Scott C. Douglas, et. Al. "Convergence Issues in the LMS Adaptive Filter."
2000 CRC Press LLC. <http://www.engnetbase.com>.

Convergence Issues in the LMS Adaptive Filter

19.1 Introduction
19.2 Characterizing the Performance of Adaptive Filters
19.3 Analytical Models, Assumptions, and Definitions
System Identification Model for the Desired Response Signal • Statistical Models for the Input Signal • The Independence Assumptions • Useful Definitions
19.4 Analysis of the LMS Adaptive Filter
Mean Analysis • Mean-Square Analysis
19.5 Performance Issues
Basic Criteria for Performance • Identifying Stationary Systems • Tracking Time-Varying Systems
19.6 Selecting Time-Varying Step Sizes
Normalized Step Sizes • Adaptive and Matrix Step Sizes • Other Time-Varying Step Size Methods
19.7 Other Analyses of the LMS Adaptive Filter
19.8 Analysis of Other Adaptive Filters
19.9 Conclusions
References

Scott C. Douglas
University of Utah

Markus Rupp
Bell Laboratories
Lucent Technologies

19.1 Introduction

In adaptive filtering, the *least-mean-square (LMS) adaptive filter* [1] is the most popular and widely used adaptive system, appearing in numerous commercial and scientific applications. The LMS adaptive filter is described by the equations

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(n)e(n)\mathbf{X}(n) \tag{19.1}$$

$$e(n) = d(n) - \mathbf{W}^T(n)\mathbf{X}(n)\,, \tag{19.2}$$

where $\mathbf{W}(n) = [w_0(n)\ \ w_1(n)\ \ \cdots\ \ w_{L-1}(n)]^T$ is the coefficient vector, $\mathbf{X}(n) = [x(n)\ \ x(n-1)\ \ \cdots\ \ x(n-L+1)]^T$ is the input signal vector, $d(n)$ is the desired signal, $e(n)$ is the error signal, and $\mu(n)$ is the step size.

There are three main reasons why the LMS adaptive filter is so popular. First, it is relatively easy to implement in software and hardware due to its computational simplicity and efficient use of memory. Second, it performs robustly in the presence of numerical errors caused by finite-precision arithmetic. Third, its behavior has been analytically characterized to the point where a user can easily set up the system to obtain adequate performance with only limited knowledge about the input and desired response signals.

Our goal in this chapter is to provide a detailed performance analysis of the LMS adaptive filter so that the user of this system understands how the choice of the step size $\mu(n)$ and filter length L affect the performance of the system through the natures of the input and desired response signals $x(n)$ and $d(n)$, respectively. The organization of this chapter is as follows. We first discuss why analytically characterizing the behavior of the LMS adaptive filter is important from a practical point of view. We then present particular signal models and assumptions that make such analyses tractable. We summarize the analytical results that can be obtained from these models and assumptions, and we discuss the implications of these results for different practical situations. Finally, to overcome some of the limitations of the LMS adaptive filter's behavior, we describe simple extensions of this system that are suggested by the analytical results. In all of our discussions, we assume that the reader is familiar with the adaptive filtering task and the LMS adaptive filter as described in Chapter 18 of this Handbook.

19.2 Characterizing the Performance of Adaptive Filters

There are two practical methods for characterizing the behavior of an adaptive filter. The simplest method of all to understand is *simulation.* In simulation, a set of input and desired response signals are either collected from a physical environment or are generated from a mathematical or statistical model of the physical environment. These signals are then processed by a software program that implements the particular adaptive filter under evaluation. By trial-and-error, important design parameters, such as the step size $\mu(n)$ and filter length L, are selected based on the observed behavior of the system when operating on these example signals. Once these parameters are selected, they are used in an adaptive filter implementation to process additional signals as they are obtained from the physical environment. In the case of a real-time adaptive filter implementation, the design parameters obtained from simulation are encoded within the real-time system to allow it to process signals as they are continuously collected.

While straightforward, simulation has two drawbacks that make it a poor sole choice for characterizing the behavior of an adaptive filter:

- *Selecting design parameters via simulation alone is an iterative and time-consuming process.* Without any other knowledge of the adaptive filter's behavior, the number of trials needed to select the best combination of design parameters is daunting, even for systems as simple as the LMS adaptive filter.
- *The amount of data needed to accurately characterize the behavior of the adaptive filter for all cases of interest may be large.* If real-world signal measurements are used, it may be difficult or costly to collect and store the large amounts of data needed for simulation characterizations. Moreover, once this data is collected or generated, it must be processed by the software program that implements the adaptive filter, which can be time-consuming as well.

For these reasons, we are motivated to develop an *analysis* of the adaptive filter under study. In such an analysis, the input and desired response signals $x(n)$ and $d(n)$ are characterized by certain properties that govern the forms of these signals for the application of interest. Often, these properties are *statistical* in nature, such as the *means* of the signals or the *correlation* between two signals at different time instants. An analytical description of the adaptive filter's behavior is then developed that is based on these signal properties. Once this analytical description is obtained, the design parameters are selected to obtain the best performance of the system as predicted by the analysis. What is considered "best performance" for the adaptive filter can often be specified directly within the analysis, without the need for iterative calculations or extensive simulations.

Usually, both analysis and simulation are employed to select design parameters for adaptive filters, as the simulation results provide a check on the accuracy of the signal models and assumptions that are used within the analysis procedure.

19.3 Analytical Models, Assumptions, and Definitions

The type of analysis that we employ has a long-standing history in the field of adaptive filters [2]– [6]. Our analysis uses *statistical models* for the input and desired response signals, such that any collection of samples from the signals $x(n)$ and $d(n)$ have well-defined joint probability density functions (p.d.f.s). With this model, we can study the *average behavior* of functions of the coefficients $\mathbf{W}(n)$ at each time instant, where "average" implies taking a statistical expectation over the ensemble of possible coefficient values. For example, the mean value of the ith coefficient $w_i(n)$ is defined as

$$E\{w_i(n)\} = \int_{-\infty}^{\infty} w\ p_{w_i}(w, n) dw \ , \tag{19.3}$$

where $p_{w_i}(w, n)$ is the probability distribution of the ith coefficient at time n. The mean value of the coefficient vector at time n is defined as $E\{\mathbf{W}(n)\} = [E\{w_0(n)\}\ E\{w_1(n)\}\ \cdots\ E\{w_{L-1}(n)\}]^T$.

While it is usually difficult to evaluate expectations such as (19.3) directly, we can employ several simplifying *assumptions* and *approximations* that enable the formation of *evolution equations* that describe the behavior of quantities such as $E\{\mathbf{W}(n)\}$ from one time instant to the next. In this way, we can predict the evolutionary behavior of the LMS adaptive filter on average. More importantly, we can study certain characteristics of this behavior, such as the stability of the coefficient updates, the speed of convergence of the system, and the estimation accuracy of the filter in steady-state. Because of their role in the analyses that follow, we now describe these simplifying assumptions and approximations.

19.3.1 System Identification Model for the Desired Response Signal

For our analysis, we assume that the desired response signal is generated from the input signal as

$$d(n) = \mathbf{W}_{opt}^T \mathbf{X}(n) + \eta(n) \ , \tag{19.4}$$

where $\mathbf{W}_{opt} = [w_{0,opt}\ \ w_{1,opt}\ \ \cdots\ \ w_{L-1,opt}]^T$ is a vector of optimum FIR filter coefficients and $\eta(n)$ is a noise signal that is independent of the input signal. Such a model for $d(n)$ is realistic for several important adaptive filtering tasks. For example, in echo cancellation for telephone networks, the optimum coefficient vector $\mathbf{W}_{opt}$ contains the impulse response of the echo path caused by the impedance mismatches at hybrid junctions within the network, and the noise $\eta(n)$ is the near-end source signal [7]. The model is also appropriate in system identification and modeling tasks such as plant identification for adaptive control [8] and channel modeling for communication systems [9]. Moreover, most of the results obtained from this model are independent of the specific impulse response values within $\mathbf{W}_{opt}$, so that general conclusions can be readily drawn.

19.3.2 Statistical Models for the Input Signal

Given the desired response signal model in (19.4), we now consider useful and appropriate statistical models for the input signal $x(n)$. Here, we are motivated by two typically conflicting concerns: (1) the need for signal models that are realistic for several practical situations and (2) the tractability of the analyses that the models allow. We consider two input signal models that have proven useful for predicting the behavior of the LMS adaptive filter.

