\xi(n) = E\{(v(n)x(n))^2\}. \quad (5.119)$$

Moreover, our model for $x(n)$ in (5.112) for $M = 2$ can be written as

$$x(n) = a_0u(n) + a_1u(n-1). \quad (5.120)$$

Since $v(n)$ depends on $x(n-1)$ and $x(n-1)$ is correlated with $x(n)$, $x(n)$ and $v(n)$ are correlated. In the independence assumption analysis, this correlation is assumed to be zero, leading to possibly inaccurate results.

To proceed further, we substitute the expression for $x(n)$ in (5.120) into (5.119). Expanding the result, we get

$$\xi(n) = a_0^2 E\{u^2(n)v^2(n)\} + 2a_0a_1E\{u(n)u(n-1)v^2(n)\} + a_1^2E\{u(n-1)v^2(n)\}. \quad (5.121)$$

The quantity $u(n)$ does not appear in the input signal value $x(n-i)$ for $i > 0$. Because $v(n)$ only depends on $x(n-i)$ and $\eta(n-i)$ for $i > 0$, the sample $u(n)$ is independent of $v(n)$ if the sequence $\{u(n)\}$ is i.i.d. Thus, we can simplify the expressions on the left-hand-side of (5.121) as

$$\begin{aligned}\xi(n) &= a_0^2E\{u^2(n)\}E\{v^2(n)\} + 2a_0a_1E\{u(n)\}E\{u(n-1)v^2(n)\} \\ &\quad + a_1^2E\{u(n-1)v^2(n)\} \\ &= a_0^2\gamma_2E\{v^2(n)\} + a_1^2E\{u(n-1)v^2(n)\}. \end{aligned} \quad (5.122)$$

Thus, the excess MSE in this case is a linear combination of the two quantities $E\{v^2(n)\}$ and $E\{u(n-1)v^2(n)\}$. To calculate $\xi(n)$, we must derive evolution equations for these two quantities. To develop an evolution equation for $E\{v^2(n)\}$, we square both sides of (5.118) and take expectations of both sides of the resulting equation. This operation gives

$$\begin{aligned}E\{v^2(n+1)\} &= E\{(1-\mu x^2(n))^2v^2(n)\} + 2E\{\mu\eta(n)(x(n)-\mu x^3(n))v(n)\} \\ &\quad +\mu^2E\{\eta^2(n)x^2(n)\}. \end{aligned} \quad (5.123)$$

Since the two sequences $\eta(n)$ and $u(n)$ are independent by our assumptions, we see that the last term on the right-hand-side of (5.123) evaluates to $\mu^2\sigma_\eta^2 r_{xx}(0)$. To evaluate the other two terms on the right-hand-side of this equation, we substitute the relationship for $x(n)$ in (5.112) into these terms, expand the expressions, and evaluate the expectations, noting that $u(n)$ is independent of all other quantities appearing in the equation. After some algebra, the resulting expression is found to be

$$\begin{aligned}&E\{v^2(n+1)\} \\ &\quad = (1-2a_0^2\mu\gamma_2 + a_0^4\mu^2\gamma_4)E\{v^2(n)\} + a_1^4\mu^2E\{u^4(n-1)v^2(n)\} \\ &\qquad + (6a_0^2a_1^2\mu^2\gamma_2 - 2a_1^2\mu)E\{u^2(n-1)v^2(n)\} \\ &\qquad + (a_0^2+a_1^2)\mu^2\sigma_\eta^2\gamma_2. \end{aligned} \quad (5.124)$$

The right-hand-side of (5.124) contains the quantities $E\{u^4(n-1)v^2(n)\}$ and $E\{u^2(n-1)v^2(n)\}$. In order to proceed, we also require evolution equations for these quantities. To derive the necessary relationships, we proceed in the same fashion as above. By expressing $E\{u^4(n)v^2(n+1)\}$ and $E\{u^2(n)v^2(n+1)\}$. as functions of quantities at time n, substituting the relationship for $x(n)$ in (5.112) into the results, and evaluating the statistical expectations after recognizing that $u(n)$ is independent of everything else in the equation, we obtain the following expressions:

$$\begin{aligned}
&E\{u^2(n)v^2(n+1)\} \\
&\quad= (\gamma_2 - 2a_0^2\mu\gamma_4 + a_0^4\mu^2\gamma_6)E\{v^2(n)\} + a_1^4\mu^2\gamma_2 E\{u^4(n-1)v^2(n)\} \\
&\qquad + (6a_0^2a_1^2\mu^2\gamma_4 - 2a_1^2\mu\gamma_2)E\{u^2(n-1)v^2(n)\} \\
&\qquad + (a_0^2\mu^2\sigma_n^2\gamma_4 + a_1^2\mu^2\sigma_n^2\gamma_2^2)
\end{aligned} \tag{5.125}$$

and

$$\begin{aligned}
&E\{u^4(n)v^2(n+1)\} \\
&\quad= (\gamma_4 - 2a_0^2\mu\gamma_6 + a_0^4\mu^2\gamma_8)E\{v^2(n)\} + a_1^4\mu^2\gamma_4 E\{u^4(n-1)v^2(n)\} \\
&\qquad + (6a_0^2a_1^2\mu^2\gamma_6 - 2a_1^2\mu\gamma_4)E\{u^2(n-1)v^2(n)\} \\
&\qquad + (a_0^2\mu^2\sigma_n^2\gamma_6 + a_1^2\mu^2\sigma_n^2\gamma_2\gamma_4).
\end{aligned} \tag{5.126}$$