Independent and Identically Distributed (I.I.D.) Random Processes

In digital communication tasks, an adaptive filter can be used to identify the dispersive characteristics of the unknown channel for purposes of decoding future transmitted sequences [9]. In this application, the transmitted signal is a bit sequence that is usually zero mean with a small number of amplitude levels. For example, a non-return-to-zero (NRZ) binary signal takes on the values of ± 1 with equal probability at each time instant. Moreover, due to the nature of the encoding of the transmitted signal in many cases, any set of L samples of the signal can be assumed to be *independent* and *identically distributed* (i.i.d.). For an i.i.d. random process, the p.d.f. of the samples $\{x(n_1),\ x(n_2),\ \ldots, x(n_L)\}$ for any choices of n_i such that $n_i \neq n_j$ is

$$p_{\mathbf{x}}\left(x(n_1),\ x(n_2),\ \ldots, x(n_L)\right) = p_x(x(n_1))\ p_x(x(n_2)) \cdots p_x(x(n_L))\ , \tag{19.5}$$

where $p_x(\cdot)$ and $p_{\mathbf{x}}(\cdot)$ are the univariate and L-variate probability densities of the associated random variables, respectively.

Zero-mean and statistically independent random variables are also *uncorrelated*, such that

$$E\{x(n_i)x(n_j)\} = 0 \tag{19.6}$$

for $n_i \neq n_j$, although uncorrelated random variables are not necessarily statistically independent. The input signal model in (19.5) is useful for analyzing the behavior of the LMS adaptive filter, as it allows a particularly simple analysis of this system.

Spherically Invariant Random Processes (SIRPs)

In acoustic echo cancellation for speakerphones, an adaptive filter can be used to electronically isolate the speaker and microphone so that the amplifier gains within the system can be increased [10]. In this application, the input signal to the adaptive filter consists of samples of bandlimited speech. It has been shown in experiments that samples of a bandlimited speech signal taken over a short time period (e.g., 5 ms) have so-called "spherically invariant" statistical properties. *Spherically invariant random processes* (SIRPs) are characterized by multivariate p.d.f.s that depend on a quadratic form of their arguments, given by $\mathbf{X}^T(n)\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{X}(n)$, where

$$\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{X}(n)\mathbf{X}^T(n)\} \tag{19.7}$$

is the L-dimensional *input signal autocorrelation matrix* of the stationary signal $x(n)$. The best-known representative of this class of stationary stochastic processes is the *jointly Gaussian random process* for which the joint p.d.f. of the elements of $\mathbf{X}(n)$ is

$$p_{\mathbf{x}}(x(n),\ \ldots,\ x(n-L+1)) = \left((2\pi)^L \det\left(\mathbf{R}_{\mathbf{xx}}\right)\right)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{X}^T(n)\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{X}(n)\right)\ , \tag{19.8}$$

where $\det(\mathbf{R}_{\mathbf{xx}})$ is the *determinant* of the matrix $\mathbf{R}_{\mathbf{xx}}$. More generally, SIRPs can be described by a weighted mixture of Gaussian processes as

$$\begin{aligned} p_{\mathbf{x}}(x(n),\ \ldots,\ x(n-L+1) \quad &= \quad \int_0^\infty \left((2\pi|u|)^L \det\left(\overline{\mathbf{R}}_{\mathbf{xx}}\right)\right)^{-1/2} \\ &\qquad \times\ p_\sigma(u) \exp\left(-\frac{1}{2u^2}\mathbf{X}^T(n)\overline{\mathbf{R}}_{\mathbf{xx}}^{-1}\mathbf{X}(n)\right)\ du\ , \end{aligned} \tag{19.9}$$

where $\overline{\mathbf{R}}_{\mathbf{xx}}$ is the autocorrelation matrix of a zero-mean, unit-variance jointly Gaussian random process. In (19.9), the p.d.f. $p_\sigma(u)$ is a weighting function for the value of u that scales the standard deviation of this process. In other words, any single realization of a SIRP is a Gaussian random process with an autocorrelation matrix $u^2\overline{\mathbf{R}}_{\mathbf{xx}}$. Each realization, however, will have a different variance u^2.

As described, the above SIRP model does not accurately depict the statistical nature of a speech signal. The variance of a speech signal varies widely from phoneme (vowel) to fricative (consonant) utterances, and this burst-like behavior is uncharacteristic of Gaussian signals. The statistics of such behavior can be accurately modeled if a *slowly varying* value for the random variable u in (19.9) is allowed. Figure 19.1 depicts the differences between a *nearly SIRP* and an SIRP. In this system, either the random variable u or a sample from the slowly varying random process $u(n)$ is created and used to scale the magnitude of a sample from an uncorrelated Gaussian random process. Depending on the position of the switch, either an SIRP (upper position) or a nearly SIRP (lower position) is created. The linear filter $F(z)$ is then used to produce the desired autocorrelation function of the SIRP. So long as the value of $u(n)$ changes slowly over time, $\mathbf{R_{xx}}$ for the signal $x(n)$ as produced from this system is approximately the same as would be obtained if the value of $u(n)$ were fixed, except for the amplitude scaling provided by the value of $u(n)$.

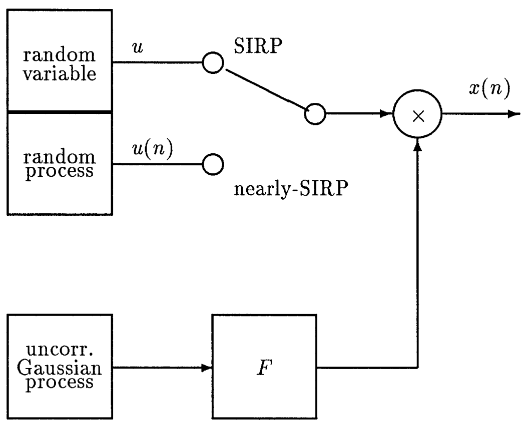

FIGURE 19.1: Generation of SIRPs and nearly SIRPs.

The random process $u(n)$ can be generated by filtering a zero-mean uncorrelated Gaussian process with a narrow-bandwidth lowpass filter. With this choice, the system generates samples from the so-called K_0 p.d.f., also known as the MacDonald function or degenerated Bessel function of the second kind [11]. This density is a reasonable match to that of typical speech sequences, although it does not necessarily generate sequences that sound like speech. Given a short-length speech sequence from a particular speaker, one can also determine the proper $p_\sigma(u)$ needed to generate $u(n)$ as well as the form of the filter $F(z)$ from estimates of the amplitude and correlation statistics of the speech sequence, respectively.

In addition to adaptive filtering, SIRPs are also useful for characterizing the performance of vector quantizers for speech coding. Details about the properties of SIRPs can be found in [12].

19.3.3 The Independence Assumptions

In the LMS adaptive filter, the coefficient vector $\mathbf{W}(n)$ is a complex function of the current and past samples of the input and desired response signals. This fact would appear to foil any attempts to develop equations that describe the evolutionary behavior of the filter coefficients from one time instant to the next. One way to resolve this problem is to make further statistical assumptions about the nature of the input and the desired response signals. We now describe a set of assumptions that have proven to be useful for predicting the behaviors of many types of adaptive filters.

The Independence Assumptions: Elements of the vector $\mathbf{X}(n)$ are statistically independent of the elements of the vector $\mathbf{X}(m)$ if $m \neq n$. In addition, samples from the noise signal $\eta(n)$ are i.i.d. and independent of the input vector sequence $\mathbf{X}(k)$ for all k and n.

A careful study of the structure of the input signal vector indicates that the independence assumptions are never true, as the vector $\mathbf{X}(n)$ shares elements with $\mathbf{X}(n-m)$ if $|m| < L$ and thus cannot be independent of $\mathbf{X}(n-m)$ in this case. Moreover, $\eta(n)$ is not guaranteed to be independent from sample to sample. Even so, numerous analyses and simulations have indicated that these assumptions lead to a reasonably accurate characterization of the behavior of the LMS and other adaptive filter algorithms for small step size values, even in situations where the assumptions are grossly violated. In addition, analyses using the independence assumptions enable a simple characterization of the LMS adaptive filter's behavior and provide reasonable guidelines for selecting the filter length L and step size $\mu(n)$ to obtain good performance from the system.

It has been shown that the independence assumptions lead to a first-order-in-$\mu(n)$ approximation to a more accurate description of the LMS adaptive filter's behavior [13]. For this reason, the analytical results obtained from these assumptions are not particularly accurate when the step size is near the stability limits for adaptation. It is possible to derive an exact statistical analysis of the LMS adaptive filter that does not use the independence assumptions [14], although the exact analysis is quite complex for adaptive filters with more than a few coefficients. From the results in [14], it appears that the analysis obtained from the independence assumptions is most inaccurate for large step sizes and for input signals that exhibit a high degree of statistical correlation.

19.3.4 Useful Definitions

In our analysis, we define the minimum mean-squared error (MSE) solution as the coefficient vector $\mathbf{W}(n)$ that minimizes the mean-squared error criterion given by

$$\xi(n) = E\{e^2(n)\} . \tag{19.10}$$

Since $\xi(n)$ is a function of $\mathbf{W}(n)$, it can be viewed as an *error surface* with a minimum that occurs at the minimum MSE solution. It can be shown for the desired response signal model in (19.4) that the minimum MSE solution is $\mathbf{W}_{opt}$ and can be equivalently defined as

$$\mathbf{W}_{opt} = \mathbf{R}_{\mathrm{xx}}^{-1}\mathbf{P}_{\mathrm{dx}} , \tag{19.11}$$

where $\mathbf{R}_{\mathrm{xx}}$ is as defined in (19.7) and $\mathbf{P}_{\mathrm{dx}} = E\{d(n)\mathbf{X}(n)\}$ is the cross-correlation of $d(n)$ and $\mathbf{X}(n)$. When $\mathbf{W}(n) = \mathbf{W}_{opt}$, the value of the *minimum MSE* is given by

$$\xi_{min} = \sigma_\eta^2 , \tag{19.12}$$

where σ_η^2 is the power of the signal $\eta(n)$.

We define the *coefficient error vector* $\mathbf{V}(n) = [v_0(n) \ \cdots \ v_{L-1}(n)]^T$ as

$$\mathbf{V}(n) = \mathbf{W}(n) - \mathbf{W}_{opt} \ , \tag{19.13}$$

such that $\mathbf{V}(n)$ represents the errors in the estimates of the optimum coefficients at time n. Our study of the LMS algorithm focuses on the statistical characteristics of the coefficient error vector. In particular, we can characterize the approximate evolution of the *coefficient error correlation matrix* $\mathbf{K}(n)$, defined as

$$\mathbf{K}(n) = E\{\mathbf{V}(n)\mathbf{V}^T(n)\} \ . \tag{19.14}$$

Another quantity that characterizes the performance of the LMS adaptive filter is the *excess mean-squared error (excess MSE)*, defined as

$$\begin{aligned} \xi_{ex}(n) &= \xi(n) - \xi_{min} \\ &= \xi(n) - \sigma_\eta^2 \ , \end{aligned} \tag{19.15}$$

where $\xi(n)$ is as defined in (19.10). The excess MSE is the power of the additional error in the filter output due to the errors in the filter coefficients. An equivalent measure of the excess MSE in steady-state is the *misadjustment*, defined as

$$M = \lim_{n\to\infty} \frac{\xi_{ex}(n)}{\sigma_\eta^2} \ , \tag{19.16}$$

such that the quantity $(1 + M)\sigma_\eta^2$ denotes the total MSE in steady-state.

Under the independence assumptions, it can be shown that the excess MSE at any time instant is related to $\mathbf{K}(n)$ as

$$\xi_{ex}(n) = \mathrm{tr}[\mathbf{R_{xx}}\mathbf{K}(n)] \ , \tag{19.17}$$

where the *trace* $\mathrm{tr}[\cdot]$ of a matrix is the sum of its diagonal values.

19.4 Analysis of the LMS Adaptive Filter

We now analyze the behavior of the LMS adaptive filter using the assumptions and definitions that we have provided. For the first portion of our analysis, we characterize the mean behavior of the filter coefficients of the LMS algorithm in (19.1) and (19.2). Then, we provide a mean-square analysis of the system that characterizes the natures of $\mathbf{K}(n)$, $\xi_{ex}(n)$, and M in (19.14), (19.15), and (19.16), respectively.

19.4.1 Mean Analysis

By substituting the definition of $d(n)$ from the desired response signal model in (19.4) into the coefficient updates in (19.1) and (19.2), we can express the LMS algorithm in terms of the coefficient error vector in (19.13) as

$$\mathbf{V}(n+1) = \mathbf{V}(n) - \mu(n)\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n) + \mu(n)\eta(n)\mathbf{X}(n) \ . \tag{19.18}$$

We take expectations of both sides of (19.18), which yields

$$E\{\mathbf{V}(n+1)\} = E\{\mathbf{V}(n)\} - \mu(n)E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\} + \mu(n)E\{\eta(n)\mathbf{X}(n)\} \ , \tag{19.19}$$

in which we have assumed that $\mu(n)$ does not depend on $\mathbf{X}(n)$, $d(n)$, or $\mathbf{W}(n)$.

In many practical cases of interest, either the input signal $x(n)$ and/or the noise signal $\eta(n)$ is zero-mean, such that the last term in (19.19) is zero. Moreover, under the independence assumptions, it can be shown that $\mathbf{V}(n)$ is approximately independent of $\mathbf{X}(n)$, and thus the second expectation on the right-hand side of (19.19) is approximately given by

$$\begin{aligned} E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\} &\approx E\{\mathbf{X}(n)\mathbf{X}^T(n)\}E\{\mathbf{V}(n)\} \\ &= \mathbf{R}_{\mathrm{xx}} E\{\mathbf{V}(n)\} . \end{aligned} \tag{19.20}$$

Combining these results with (19.19), we obtain

$$E\{\mathbf{V}(n+1)\} = \left(\mathbf{I} - \mu(n)\mathbf{R}_{\mathrm{xx}}\right) E\{\mathbf{V}(n)\} . \tag{19.21}$$

The simple expression in (19.21) describes the evolutionary behavior of the mean values of the errors in the LMS adaptive filter coefficients. Moreover, if the step size $\mu(n)$ is constant, then we can write (19.21) as

$$E\{\mathbf{V}(n)\} = \left(\mathbf{I} - \mu\mathbf{R}_{\mathrm{xx}}\right)^n E\{\mathbf{V}(0)\} , \tag{19.22}$$

To further simplify this matrix equation, note that $\mathbf{R}_{\mathrm{xx}}$ can be described by its *eigenvalue decomposition* as

$$\mathbf{R}_{\mathrm{xx}} = \mathbf{Q}\Lambda\mathbf{Q}^T , \tag{19.23}$$

where $\mathbf{Q}$ is a matrix of the eigenvectors of $\mathbf{R}_{\mathrm{xx}}$ and Λ is a diagonal matrix of the eigenvalues $\{\lambda_0, \lambda_1, \ldots, \lambda_{L-1}\}$ of $\mathbf{R}_{\mathrm{xx}}$, which are all real valued because of the symmetry of $\mathbf{R}_{\mathrm{xx}}$. Through some simple manipulations of (19.22), we can express the $(i+1)$th element of $E\{\mathbf{W}(n)\}$ as

$$E\{w_i(n)\} = w_{i,opt} + \sum_{j=0}^{L-1} q_{ij}(1-\mu\lambda_j)^n E\{\widetilde{v}_j(0)\} , \tag{19.24}$$

where q_{ij} is the $(i+1, j+1)$th element of the eigenvector matrix $\mathbf{Q}$ and $\widetilde{v}_j(n)$ is the $(j+1)$th element of the *rotated coefficient error vector* defined as

$$\widetilde{\mathbf{V}}(n) = \mathbf{Q}^T\mathbf{V}(n) . \tag{19.25}$$

From (19.21) and (19.24), we can state several results concerning the mean behaviors of the LMS adaptive filter coefficients:

- *The mean behavior of the LMS adaptive filter as predicted by (19.21) is identical to that of the method of steepest descent for this adaptive filtering task.* Discussed in Chapter 18 of this Handbook, the method of steepest descent is an iterative optimization procedure that requires precise knowledge of the statistics of $x(n)$ and $d(n)$ to operate. That the LMS adaptive filter's average behavior is similar to that of steepest descent was recognized in one of the earliest publications of the LMS adaptive filter [1].
- *The mean value of any LMS adaptive filter coefficient at any time instant consists of the sum of the optimal coefficient value and a weighted sum of exponentially converging and/or diverging terms.* These error terms depend on the elements of the eigenvector matrix $\mathbf{Q}$, the eigenvalues of $\mathbf{R}_{\mathrm{xx}}$, and the mean $E\{\mathbf{V}(0)\}$ of the initial coefficient error vector.
- *If all of the eigenvalues $\{\lambda_j\}$ of $\mathbf{R}_{\mathrm{xx}}$ are strictly positive and*

$$0 < \mu < \frac{2}{\lambda_j} \tag{19.26}$$

for all $0 < j < L-1$, then the means of the filter coefficients converge exponentially to their optimum values. This result can be found directly from (19.24) by noting that the quantity $(1-\mu\lambda_j)^n \to 0$ as $n \to \infty$ if $|1-\mu\lambda_j| < 1$.