Let us define a state vector $\mathbf{\Phi}(n)$ as

$$\mathbf{\Phi}(n) = \begin{bmatrix} E\{v^2(n)\} \\ E\{u^2(n-1)v^2(n)\} \\ E\{u^4(n-1)v^2(n)\} \end{bmatrix}. \tag{5.127}$$

The evolution equations in (5.124)-(5.126) can be written in matrix form using the state vector as

$$\mathbf{\Phi}(n+1) = \mathbf{A}_e\mathbf{\Phi}(n) + \mathbf{B}_e, \tag{5.128}$$

where the matrix $\mathbf{A}_e$ and vector $\mathbf{B}_e$ are given by

$$\mathbf{A}_e = \begin{bmatrix} 1 - 2a_0^2\mu\gamma_2 + a_0^4\mu^2\gamma_4 & 6a_0^2a_1^2\mu\gamma_2 - 2a_1^2\mu & a_1^4\mu^2 \\ \gamma_2 - 2a_0^2\mu\gamma_4 + a_0^4\mu^2\gamma_6 & 6a_0^2a_1^2\mu^2\gamma_4 - 2a_1^2\mu\gamma_2 & a_1^4\mu^2\gamma_2 \\ \gamma_4 - 2a_0^2\mu\gamma_6 + a_0^4\mu^2\gamma_8 & 6a_0^2a_1^2\mu^2\gamma_6 - 2a_1^2\mu\gamma_4 & a_1^4\mu^2\gamma_4 \end{bmatrix} \tag{5.129}$$

and

$$\mathbf{B}_e = \begin{bmatrix} a_0^2\mu^2\sigma_n^2\gamma_2 + a_1^2\mu^2\sigma_n^2\gamma_2 \\ a_0^2\mu^2\sigma_n^2\gamma_4 + a_1^2\mu^2\sigma_n^2\gamma_2^2 \\ a_0^2\mu^2\sigma_n^2\gamma_6 + a_1^2\mu^2\sigma_n^2\gamma_2\gamma_4 \end{bmatrix}, \tag{5.130}$$

respectively.

Equation (5.128) provides an exact description of the evolution of the mean-square error for this single-coefficient adaptive filter for $M = 2$. From this equation, we can evaluate the sequence of vectors $\mathbf{\Phi}(n)$ given numerical values of a_0, a_1, γ_2, γ_4, γ_6, γ_8, σ_η^2, μ, and the elements of $\mathbf{\Phi}(0)$. By performing an eigenvalue decomposition of the matrix $\mathbf{A}_e$, we can also numerically determine the range of step sizes that guarantee that all of the eigenvalues of $\mathbf{A}_e$ are less than one, producing exact stability conditions on the step size for this system. To find the steady-state value of the excess MSE, we solve (5.128) for the case $\mathbf{\Phi}(n+1) = \mathbf{\Phi}(n) = \mathbf{\Phi}_{ss}$ and substitute the resulting steady-state expressions for $E\{v^2(n)\}$ and $E\{u^2(n-1)v^2(n)\}$ in (5.122).

For general L and M, the procedure to generate the state vector $\mathbf{\Phi}(n)$ is analogous to the case given above. Whenever L and M are finite, the procedure produces a finite number of equations that are propagated using a linear matrix equation, as in (5.128). However, the number and form of the elements of $\mathbf{\Phi}(n)$ are different for each pair $\{L, M\}$.

Example 5.7: Comparison of the Exact Analysis with the Analysis Using the Independence Assumption

In this example, we compare the behaviors of the LMS adaptive filter as predicted by the analyses using the exact method and the independence assumption, respectively. For this example, consider a two-tap FIR filter whose input signal is generated from a two-tap correlation filter $A(z)$ with coefficients

$$a_0 = \sqrt{\frac{1+\sqrt{1-4\rho^2}}{2}} \tag{5.131}$$

$$a_1 = \frac{\rho}{a_0} \tag{5.132}$$

with a zero mean, white Gaussian signal with unit-variance as its input. Here, we restrict the value of the *correlation parameter* ρ to the range $-0.5 < \rho < 0.5$. Using (5.113), it can be shown that the signal $x(n)$ produced from this filter is stationary zero-mean Gaussian with autocorrelation sequence

$$r_{xx}(k) = \begin{cases} 1, & k = 0 \\ \rho, & |k| = 1 \\ 0 & \text{otherwise,} \end{cases}$$

We generate the desired response signal as

$$d(n) = x(n) + x(n-1) + \eta(n),$$

where $\eta(n)$ is a zero-mean white Gaussian signal with $\sigma_\eta^2 = 0.01$. We can generate the evolution equations necessary for the exact analysis using the procedure outlined above. The number of equations needed to describe the evolution of the excess MSE in this case is 48. These equations were generated using a program written in the MAPLE symbolic manipulation language. For details on this program, the reader is referred to [Douglas 1995]. We omit the explicit expressions for brevity.