- *The speeds of convergence of the means of the coefficient values depend on the eigenvalues* λ_i *and the step size* μ. In particular, we can define the time constant τ_j of the jth term within the summation on the right hand side of (19.24) as the approximate number of iterations it takes for this term to reach $(1/e)$th its initial value. For step sizes in the range $0 < \mu \ll 1/\lambda_{max}$ where λ_{max} is the maximum eigenvalue of $\mathbf{R}_{xx}$, this time constant is

$$\tau_j \;=\; -\frac{1}{\ln(1-\mu\lambda_j)} \;\approx\; \frac{1}{\mu\lambda_j}\,. \tag{19.27}$$

Thus, faster convergence is obtained as the step size is increased. However, for step size values greater than $1/\lambda_{max}$, the speeds of convergence can actually decrease. Moreover, the convergence of the system is limited by its *mean-squared behavior,* as we shall indicate shortly.

An Example

Consider the behavior of an $L = 2$-coefficient LMS adaptive filter in which $x(n)$ and $d(n)$ are generated as

$$\begin{aligned} x(n) &= 0.5x(n-1) + \frac{\sqrt{3}}{2}z(n) && (19.28)\\ d(n) &= x(n) + 0.5x(n-1) + \eta(n)\,, && (19.29) \end{aligned}$$

where $z(n)$ and $\eta(n)$ are zero-mean uncorrelated jointly Gaussian signals with variances of one and 0.01, respectively. It is straightforward to show for these signal statistics that

$$\mathbf{W}_{opt} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \text{ and } \mathbf{R}_{xx} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}. \tag{19.30}$$

Figure 19.2(a) depicts the behavior of the mean analysis equation in (19.24) for these signal statistics, where $\mu(n) = 0.08$ and $\mathbf{W}(0) = [4 \;\; -0.5]^T$. Each circle on this plot corresponds to the value of $E\{\mathbf{W}(n)\}$ for a particular time instant. Shown on this $\{w_0,\ w_1\}$ plot are the coefficient error axes $\{v_0,\ v_1\}$, the rotated coefficient error axes $\{\widetilde{v}_0,\ \widetilde{v}_1\}$, and the contours of the excess MSE error surface ξ_{ex} as a function of w_0 and w_1 for values in the set $\{0.1,\ 0.2,\ 0.5,\ 1,\ 2,\ 5,\ 10,\ 20\}$. Starting from the initial coefficient vector $\mathbf{W}(0)$, $E\{\mathbf{W}(n)\}$ converge toward $\mathbf{W}_{opt}$ by reducing the components of the mean coefficient error vector $E\{\mathbf{V}(n)\}$ along the rotated coefficient error axes $\{\widetilde{v}_0,\ \widetilde{v}_1\}$ according to the exponential weighting factors $(1-\mu\lambda_0)^n$ and $(1-\mu\lambda_1)^n$ in (19.24).

For comparison, Fig. 19.2(b) shows five different simulation runs of an LMS adaptive filter operating on Gaussian signals generated according to (19.28) and (19.29), where $\mu(n) = 0.08$ and $\mathbf{W}(0) = [4 - 0.5]^T$ in each case. Although any single simulation run of the adaptive filter shows a considerably more erratic convergence path than that predicted by (19.24), one observes that the average of these coefficient trajectories roughly follows the same path as that of the analysis.

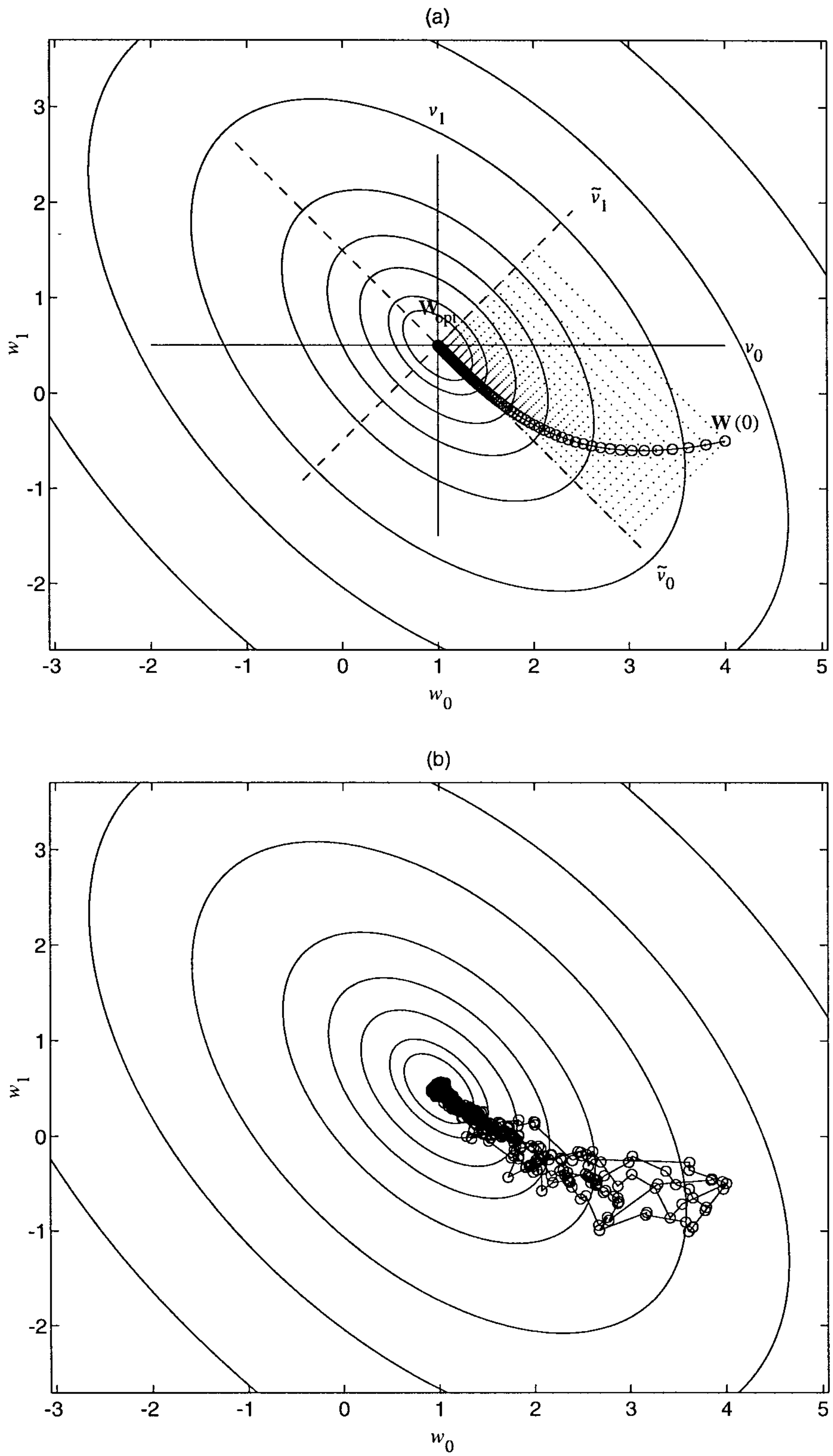

FIGURE 19.2: Comparison of the predicted and actual performances of the LMS adaptive filter in the two-coefficient example: (a) the behavior predicted by the mean analysis, and (b) the actual LMS adaptive filter behavior for five different simulation runs.

19.4.2 Mean-Square Analysis

Although (19.24) characterizes the mean behavior of the LMS adaptive filter, it does not indicate the nature of the fluctuations of the filter coefficients about their mean values, as indicated by the actual behavior of the LMS adaptive filter in Fig. 19.2(b). The magnitudes of these fluctuations can be accurately characterized through a *mean-square analysis* of the LMS adaptive filter. Because the coefficient error correlation matrix $\mathbf{K}(n)$ as defined in (19.14) is the basis for our mean-square analysis, we outline methods for determining an evolution equation for this matrix. Then, we derive the forms of this evolution equation for both the i.i.d. and SIRP input signal models described previously, and we summarize the resulting expressions for the steady-state values of the misadjustment and excess MSE in (19.16) and (19.17), respectively, for these different signal types. Finally, several conclusions regarding the mean-square behavior of the LMS adaptive filter are drawn.

Evolution of the Coefficient Error Correlation Matrix

To derive an evolution equation for $\mathbf{K}(n)$, we post-multiply both sides of (19.18) by their respective transposes, which gives

$$\begin{aligned}\mathbf{V}(n+1)\mathbf{V}^T(n+1) &= \left(\mathbf{I}-\mu(n)\mathbf{X}(n)\mathbf{X}^T(n)\right)\mathbf{V}(n)\mathbf{V}^T(n)\left(\mathbf{I}-\mu(n)\mathbf{X}(n)\mathbf{X}^T(n)\right)\\ &\quad+\mu^2(n)\eta^2(n)\mathbf{X}(n)\mathbf{X}^T(n)\\ &\quad+\mu(n)\eta(n)\left(\mathbf{I}-\mu(n)\mathbf{X}(n)\mathbf{X}^T(n)\right)\mathbf{V}(n)\mathbf{X}^T(n)\\ &\quad+\mu(n)\eta(n)\mathbf{X}(n)\mathbf{V}^T(n)\left(\mathbf{I}-\mu(n)\mathbf{X}(n)\mathbf{X}^T(n)\right)\,. \end{aligned} \tag{19.31}$$

Taking expectations of both sides of (19.31), we note that $\eta(n)$ is zero mean and independent of both $\mathbf{X}(n)$ and $\mathbf{V}(n)$ from our models and assumptions, and thus the expectations of the third and fourth terms on the right hand side of (19.31) are zero. Moreover, by using the independence assumptions, it can be shown that

$$\begin{aligned}E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{V}(n)\mathbf{V}^T(n)\} &\approx E\{\mathbf{X}(n)\mathbf{X}^T(n)\}E\{\mathbf{V}(n)\mathbf{V}^T(n)\}\\ &= \mathbf{R}_{\mathrm{xx}}\mathbf{K}(n)\,.\end{aligned} \tag{19.32}$$

Thus, we obtain from (19.31) the expression

$$\begin{aligned}\mathbf{K}(n+1) &= \mathbf{K}(n)-\mu(n)\left(\mathbf{R}_{\mathrm{xx}}\mathbf{K}(n)+\mathbf{K}(n)\mathbf{R}_{\mathrm{xx}}\right)\\ &\quad+\mu^2(n)E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}+\mu^2(n)\sigma_\eta^2\mathbf{R}_{\mathrm{xx}}\,,\end{aligned} \tag{19.33}$$

where σ_η^2 is as defined in (19.12).

At this point, the analysis can be simplified depending on how the third term on the right hand side of (19.33) is evaluated according to the signal models and assumptions.

Analysis for SIRP Input Signals: In this case, the value of $E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$ can be expressed as

$$E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\} = m_z^{(2,2)}\left[2\mathbf{R}_{\mathrm{xx}}\mathbf{K}(n)\mathbf{R}_{\mathrm{xx}}+\mathbf{R}_{\mathrm{xx}}\mathrm{tr}\left\{\mathbf{R}_{\mathrm{xx}}\mathbf{K}(n)\right\}\right]\,, \tag{19.34}$$

where the moment term $m_z^{(2,2)}$ is given by

$$m_z^{(2,2)} = E\{z_i^2(n)z_j^2(n)\} \tag{19.35}$$

for any $0 \le i \ne j \le (L-1)$ and

$$z_i(n) = \lambda_i^{1/2}\sum_{l=0}^{L-1} q_{li}\,x(n-l)\,. \tag{19.36}$$

If $x(n)$ is a Gaussian random process, then $m_z^{(2,2)} = 1$, and it can be shown that $m_z^{(2,2)} \geq 1$ for SIRPs in general. For more details on these results, see [15].

Analysis for I.I.D. Input Signals: In this case, we can express the (i, j)th element of the matrix $E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$ as

$$\left[E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}\right]_{i,j} = \begin{cases} 2\sigma_x^4[\mathbf{K}(n)]_{i,j}, & \text{if } i \neq j \\ \sigma_x^4\left(\gamma[\mathbf{K}(n)]_{i,i} + \sum_{m=1,\, m\neq i}^{L}[\mathbf{K}(n)]_{m,m}\right), & \text{if } i = j\,, \end{cases} \tag{19.37}$$

where $[\mathbf{K}(n)]_{i,j}$ is the (i, j)th element of $\mathbf{K}(n)$,

$$\sigma_x^2 = E\{x^2(n)\}, \text{ and } \gamma = \frac{E\{x^4(n)\}}{\sigma_x^4}\,, \tag{19.38}$$

respectively. For details, see [5].

Zeroth-Order Approximation Near Convergence: For small step sizes, it can be shown that the elements of $\mathbf{K}(n)$ are approximately proportional to both the step size and the noise variance σ_η^2 in steady-state. Thus, the magnitudes of the elements in the third term on the right hand side of (19.33) are about a factor of $\mu(n)$ smaller than those of any other terms in this equation at convergence. Such a result suggests that we could set

$$\mu^2(n)E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\} \approx \mathbf{0} \tag{19.39}$$

in the steady-state analysis of (19.33) without perturbing the analytical results too much. If this approximation is valid, then the form of (19.33) no longer depends on the form of the amplitude statistics of $x(n)$, as in the case of the mean analysis.

Excess MSE, Mean-Square Stability, and Misadjustment

Given the results in (19.34) through (19.39), we can use the evolution equation for $\mathbf{K}(n)$ in (19.33) to explore the mean-square behavior of the LMS adaptive filter in several ways:

- By studying the structure of (19.33) for different signal types, we can determine conditions on the step size $\mu(n)$ to guarantee the stability of the mean-square analysis equation.
- By setting $\mathbf{K}(n + 1) = \mathbf{K}(n)$ and fixing the value of $\mu(n)$, we can solve for the steady-state value of $\mathbf{K}(n)$ at convergence, thereby obtaining a measure of the fluctuations of the coefficients about their optimum solutions.
- Given a value for $\mathbf{V}(0)$, we can write a computer program to simulate the behavior of this equation for different signal statistics and step size sequences.

Moreover, once the matrix sequence $\mathbf{K}(n)$ is known, we can obtain the values of the excess MSE and misadjustment from $\mathbf{K}(n)$ by employing the relations in (19.16) and (19.17), respectively.