Figure 5.10 shows the evolution of the excess MSE as predicted by the analyses employing the exact method and the independence assumption for this system for $\mu = 0.2$ and $\rho = 0.5$. Note that this step size is close to the upper stability bound of $2/(3L\sigma_x^2) = 0.33$ as predicted by the

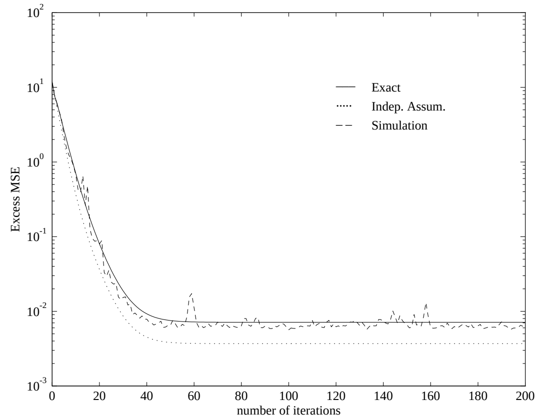

Figure 5.10: Evolution of the excess MSE for $\mu = 0.2$ and $\rho = 0.5$ in Example 5.7.

analysis using the independence assumption. Also shown are simulations obtained by averaging 10000 individual runs of the LMS adaptive filter. Clearly, the exact analysis is more accurate in predicting the actual statistical behavior than the approximate analysis using the independence assumption.

Table 5.3 compares the error in the prediction of the steady-state excess MSE for the exact and approximate analyses for different values of the correlation parameter ρ. The experimental results were evaluated by averaging the steady-state values of the excess MSE over the range $401 \leq n \leq 500$ as obtained from ensemble averages of 10,000 independent experiments. In the table, μ_{max} is the upper step size bound for the system as predicted by the exact analysis. The exact analysis is clearly more accurate than the analyses based on the independence assumption, particularly for values of ρ near 0.5 and larger step size values. Note that the independence assumption and exact analyses are equally-accurate in predicting simulated behavior for step sizes in the range $0 \leq \mu \leq 0.3\mu_{max}$. For step size values approaching μ_{max}, it is difficult to get accurate estimates of the excess MSE from simulations, which results in the discrepancy between the exact analysis and simulation results in these cases.

Comparison of the Exact and Independence Assumption Analyses

The exact analysis offers a unique opportunity to evaluate the accuracy of the approximate analysis using the independence assumption. One chief shortcoming of analyses using the independence assumption is that they are typically inaccurate in predicting the step size range for stable behavior of the LMS adaptive filter. We demonstrate this statement with the following example.

Table 5.3: Comparison of exact and independence assumption analyses in predicting steady-state excess MSEs for the two-tap system described in Example 5.7.

	% Error in Predicting Final EMSE Values, Indep.					*% Error in Predicting Final EMSE Values, Exact*				
step size	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.5$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.5$
$0.1\mu_{max}$	0.25	0.089	−0.057	0.87	−0.11	0.23	0.068	−0.080	0.85	−0.13
$0.2\mu_{max}$	0.61	0.72	0.83	0.24	0.57	0.43	0.54	0.63	0.022	0.34
$0.3\mu_{max}$	0.85	0.77	0.70	0.41	0.54	0.16	0.059	−0.062	−0.42	−0.35
$0.4\mu_{max}$	1.80	1.87	1.93	1.73	1.74	−0.093	−0.085	−0.18	−0.54	−0.68
$0.5\mu_{max}$	4.82	5.01	5.38	5.32	6.10	0.37	0.38	0.36	−0.06	0.35
$0.6\mu_{max}$	10.11	11.20	12.48	11.32	12.87	0.52	1.04	1.41	−0.39	0.33
$0.7\mu_{max}$	21.25	21.32	22.01	18.69	26.71	1.00	−0.078	−1.11	−5.19	−0.042
$0.8\mu_{max}$	85.01	102.7	112.6	41.12	56.02	28.27	36.92	39.19	−10.06	−2.79
$0.9\mu_{max}$	58.34	58.37	63.00	112.1	148.5	−28.96	−32.88	−34.84	−19.16	−8.91

Table 5.4: Maximum step sizes for i.i.d. Gaussian, binary, and Laplacian inputs for different filter lengths, as predicted by the exact and independence assumption analyses.

	Gaussian		*binary*		*Laplacian*	
L	*Indep.*	*Exact*	*Indep.*	*Exact*	*Indep.*	*Exact*
2	0.5000	0.4023	1.000	1.000	0.2857	0.2121
3	0.4000	0.2756	0.6667	0.6667	0.2500	0.1378
4	0.3333	0.2092	0.5000	0.5000	0.2222	0.0962
5	0.2857	0.1674	0.4000	0.4000	0.2000	0.0706

Example 5.8: Comparison of the Exact Analysis with the Analysis Using the Independence Assumption

Using the exact analysis method, we compare the maximum step sizes as predicted by the exact and independence assumption analyses for LMS adaptive filters when the input signal is i.i.d. Gaussian, binary, and Laplacian, respectively, with zero mean and unit variance in each case. Table 5.4 shows these steps sizes for $L = 2$, 3, 4, and 5-tap LMS adaptive filters. The exact analysis predicts a smaller step size bound for the Gaussian and Laplacian input cases, whereas the two analyses predict the same step size bound for binary input. Note that the higher-order moments γ_i of the Laplacian distribution are uniformly larger than those for the Gaussian distribution for even $i \geq 4$, and the binary distribution has all of its even-order moments equal to unity. Thus, the step size bounds predicted by the analysis employing the independence assumption are seen to be inaccurate for "heavy-tailed" input signal distributions that produce large higher-order moments.

From the preceeding example and similar examples in [Douglas 1995], we can infer the following about analyses that use the independence assumption:

- They can accurately describe the mean- and mean-square behavior of the LMS adaptive filter for small step sizes. Typically, it is quite accurate for step sizes between zero and one-tenth of the upper step size bounds predicted by the independence assumption.
- They are inaccurate in predicting upper step size bounds, and the true step size bound in any particular situation is always smaller than that predicted by the independence assumption. Thus, the bounds in (5.20), (5.89), and (5.95) can only serve as guidelines for choosing μ for stable behavior.
- They are inaccurate for large step sizes if the input signal is impulsive or otherwise has large higher-order moments.
- They are inaccurate for large step sizes if the input signal is highly-correlated.
- The analysis using the independence assumption is the most accurate in the situation where the input signal is i.i.d. binary$\{\pm 1\}$-distributed.

5.6.2 Averaging Analysis

In this section, we describe a form of analysis that shares many similarities with the analyses of the previous sections. Popular in the field of adaptive control, the *averaging analysis* technique can be viewed as a generalization of the analyses used in Sections 5.2 and 5.3 that are based on the independence assumption. It can be used in the case where the input and desired response signals do not fit a stochastic model. One example of such a situation is when $x(n)$ and $d(n)$ are periodic signals. As in the analysis employing the independence assumption, an averaging analysis produces only an approximate characterization of the adaptive filter's behavior, and the ability of the analysis to predict the actual behavior of the system depends on the signals being processed and the value of the step size μ.

A General Adaptive System

To develop the concepts used in averaging analysis techniques, we consider a more-general form of the adaptive filter than that given in (4.26) The update for this system is described by

$$\mathbf{\Theta}(n+1) \quad = \quad \mathbf{\Theta}(n) + \mu h(d(n), \mathbf{X}(n), \mathbf{\Theta}(n)), \tag{5.133}$$

where $\mathbf{\Theta}(\mathbf{n})$ describes a *state vector* of the adaptive system at time n and $h(\cdot)$ is a scalar function with three arguments. The state vector $\mathbf{\Theta}(n)$ contains the parameters of the adaptive filter as well as any other quantities that are maintained within the memory of the system, except signal values. As an example, the coefficients in the vector $\mathbf{W}(n)$ can be regarded as the states of the LMS adaptive filter at time n. Thus,

$$\mathbf{\Theta}(n) \quad = \quad [w_0(n)\; w_1(n)\; \cdots\; w_{L-1}(n)]^T \tag{5.134}$$

for this system. The form of $h(\cdot)$ depends on the type of algorithm used to adjust the system's parameters.

Analysis

To analyze the behavior of this system, we make the following assumption:

- *The value of the step size μ is sufficiently small such that the values of the states change slowly with respect to variations in the signals $x(n)$ and $d(n)$.*

This assumption implies that the averaging analysis gives an accurate description of the adaptive system's behavior only for slow adaptation situations. In fact, the averaging analysis is only guaranteed to be accurate when $\mu \to 0$. However, experience has shown that this form of analysis is reasonably accurate for even moderate values of μ.

Under the assumption above, we form the averaged system as

$$\text{avg}\{\mathbf{\Theta}(n+1)\} \quad = \quad \text{avg}\{\mathbf{\Theta}(n)\} + \mu \text{avg}\{h(d(n), \mathbf{X}(n), \text{avg}\{\mathbf{\Theta}(n)\})\}, \tag{5.135}$$

where $\text{avg}\{\cdot\}$ denotes a particular choice of averaging operation. The averaging operation on a vector corresponds to the averaging operation on each of its elements, *i.e.*, $\text{avg}\{\mathbf{U}(n)\} = [\text{avg}\{u_1(n)\} \cdots \text{avg}\{u_N(n)\}]^T$. The form of the averaging operation $\text{avg}\{u(n)\}$ for a signal $u(n)$ varies, but is usually chosen to be one of the following:

1. $\text{avg}\{u(n)\} = \dfrac{1}{N} \displaystyle\sum_{m=0}^{N-1} u(n-m)$

2. $\text{avg}\{u(n)\} = \lim_{N\to\infty} \frac{1}{N}\sum_{m=0}^{N-1} u(n-m)$

3. $\text{avg}\{u(n)\} = E\{u(n)\}$.

The first form of averaging is useful in cases where the signals are both deterministic and periodic. The second form of averaging is useful for slowly varying systems operating on deterministic signals. The third form of averaging assumes that the signals are random processes with known distributions. In fact, this last form of averaging is quite similar to that used in the independence assumption analysis, although no assumptions about the input and desired response signals are explicitly made in this case.

Although no rigorous proofs of the averaging analyses' accuracy have been given here, it has proven to be useful for understanding the behaviors of both continuous- and discrete-time linear and nonlinear systems [Aström 1995]. We now illustrate the method of the averaging analysis through an example.

Example 5.8: Averaging Analysis for an Adaptive Gain Operating on Periodic Signals

We consider a single-coefficient LMS adaptive filter driven by the input and desired response signals $d(n) = Acos(\omega_0 n + \phi)$ and $x(n) = cos(\omega_0 \pi n)$, respectively, where $A = 2$, $\omega_0 = 0.25\pi$, and and $0 \leq \phi < 2\pi$. To study the behavior of this system, we use the first averaging criterion given above, where $N = 2\pi/\omega_0 = 8$. In this case, we find that the averaged equation describing the evolution of $w(n)$ is

$$\text{avg}\{w(n+1)\} = (1 - 0.5\mu)\text{avg}\{w(n)\} + \mu\cos(\phi).$$

Figure 5.11(a) and (b) show the evolutions of the filter coefficients $w(n)$ and the averaged filter coefficient $\text{avg}\{w(n)\}$ for different values of $\phi = n\pi/10$, $n = \{0, 1, 2, 3, 4, 5\}$. In this example, the step size is chosen as $\mu = 0.05$. The average analysis is accurate in predicting the mean behavior of this system for this step size value. The actual value $w(n)$ oscillates with a small amplitude around the value $\text{avg}\{w(n)\}$ in this case for $\phi \neq 0$.