Table 19.1 summarizes many of the analytical results that can be obtained from a careful study of (19.33). Shown in the table are the conditions on the step size $\mu(n) = \mu$ to guarantee stability, sufficient stability conditions on the step size that can be easily calculated, and the misadjustment in steady-state for the three different methods of evaluating $E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$ in (19.34) through (19.39). In the table, the quantity C is defined as

$$C = \sum_{i=0}^{L-1} \frac{\lambda_i}{1 - \mu m_z^{(2,2)}\lambda_i}\,. \tag{19.40}$$

From these results and others that can be obtained from (19.33), we can infer several facts about the mean-square performance of the LMS adaptive filter:

TABLE 19.1 Summary of MSE Analysis Results

Assumption	MSE Stability Conditions	Sufficient Conditions	Misadjustment
I.I.D. input	$0 < \mu < \frac{2}{(L-1+\gamma)\sigma_x^2}$	$0 < \mu < \frac{2}{(L-1+\gamma)\sigma_x^2}$	$M = \frac{\mu\sigma_x^2 L}{2-\mu\sigma_x^2(L-1+\gamma)}$
SIRP input	$0 < \mu < \frac{1}{m_z^{(2,2)}\lambda_{max}}$ and $\mu m_z^{(2,2)} C < 2$	$0 < \mu < \frac{2}{3Lm_z^{(2,2)}\sigma_x^2}$	$M = \frac{\mu C}{2-\mu m_z^{(2,2)} C}$
Approx.	$0 < \mu < \frac{1}{\lambda_{max}}$	$0 < \mu < \frac{1}{L\sigma_x^2}$	$M = \frac{\mu\sigma_x^2 L}{2}$

- *The value of the excess MSE at time n consists of the sum of the steady-state excess MSE, given by $M\sigma_\eta^2$, and a weighted sum of L exponentially converging and/or diverging terms.* Similar to the mean analysis case, these additional terms depend on the elements of the eigenvector matrix $\mathbf{Q}$, the eigenvalues of $\mathbf{R_{xx}}$, the eigenvalues of $\mathbf{K}(0)$, and the values of $m_z^{(2,2)}$ or γ for the SIRP or i.i.d. input signal models, respectively.
- *For all input signal types, approximate conditions on the fixed step size value to guarantee convergence of the evolution equations for $\mathbf{K}(n)$ are of the form*

$$0 < \mu < \frac{K}{L\sigma_x^2}, \tag{19.41}$$

  where σ_x^2 is the input signal power and where the constant K depends weakly on the nature of the input signal statistics and not on the magnitude of the input signal. All of the sufficient stability bounds on μ as shown in Table 19.1 can be put in the form of (19.26). Because of the inaccuracies within the analysis that are caused by the independence assumptions, however, the actual step size chosen for stability of the LMS adaptive filter should be somewhat smaller than these values, and step sizes in the range $0 < \mu(n) < 0.1/(L\sigma_x^2)$ are often chosen in practice.
- *The misadjustment of the LMS adaptive filter increases as the filter length L and step size μ are increased.* Thus, a larger step size causes larger fluctuations of the filter coefficients about their optimum solutions in steady-state.

19.5 Performance Issues

When using the LMS adaptive filter, one must select the filter length L and the step size $\mu(n)$ to obtain the desired performance from the system. In this section, we explore the issues affecting the choices of these parameters using the analytical results for LMS adaptive filter's behavior derived in the last section.

19.5.1 Basic Criteria for Performance

The performance of the LMS adaptive filter can be characterized in three important ways: the *adequacy of the FIR filter model*, the *speed of convergence* of the system, and the *misadjustment* in steady-state.

Adequacy of the FIR Model

The LMS adaptive filter relies on the linearity of the FIR filter model to accurately characterize the relationship between the input and desired response signals. When the relationship between $x(n)$ and $d(n)$ deviates from the linear one given in (19.4), then the performance of the overall system suffers. In general, it is possible to use a nonlinear model in place of the adaptive FIR filter model considered here. Possible nonlinear models include polynomial-based filters such as Volterra and bilinear filters [16] as well as neural network structures [17].

Another source of model inaccuracy is the finite impulse response length of the adaptive FIR filter. It is typically necessary to tune both the length of the filter L and the relative delay between the input and desired response signals so that the input signal values *not* contained in $\mathbf{X}(n)$ are largely uncorrelated with the desired response signal sample $d(n)$. However, such a situation may be impossible to achieve when the relationship between $x(n)$ and $d(n)$ is of an infinite-impulse response (IIR) nature. Adaptive IIR filters can be considered for these situations, although the stability and performance behaviors of these systems are much more difficult to characterize. Adaptive IIR filters are discussed in Chapter 23 of this Handbook.

Speed of Convergence

The rate at which the coefficients approach their optimum values is called the *speed of convergence.* As the analytical results show, there exists no one quantity that characterizes the speed of convergence, as it depends on the initial coefficient values, the amplitude and correlation statistics of the signals, the filter length L, and the step size $\mu(n)$. However, we can make several qualitative statements relating the speed of convergence to both the step size and the filter length. All of these results assume that the desired response signal model in (19.4) is reasonable and that the errors in the filter coefficients are uniformly distributed across the coefficients on average.

- *The speed of convergence increases as the value of the step size is increased, up to step sizes near one-half the maximum value required for stable operation of the system.* This result can be obtained from a careful analysis of (19.33) for different input signal types and correlation statistics. Moreover, by simulating the behavior of (19.33) and (19.34) for typical signal scenarios, it is observed that the speed of convergence of the excess MSE actually *decreases* for large enough step size values. For i.i.d. input signals, the fixed step size providing fastest convergence of the excess MSE is exactly one-half the MSE step size bound as given in Table 19.1 for this type of input signal.
- *The speed of convergence decreases as the length of the filter is increased.* The reasons for this behavior are twofold. First, if the input signal is correlated, the *condition number* of $\mathbf{R}_{xx}$, defined as the ratio of the largest and smallest eigenvalues of this matrix, generally increases as L is increased for typical real-world input signals. A larger condition number for $\mathbf{R}_{xx}$ makes it more difficult to choose a good step size to obtain fast convergence of all of the elements of either $E\{\mathbf{V}(n)\}$ or $\mathbf{K}(n)$. Such an effect can be seen in (19.24), as a larger condition number leads to a larger disparity in the values of $(1 - \mu\lambda_j)$ for different j. Second, in the MSE analysis equation of (19.33), the overall magnitude of the expectation term $E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$ is larger for larger L, due to the fact that the scalar quantity $\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)$ within this expectation increases as the size of the filter is increased. Since this quantity is always positive, it limits the amount that the excess MSE can be decreased at each iteration, and it reduces the maximum step size that is allowed for mean-square convergence as L is increased.
- *The maximum possible speed of convergence is limited by the largest step size that can be chosen for stability for moderately correlated input signals.* In practice, the actual step size needed for stability of the LMS adaptive filter is smaller than one-half the maximum values given in Table 19.1 when the input signal is moderately correlated. This effect is due to the actual statistical relationships between the current coefficient vector $\mathbf{W}(n)$ and the signals $\mathbf{X}(n)$ and $d(n)$, relationships that are neglected via the independence assumptions. Since the convergence speed increases as μ is increased over this allowable step size range, the maximum stable step size provides a practical limit on the speed of convergence of the system.

- *The speed of convergence depends on the desired level of accuracy that is to be obtained by the adaptive filter.* Generally speaking, the speed of convergence of the system decreases as the desired level of misadjustment is decreased. This result is due to the fact that the behavior of the system is dominated by the slower-converging modes of the system as the length of adaptation time is increased. Thus, if the desired level of misadjustment is low, the speed of convergence is dominated by the slower-converging modes, thus limiting the overall convergence speed of the system.

Misadjustment

The misadjustment, defined in (19.16), is the additional fraction of MSE in the filter output above the minimum MSE value σ_η^2 caused by a nonzero adaptation speed. We can draw the following two conclusions regarding this quantity:

- *The misadjustment increases as the step size is increased.*
- *The misadjustment increases as the filter length is increased.*

Both results can be proven by direct study of the analytical results for M in Table 19.1.

19.5.2 Identifying Stationary Systems

We now evaluate the basic criteria for performance to provide qualitative guidance as to how to choose μ and L to identify a stationary system.

Choice of Filter Length

We have seen that as the filter length L is increased, the speed of convergence of the LMS adaptive filter decreases, and the misadjustment in steady-state increases. Therefore, the filter length should be chosen as short as possible but long enough to adequately model the unknown system, as too short a filter model leads to poor modeling performance. In general, there exists an optimal length L for a given μ that exactly balances the penalty for a finite-length filter model with the increase in misadjustment caused by a longer filter length, although the calculation of such a model order requires more information than is typically available in practice. Modeling criteria, such as Akaike's Information Criterion [18] and minimum description length (MDL) [19] could be used in this situation.

Choice of Step Size

We have seen that the speed of convergence increases as the step size is increased, up to values that are roughly within a factor of 1/2 of the step size stability limits. Thus, if fast convergence is desired, one should choose a large step size according to the limits in Table 19.1. However, we also observe that the misadjustment increases as the step size is increased. Therefore, if highly accurate estimates of the filter coefficients are desired, a small step size should be chosen. This classical tradeoff in convergence speed vs. the level of error in steady state dominates the issue of step size selection in many estimation schemes.