5.6.3 Robustness Analysis of the LMS Adaptive Filter

The LMS adaptive filter is known to behave robustly when approximately implemented using finite-precision calculations. As will become apparent in the following chapters, not all adaptive filters share this robustness property of the LMS adaptive filter. In this section, we indicate why the LMS adaptive filter is a robust processor through a *deterministic analysis* of its properties. This analysis does not make use of the simplifying assumptions and approximations used in previous analyses and is therefore more general. While it does not give specific details about the LMS algorithm's convergence behavior, it provides another useful interpretation of the algorithm's behavior in a general context.

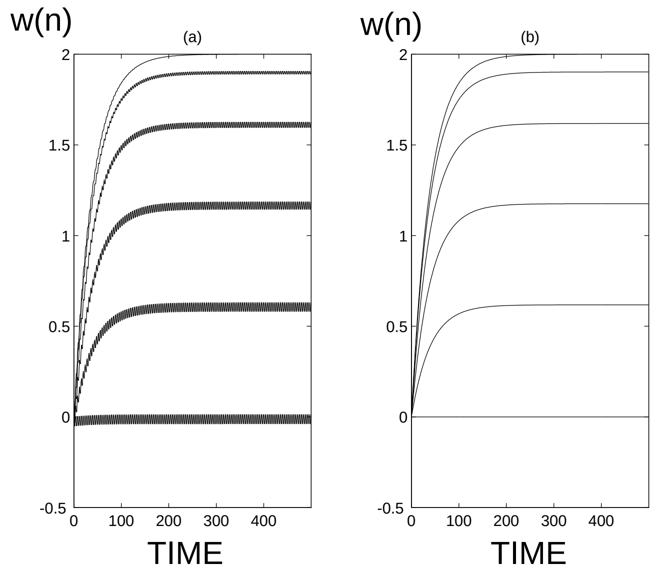

Figure 5.11: Evolution of (a) the filter coefficient $w(n)$ and (b) the predicted behavior from the averaging analysis in Example 5.8.

H_∞ or Minimax Filters

For this analysis, we define the desired response signal $d(n)$ to be of the form

$$d(n) \quad = \quad \mathbf{W}_d^T\mathbf{X}(n) + \eta_d(n), \tag{5.136}$$

where $\mathbf{W}_d$ is a set of L coefficients to be identified and $\eta_d(n)$ is a general noise signal. Note that we make no assumptions on the nature of $\eta_d(n)$. In particular, it may be correlated with the input signal $x(n)$. Define a residual signal as

$$\begin{aligned} e_d(n) &= e(n) - \eta_d(n) \\ &= \mathbf{W}_d^T\mathbf{X}(n) - \mathbf{W}^T(n)\mathbf{X}(n). \end{aligned} \tag{5.137}$$

The residual signal $e_d(n)$ is the portion of the error signal $e(n)$ that can be affected by the adaptive filter coefficients.

Let us form a cost function given by

$$J_d(n) \quad = \quad \frac{\displaystyle\sum_{n=0}^{N} e_d^2(n)}{c + \displaystyle\sum_{n=0}^{N} \eta_d^2(n)}. \tag{5.138}$$

where c is a constant that does not depend on the time-varying filter coefficients $\mathbf{W}(n)$. The cost function $J_d(n)$ depends on the coefficient sequence $\{\mathbf{W}(n)\}$ for $1 \leq n \leq N$. The criterion in (5.138) measures the overall estimation accuracy of the adaptive filter from the initial stages of adaptation. It measures both the system's convergence speed and the quality of its estimates at convergence in some sense.

An H_∞ or *minimax filter of order* L for the sequence triplet $\{d(n), \mathbf{X}(n), \eta_d(n)\}$ for $1 \leq n \leq N$ is the set of coefficients $\{\mathbf{W}(n),\ 1 \leq n \leq N\}$ that minimizes the maximum value of $J_d(n)$ over all possible finite-valued sequences $\eta_d(n)$. Thus, the minimax filter provides good adaptation behavior over the widest possible range of signals. The drawback of such a solution is that the minimax filter may not perform extremely well for any particular set of signals being processed. This tradeoff between the conflicting requirements of *precise estimation* and *robust adaptation* is common in many engineering design problems.

Robustness of the LMS Algorithm

It can be shown [Sayed 1992] for the LMS adaptive filter that

$$\frac{\sum\limits_{n=0}^{N} e_d^2(n)}{\mu^{-1}||\mathbf{W}_d - \mathbf{W}(0)||^2 + \sum\limits_{n=0}^{N} \eta_d^2(n)} \quad \leq \quad 1, \tag{5.139}$$

whenever μ satisfies the bounds

$$0 < \mu \leq \min_{0 \leq n \leq N-1} \frac{1}{||\mathbf{X}(n)||^2} \tag{5.140}$$

for all $0 \leq n \leq N$. Moreover, the LMS adaptive filter is the only adaptation method that satisfies the upper bound in (5.139). Other adaptive filters do not satisfy this bound in their worst-case situations.