If the user knows that the relationship between $x(n)$ and $d(n)$ is linear and time-invariant, then one possible solution to the above tradeoff is to choose a large step size initially to obtain fast convergence, and then switch to a smaller step size to obtain a more accurate estimate of $\mathbf{W}_{opt}$ near convergence. The point to switch to a smaller step size is roughly when the excess MSE becomes a small fraction (approximately 1/10th) of the minimum MSE of the filter. This method of *gearshifting*, as it is commonly known, is part of a larger class of time-varying step size methods that we shall explore shortly.

Although we have discussed qualitative criteria by which to choose a fixed step size, it is possible to define specific performance criteria by which to choose μ. For one study of this problem for i.i.d. input signals, see [20].

19.5.3 Tracking Time-Varying Systems

Since the LMS adaptive filter continually adjusts its coefficients to approximately minimize the mean-squared error criterion, it can adjust to changes in the relationship between $x(n)$ and $d(n)$. This behavior is commonly referred to as *tracking.* In such situations, it clearly is not desirable to reduce the step size to an extremely small value in steady-state, as the LMS adaptive filter would not be able to follow any changes in the relationship between $x(n)$ and $d(n)$.

To illustrate the issues involved, consider the desired response signal model given by

$$d(n) = \mathbf{W}_{opt}^T(n)\mathbf{X}(n) + \eta(n) \ , \tag{19.42}$$

where the optimum coefficient vector $\mathbf{W}_{opt}(n)$ varies with time according to

$$\mathbf{W}_{opt}(n+1) = \mathbf{W}_{opt}(n) + \mathbf{M}(n) \ , \tag{19.43}$$

and $\{\mathbf{M}(n)\}$ is a sequence of vectors whose elements are all i.i.d. This nonstationary model is similar to others used in other tracking analyses of the LMS adaptive filter [4, 21]; it also enables a simple analysis that is similar to the stationary system identification model discussed earlier.

Applying the independence assumptions, we can analyze the behavior of the LMS adaptive filter for this desired response signal model. For brevity, we only summarize the results of an approximate analysis in which terms of the form $\mu^2 E\{\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{K}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$ are neglected in the MSE behavioral equations [4]. The misadjustment of the system in steady-state is

$$M_{non} = \frac{L}{2}\left(\mu\sigma_x^2 + \frac{\sigma_m^2}{\mu\sigma_\eta^2}\right) \ , \tag{19.44}$$

where σ_m^2 is the power in any one element of $\mathbf{M}(n)$. Details of a more-accurate tracking analysis can be found in [21].

In this case, the misadjustment is the sum of two terms. The first term is the same as that for the stationary case and is proportional to μ. The second term is the *lag error* and is due to the fact that the LMS coefficients follow or "lag" behind the optimum coefficient values. The lag error is proportional to the speed of variation of the unknown system through σ_m^2 and is inversely proportional to the step size, such that its value increases as the step size is decreased.

In general, there exists an optimum fixed step size that minimizes the misadjustment in steady-state for an LMS adaptive filter that is tracking changes in $\mathbf{W}_{opt}(n)$. For the approximate analysis used to derive (19.44), the resulting step size is

$$\mu_{opt} = \frac{\sigma_m}{\sigma_\eta \sigma_x} \ . \tag{19.45}$$

As the value of σ_m^2 increases, the level of nonstationarity increases such that a larger step size is required to accurately track changes in the unknown system. Similar conclusions can be drawn from other analyses of the LMS adaptive filter in tracking situations [22].

19.6 Selecting Time-Varying Step Sizes

The analyses of the previous sections enable one to choose a fixed step size μ for the LMS adaptive filter to meet the system's performance requirements when the general characteristics of the input and desired response signals are known. In practice, the exact statistics of $x(n)$ and $d(n)$ are unknown or vary with time. A time-varying step size $\mu(n)$, if properly computed, can provide stable, robust, and accurate convergence behavior for the LMS adaptive filter in these situations. In this section, we consider useful on-line procedures for computing $\mu(n)$ in the LMS adaptive filter to meet these performance requirements.

19.6.1 Normalized Step Sizes

For the LMS adaptive filter to be useful, it must operate in a stable manner so that its coefficient values do not diverge. From the stability results in Table 19.1 and the generalized expression for these stability bounds in (19.41), the upper bound for the step size is inversely proportional to the input signal power σ_x^2 in general. In practice, the input signal power is unknown or varies with time. Moreover, if one were to choose a small fixed step size value to satisfy these stability bounds for the largest anticipated input signal power value, then the convergence speed of the system would be unnecessarily slow during periods when the input signal power is small.

These concerns can be addressed by calculating a *normalized step size* $\mu(n)$ as

$$\mu(n) = \frac{\overline{\mu}}{\delta + L\widehat{\sigma_x^2}(n)} , \tag{19.46}$$

where $\widehat{\sigma_x^2}(n)$ is an *estimate* of the input signal power, $\overline{\mu}$ is a constant somewhat smaller than the value of K required for system stability in (19.41), and δ is a small constant to avoid a divide-by-zero should $\widehat{\sigma_x^2}(n)$ approach zero. To estimate the input signal power, a lowpass filter can be applied to the sequence $x^2(n)$ to track its changing envelope. Typical estimators include

- *exponentially weighted estimate:*

$$\widehat{\sigma_x^2}(n) = (1-c)\widehat{\sigma_x^2}(n-1) + cx^2(n) , \tag{19.47}$$

- *sliding-window estimate:*

$$\widehat{\sigma_x^2}(n) = \frac{1}{N}\sum_{i=0}^{N-1} x^2(n-i) , \tag{19.48}$$

where the parameters c, $0 < c \ll 1$ and N, $N \geq L$ control the effective memories of the two estimators, respectively.

The Normalized LMS Adaptive Filter

By choosing a sliding window estimate of length $N = L$, the LMS adaptive filter with $\mu(n)$ in (19.46) becomes

$$\begin{aligned} \mathbf{W}(n+1) &= \mathbf{W}(n) + \frac{\overline{\mu}e(n)}{p(n)}\mathbf{X}(n) & (19.49) \\ p(n) &= \delta + ||\mathbf{X}(n)||^2 , & (19.50) \end{aligned}$$

where $||\mathbf{X}(n)||^2$ is the L_2-norm of the input signal vector. The value of $p(n)$ can be updated recursively as

$$p(n) = p(n-1) + x^2(n) - x^2(n-L) , \tag{19.51}$$

where $p(0) = \delta$ and $x(n) = 0$ for $n \leq 0$. The adaptive filter in (19.49) is known as the *normalized LMS (NLMS) adaptive filter.* It has two special properties that make it useful for adaptive filtering tasks:

- *The NLMS adaptive filter is guaranteed to converge for any value of* $\overline{\mu}$ *in the range*

$$0 < \overline{\mu} < 2, \tag{19.52}$$

  regardless of the statistics of the input signal. Thus, selecting the value of $\overline{\mu}$ for stable behavior of this system is much easier than selecting μ for the LMS adaptive filter.
- *With the proper choice of* $\overline{\mu}$*, the NLMS adaptive filter can often converge faster than the LMS adaptive filter.* In fact, for noiseless system identification tasks in which $\eta(n)$ in (19.4) is zero, one can obtain $\mathbf{W}_{opt}(n) = \mathbf{W}_{opt}$ after L iterations of (19.49) for $\overline{\mu} = 1$. Moreover, for SIRP input signals, the NLMS adaptive filter provides more uniform convergence of the filter coefficients, making the selection of $\overline{\mu}$ an easier proposition than the selection of μ for the LMS adaptive filter.

A discussion of these and other results on the NLMS adaptive filter can be found in [15, 23]– [25].

19.6.2 Adaptive and Matrix Step Sizes

In addition to stability, the step size controls both the speed of convergence and the misadjustment of the LMS adaptive filter through the statistics of the input and desired response signals. In situations where the statistics of $x(n)$ and/or $d(n)$ are changing, the value of $\mu(n)$ that provides the best performance from the system can change as well. In these situations, it is natural to consider $\mu(n)$ as an adaptive parameter to be optimized along with the coefficient vector $\mathbf{W}(n)$ within the system. While it may seem novel, the idea of computing an *adaptive step size* has a long history in the field of adaptive filters [2]. Numerous such techniques have been proposed in the scientific literature. One such method uses a stochastic gradient procedure to adjust the value of $\mu(n)$ to iteratively minimize the MSE within the LMS adaptive filter. A derivation and performance analysis of this algorithm is given in [26].