To understand the implications of (5.139), we rewrite this inequality as

$$\sum_{n=0}^{N} e_d^2(n) \quad \leq \quad \mu^{-1}||\mathbf{W}_d - \mathbf{W}(0)||^2 + \sum_{n=0}^{N} \eta_d^2(n). \tag{5.141}$$

By dividing both sides by N and taking limits as the window length N tends to infinity, we get

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} e_d^2(n) \quad \leq \quad \frac{1}{N} \lim_{N\to\infty} \sum_{n=0}^{N} \eta_d^2(n) \tag{5.142}$$

for the LMS adaptive filter. In other words, the average of the squared excess error residuals of the LMS adaptive filter is no greater than the average of the squared noise signals. If

$\eta_d(n)$ is a signal with finite power, this result guarantees that the power of the error signal will be finite on average, and the resulting system is always stable, as long as the step size is selected such that (5.140) is satisfied. Therefore, we can expect the LMS adaptive filter to perform properly in a wide variety of situations, even when using finite-precision calculations to implement the system. In Chapter 13, we explore the effects of finite precision calculations of the LMS adaptive filter and indicate its robustness to errors in these calculations through several examples. We explore the choice of step sizes that satisfies the bounds in (5.140) in Chapter 7.

5.7 Main Points of This Chapter

- The evolutions of the mean values of the LMS adaptive filter coefficients are quite similar to the evolutions of the coefficients of the steepest descent algorithm. In particular the evolution of each of the filter coefficients can be decomposed into a sum of exponentially-converging or diverging terms whose speed of convergence is controlled by the eigenvalues of the autocorrelation matrix and the step size.

- When the desired response signal cannot be modeled exactly using the system model and the input signal, the LMS adaptive filter coefficients fluctuate around the optimum coefficients in steady-state.

- The stable range of step sizes for the LMS algorithm can be reasonably characterized as

$$0 < \mu < \frac{K}{L\sigma_x^2}. \tag{5.143}$$

  where K is a constant that depends on the probability distribution of the input signal.

- For stationary system identification tasks, the errors in the filter coefficients are approximately uncorrelated in steady-state for small step size values. Moreover, the variances of these errors are approximately equal and are proportional to the step size μ.

- Time-varying operating environments cause a lag error in the coefficients of the LMS adaptive filter. For the nonstationary model studied in this chapter, the lag error is approximately inversely proportional to the step size.

- Other analyses of the LMS adaptive filter include an exact expectation analysis, the averaging analysis technique, and a robustness analysis for deterministic signals.

- From the exact analysis, the analysis using the independence assumption is seen to be inaccurate if i) the step size is large, ii) the input signal is highly-correlated, and/or iii) the input signal has large higher-order moments $E\{x^i(n)\}$.

- The LMS adaptive filter satisfies a robustness criterion that guarantees well-behaved adaptation under different signal conditions.

5.8 Bibliographical Notes

Mean Analyses of the LMS Adaptive Filter. The mean analysis described in Section 5.3 is similar to other classical viewpoints of the LMS adaptive filter's behavior. For a discussion of these methods, see [Widrow 1975, Widrow 1985]. Similar results can be found in discussions of the method of steepest descent for quadratic error surfaces [Luenberger 1984].

The Independence Assumption. Widrow and Hoff's recognition that the LMS adaptive filter's behavior is similar to that of the method of steepest descent was the first implicit use of the independence assumption [Widrow 1960]. Theoretical justification for the use of the independence assumption was first presented in [Mazo 1972]. One can guarantee that the independence assumption holds in certain contexts by changing the frequency of the LMS coefficient updates [Gardner 1984], although the system usually suffers from poorer convergence behavior in this case. More recent work has focused on analyses that assume M-sample independence of the input signal; see [Eweda 1987] and [Davisson 1984]. [Florian 1986] and [Douglas 1995] present exact statistical analyses that do not require the independence assumption.

Mean-Square Analyses of the LMS Adaptive Filter The early work of Senne [Senne 1968, Horowitz 1980] has motivated the use of Gaussian statistics in the mean-square analysis of adaptive algorithms. More recent works based upon these concepts include [Gardner 1984] and [Feuer 1984]. [Rupp 1993] presents an extension of these methods to spherically-invariant random processes useful in modelling speech and other real-world signals.

The system identification model for the desired response signal is widely-used in mean-square analyses of adaptive filters, although it is not necessary if the input and desired response signals are jointly Gaussian. For an alternative view of this analysis, see [Haykin 1995]. The mean-square analysis for i.i.d. signals is based on the work in [Gardner 1984]. A somewhat different analysis method than those described above is presented in [Slock 1993]. This analysis is about as accurate as the independence assumption analysis in predicting LMS adaptation behavior.

Analysis of the LMS Adaptive Filter in Time-Varying Environments. Our model for the nonstationary behavior of the desired response signals is similar to those used in [Widrow, McCool, Larimore, and Johnson 1976]. For a critique of this model, see [Gardner 1987]. For an alternative study of tracking properties of these algorithms, see [Farden 1981].

Bershad and Macchi have studied the behavior of the LMS algorithm in tracking a chirped sinusoid for adaptive line enhancement [Bershad 1991].

Other Analyses of the LMS Adaptive Filter. The exact expectation analysis method was first presented in [Florian 1986] for i.i.d. input signals, and it was extended to the correlated input signal case in [Douglas 1995]. The two texts by øAström and Wittenmark [øAström 1995] and Anderson *et al* [Anderson 1987] provide excellent discussions of the averaging analysis technique. The robustness analysis of the LMS adaptive filter is presented in [Sayed 1994, Sayed 1996], and extensions of the technique to other algorithms are presented in [Rupp 1996].

The LMS adaptive filter appears in the statistics literature in the context of stochastic approximation theory. For a description of these methods, the reader is referred to the work of Kushner and Clark [Kushner 1978, Kushner 1984]. The desire to understand stochastic gradient algorithms in a general context have led to several useful methods for characterizing the behaviors of these algorithms. Among those are approaches based on ordinary differential equations (ODE) as described in books by Ljung and Söderström [Ljung 1984], Benveniste, Métivier, and Priouret [Benveniste 1987], and Macchi [Macchi 1995]. The work of Bucklew, Kurtz, and Sethares focuses on the comparison of several different stochastic gradient algorithms using probabilistic averaging techniques [Bucklew 1993]. Bershad and Qu study the probability density function of the LMS adaptive filter coefficients [Bershad 1989]. Solo [Solo 1989] provides an alternative analysis of the LMS adaptive filter's behavior to determine the asymptotic dependence of the filter's performance on the step size in a general context.

Although our analyses have focused on stochastic input and desired signal models, it is possible to develop similar understanding using deterministic (usually periodic) signal models. The early work of Glover [Glover 1972] is a good introduction to these methods. For refinements of these ideas, see [Bershad 1995].

5.9 Exercises

5.1. *Mean Behavior of the LMS Adaptive Filter:* Derive (5.14).

5.2. *Autocorrelation Matrix for a Simple Filtering Model:* Verify that the autocorrelation matrix of the input signal in Example 5.2 is given by

$$\mathbf{R_{xx}} = \frac{1}{2}\begin{bmatrix} 1+\epsilon & 1-\epsilon \\ 1-\epsilon & 1+\epsilon \end{bmatrix}.$$

5.3. *Evolution of the Coefficient Error Correlation Matrix for Gaussian Input Signals:* Show via substitution of (5.50) into (5.45) and (5.45) into (5.42) that the evolution equation for $\mathbf{K}(n)$ in (5.42) simplifies to (5.51).

5.4. *Steady-State Variance of the LMS Adaptive Filter Coefficients:* By application of the matrix inversion lemma, derive the expression for the steady-state variance of the ith filter coefficient in (5.63) from (5.62).

5.5 *Dependence of the Misadjustment on the Step Size:* Using the expression in (5.65), prove that the misadjustment of the LMS adaptive filter is an increasing function of the step size μ for Gaussian input signals for step size values satisfying the conditions in (5.80)–(5.81).

5.6. *Bounds on the Step Size for I.I.D. Input Signals:* Show that the upper bound in (5.94) is always satisfied.
Hint: Plot the quadratic argument in μ in the equation for a specific data case, and use your newfound insight to develop a proof.

5.7. *Equivalence of the Steady-State Excess MSE Expressions for White Gaussian Input Signals:* Verify that, for zero-mean and white Gaussian inputs, the expressions for the steady-state excess MSE in (5.64) and (5.72) are identical.

5.8. *Step Size Bounds for Correlated vs. White Gaussian Input Signals:* Prove that (5.95) is never smaller than (5.89) for Gaussian input signals.

5.9. *Comparison of Stability Bounds on the Step Size:* Find examples of situations where the necessary and sufficient conditions for convergence of the mean values of the LMS adaptive filter coefficients is satisfied but not those for convergence of the coefficients in mean-square.

5.10. *Calculating Correlation Statistics for an FIR Filtered Input Signal:* Consider the FIR filter model for the autocorrelation function $r_{xx}(n)$ as given in (5.113).

   a. Find an expression for a_0 in terms of $r_{xx}(0)$ if $M = 1$. Are there multiple possble solutions for the value of a_0? Why?

b. Find general expressions for a_0 and a_1 in term of $r_{xx}(0)$ and $r_{xx}(1)$ if $M = 2$.

c. Find general expressions for a_0, a_1, and a_2 in terms of $r_{xx}(0)$, $r_{xx}(1)$, and $r_{xx}(2)$ if $M = 3$.

d. What can you say about the values of $r_{xx}(n)$ in this model for $|n| > M$?

5.11 *Exact Analysis of a Two-Tap LMS Adaptive Filter With I.I.D. Input Signals:* Using a similar procedure as shown in Section 5.6.1, derive a linear system of equations that describe the exact mean-square behavior of the LMS adaptive filter for a two-tap filter with an i.i.d. input signal, such that $L = 2$ and $M = 1$. How many states in $\mathbf{\Phi}(n)$ are needed in this case?

5.12 *Computing Assignment on the Mean-Square Analysis of the LMS Adaptive Filter for Gaussian Signals.*

a. Write a MATLAB function that evaluates the mean-square analysis equation in (5.51) for Gaussian input and desired response signals. Your function should take as input the step size μ, the initial coefficient error vector $\mathbf{V}(0)$, the autocorrelation matrix $\mathbf{R_{xx}}$, the noise variance σ_η^2, and the number of iterations of the simulation. It should output the total coefficient error power and excess MSE sequences $\mathrm{tr}[\mathbf{K}(n)]$ and $\xi_{ex}(n)$, respectively.

b. Consider a ten-coefficient LMS adaptive filter, where $x(n)$ is generated as

$$x(n) \quad = \quad ax(n-1) + bu(n),$$

where $u(n)$ is a zero-mean and white Gaussian sequence with unit variance and a and b are 0.9 and $\sqrt{0.19}$, respectively. What is the autocorrelation sequence for this input signal? What are the mean-square stability bounds for μ for this signal as predicted by (5.89) in this case?

c. Using the LMS adaptive filter, generate estimates of both $\mathrm{tr}[\mathbf{K}(n)]$ and $\xi_{ex}(n)$ from an average of one hundred separate simulation runs of this system, where $d(n)$ is generated as

$$d(n) \quad = \quad \sum_{i=0}^{9} x(n-i) + \eta(n),$$

and $\eta(n)$ is a zero-mean and white Gaussian sequence with variance $\sigma_\eta^2 = 0.01$. In your simulations, choose μ to be 1/10th the value of the upper bound as predicted in (5.89), and choose the number of iterations to be large enough to give at least 500 samples. Note that

$$\xi_{ex}(n) \quad = \quad E\{(e(n) - \eta(n))^2\}.$$

Consequently, one can use an ensemble average of one hundred separate sequences of $(e(n) - \eta(n))^2$ to estimate the $\xi_{ex}(n)$ sequence. Plot the ensemble average estimates of both the excess MSE and the total coefficient error power sequences on the same graphs as the theoretical predictions of performance, as determined from the program you wrote for part b. How well do the simulated results match the theoretical predictions?

d. Using the expression for the steady-state excess MSE in (5.64), calculate the theoretical prediction of the steady-state excess MSE for the example above. To determine the eigenvalues of $\mathbf{R_{xx}}$, use MATLAB's `eig` function. Then, using an average of 500 values of your ensemble average of the sequence $\xi_{ex}(n)$, compare the resulting estimated value of $\xi_{ex}(n)$ with the calculated value. How close are these values? How well does the analysis predict the actual performance of the system?

e. Repeat parts b. through d. above for the case where $a = 0.5$ and $b = \sqrt{0.75}$. Be sure to choose a new step size that is 1/10th that of the upper bound in (5.89). Is the analysis more accurate or less accurate for this case than for the previous case above?

5.13 *Computing Assignment on the Mean-Square Analysis of the LMS Adaptive Filter for I.I.D. Signals.*

a. Write a MATLAB function that evaluates the mean-square analysis equation in (5.69) for i.i.d. input signals in a system identification task. Your function should take as input the step size μ, the initial coefficient error power $\mathrm{tr}[\mathbf{K}(0)]$, the values of σ_x^2 and κ for the input signal, the noise variance σ_η^2, and the number of iterations of the simulation. It should output the total coefficient error power and excess MSE sequences $\mathrm{tr}[\mathbf{RK}(n)]$ and $\xi_{ex}(n)$, respectively.

b. Repeat parts b and c of Exercise 5.12, with the exception that the input signal $x(n)$ is a zero-mean and white Gaussian sequence with unit variance. What is the value of κ in this case? In your simulations, choose μ to be 1/10th the value of the upper bound as predicted in (5.95), and choose the number of iterations to be large enough to give at least 500 samples of the steady-state excess MSE. How well does the analysis predict the simulated behavior of the system?

c. Using the expression for the steady-state excess MSE in (5.72), calculate the theoretical prediction of the steady-state excess MSE for the example above. Then, using an average of 500 values of your ensemble average of the sequence $\xi_{ex}(n)$, compare the resulting estimated value of $\xi_{ex}(n)$ with the calculated value. How close are these values? How well does the analysis predict the actual performance of the system?

d. Repeat parts b and c above for the case where $x(n)$ is an i.i.d. binary-$\{\pm 1\}$ sequence. You can generate this sequence in MATLAB using the statement `x =`

`2 * (rand(1,numiter)>0.5) - 1`, where `numiter` is the length of the vector `x`. What is the value of κ for this sequence? Be sure to choose a new step size that is 1/10th that of the upper bound in (5.95). Is the analysis more accurate or less accurate for this case than for the previous case above?

5.14 *Computing Assignment on the Exact Analysis of the LMS Adaptive Filter.* This problem explores the nature of the exact analysis for the example provided in Section 5.6.1 when $u(n)$ is a zero-mean white Gaussian signal with unit variance. In this case, the $(2j)$th moment in (5.114) for $j \geq 1$ can be expressed as

$$\gamma_{2j} \quad = \quad 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2i-1).$$

a. For the case $M = 2$, express the values of a_0 and a_1 for the correlation filter $A(z)$ in terms of the input signal power $\sigma_x^2 = r_{xx}(0)$ and the correlation coefficient ρ_x, defined as

$$\rho_x \quad = \quad \frac{r_{xx}(1)}{r_{xx}(0)}.$$

In this case, select a_1 such that $a_0 \geq a_1$ for all values of a_0 in the range $(2)^{-1/2} \leq a_0 \leq 1$.

b. Using MATLAB, determine and plot the maximum eigenvalue of the matrix $\mathbf{A}_e$ given in (5.129) as a function of μ for $\rho_x = \{0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5\}$. For what numerical ranges of μ is the system stable in each case?

c. For comparison, determine and plot the maximum eigenvalue of the matrix $\mathbf{A}$ given in (5.60), and numerically evaluate the ranges of μ that guarantee stability of the evolution equations based on the independence assumption. How different are the maximum eigenvalues curves in each case as compared to those you plotted in part b? How do the step size ranges compare to their exact counterparts?

d. The step size that gives fastest convergence is the step size that minimizes the maximum value of the matrix $\mathbf{A}_e$ in each case. Using MATLAB's `eig` function, find the values of the step sizes that gives fastest convergence for the six cases described above.

e. Generate sequences of 1000 Gaussian random variables using the model in (5.112), and simulate the behavior of a single-tap LMS adaptive filter with this input signal. Choose $w_{opt} = 1$, $w(0) = 0$, $\rho_x = 0.5$ and a step size μ that gives the fastest convergence for this system as indicated in part d. Produce an averaged evolution curve for the quantity $v(n)$ from 100 simulation experiments on different identically-distributed input sequences. Then, plot the predicted evolution curves from the exact analysis and the approximate analysis employing the independence assumption, respectively, on the same plot. How do the simulated and analytical results differ? Which analysis is more accurate?