In some applications, the task at hand suggests a particular strategy for adjusting the step size $\mu(n)$ to obtain the best performance from an LMS adaptive filter. For example, in echo cancellation for telephone networks, the signal-to-noise ratio of $d(n)$ falls to extremely low values when the near-end talker signal is present, making accurate adaptation during these periods difficult. Such systems typically employ *double-talk detectors*, in which estimates of the statistical characteristics of $x(n)$, $d(n)$, and/or $e(n)$ are used to raise and lower the value of $\mu(n)$ in an appropriate manner. A discussion of this problem and a method for its solution are given in [27].

While our discussion of the LMS adaptive filter has assumed a single step size value for each of the filter coefficients, it is possible to select L *different* step sizes $\mu_i(n)$ for each of the L coefficient updates within the LMS adaptive filter. To select fixed values for each $\mu_i(n) = \mu_i$, these *matrix step size* methods require prior knowledge about the statistics of $x(n)$ and $d(n)$ and/or the approximate values of $\mathbf{W}_{opt}$. It is possible, however, to adapt each $\mu_i(n)$ according to a suitable performance criterion to obtain improved convergence behavior from the overall system. A particularly simple adaptive method for calculating matrix step sizes is provided in [28].

19.6.3 Other Time-Varying Step Size Methods

In situations where the statistics of $x(n)$ and $d(n)$ do not change with time, choosing a variable step size sequence $\mu(n)$ is still desirable, as one can decrease the misadjustment over time to obtain an accurate estimate of the optimum coefficient vector $\mathbf{W}_{opt}$. Such methods have been derived and characterized in a branch of statistical analysis known as *stochastic approximation* [29]. Using this formalism, it is possible to prove under certain assumptions on $x(n)$ and $d(n)$ that the value of $\mathbf{W}(n)$ for the LMS adaptive filter converges to $\mathbf{W}_{opt}$ as $n \to \infty$ if $\mu(n)$ satisfies

$$\sum_{n=0}^{\infty} |\mu(n)| \to \infty \text{ and } \sum_{n=0}^{\infty} \mu^2(n) < \infty \, , \tag{19.53}$$

respectively. One step size function satisfying these constraints is

$$\mu(n) = \frac{\mu(0)}{n+1} \, , \tag{19.54}$$

where $\mu(0)$ is an initial step size parameter. The gearshifting method described in Section 19.5.2 can be seen as a simple heuristic approximation to (19.54). Moreover, one can derive an optimum step size sequence $\mu_{opt}(n)$ that minimizes the excess MSE at each iteration under certain situations, and the limiting form of the resulting step size values for stationary signals are directly related to (19.54) as well [24].

19.7 Other Analyses of the LMS Adaptive Filter

While the analytical techniques employed in this section are useful for selecting design parameters for the LMS adaptive filter, they are but one method for characterizing the behavior of this system. Other forms of analyses can be used to determine other characteristics of this system, such as the p.d.f.s of the adaptive filter coefficients [30] and the probability of large excursions in the adaptive filter coefficients for different types of input signals [31]. In addition, much research effort has focused on characterizing the stability of the system without extensive assumptions about the signals being processed. One example of such an analysis is given in [32]. Other methods for analyzing the LMS adaptive filter include the method of ordinary differential equations (ODE) [33], the stochastic approximation methods described previously [29], computer-assisted symbolic derivation methods [14], and averaging techniques that are particularly useful for deterministic signals [34].

19.8 Analysis of Other Adaptive Filters

Because of the difficulties in performing multiplications in the first digital hardware implementations of adaptive filters, many of these systems employed nonlinearities in the coefficient update terms to simplify their hardware requirements. An example of one such algorithm is the *sign-error* adaptive filter, in which the coefficient update is

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \mu(n)\mathrm{sgn}(e(n))\mathbf{X}(n) \, , \tag{19.55}$$

where the value of $\mathrm{sgn}(e(n))$ is either 1 or -1 depending on whether $e(n)$ is positive or negative, respectively. If $\mu(n)$ is chosen as a power-of-two, this algorithm only requires a comparison and bit shift per coefficient to implement in hardware. Other algorithms employing nonlinearities of the input signal vector $\mathbf{X}(n)$ in the updates are also useful [35].

Many of the analysis techniques developed for the LMS adaptive filter can be applied to algorithms with nonlinearities in the coefficient updates, although such methods require additional assumptions to obtain accurate results. For presentations of two such analyses, see [36, 37]. It should be noted that the performance characteristics and stability properties of these nonlinearly modified versions of the LMS adaptive filter can be quite different from those of the LMS adaptive filter. For example, the sign-error adaptive filter in (19.55) is guaranteed to converge for any fixed positive step size value under fairly loose assumptions on $x(n)$ and $d(n)$ [38].

19.9 Conclusions

In summary, we have described a statistical analysis of the LMS adaptive filter, and through this analysis, suggestions for selecting the design parameters for this system have been provided. While useful, analytical studies of the LMS adaptive filter are but one part of the system design process. As in all design problems, sound engineering judgment, careful analytical studies, computer simulations, and extensive real-world evaluations and testing should be combined when developing an adaptive filtering solution to any particular task.

References

[1] Widrow, B. and Hoff, M.E., Adaptive switching circuits, *IRE WESCON Conv. Rec.*, 4, 96–104, Aug. 1960.

[2] Widrow, B., Adaptive sampled-data systems — a statistical theory of adaptation, *IRE WESCON Conv. Rec.*, 4, 74–85, Aug. 1959.

[3] Senne, K.D., Adaptive linear discrete-time estimation, Ph.D. thesis, Stanford University, Stanford, CA, June 1968.

[4] Widrow, B., McCool, J., Larimore, M.G., and Johnson, C.R., Jr., Stationary and nonstationary learning characteristics of the LMS adaptive filter, *Proc. IEEE*, 64(8), 1151–1162, Aug. 1976.

[5] Gardner, W.A., Learning characteristics of stochastic-gradient-descent algorithms: a general study, analysis, and critique, *Signal Processing*, 6(2), 113–133, April 1984.

[6] Feuer, A. and Weinstein, E., Convergence analysis of LMS filters with uncorrelated data, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-331, 222–230, Feb. 1985.

[7] Messerschmitt, D.G., Echo cancellation in speech and data transmission, *IEEE J. Selected Areas Comm.*, 2(2), 283–301, Mar. 1984.

[8] Widrow, B. and Walach, E., *Adaptive Inverse Control*, Prentice-Hall, Upper Saddle River, NJ, 1996.

[9] Proakis, J.G., *Digital Communications*, 3rd ed., McGraw-Hill, New York, 1995.

[10] Murano, K., Unagami, S., and Amano, F., Echo cancellation and applications, *IEEE Commun. Mag.*, 28(1), 49–55, Jan. 1990.

[11] Gradsteyn, I.S. and Ryzhik, I.M., *Table of Integrals, Series and Products*, Academic Press, New York, 1980.

[12] Brehm, H. and Stammler, W., Description and generation of spherically invariant speech-model signals, *Signal Processing*, 12(2), 119–141, Mar. 1987.

[13] Mazo, J.E., On the independence theory of equalizer convergence, *Bell Sys. Tech. J.*, 58(5), 963–993, May-June 1979.

[14] Douglas, S.C. and Pan, W., Exact expectation analysis of the LMS adaptive filter, *IEEE Trans. Signal Processing*, 43(12), 2863–2871, Dec. 1995.

[15] Rupp, M., The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes, *IEEE Trans. Signal Processing*, 41(3), 1149–1160, Mar. 1993.

[16] Mathews, V.J., Adaptive polynomial filters, *IEEE Signal Proc. Mag.*, 8(3), 10–26, July 1991.

[17] Haykin, S., *Neural Networks*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[18] Akaike, H., A new look at the statistical model identification, *IEEE Trans. Automatic Control*, AC-19(6), 716–723, Dec. 1974.

[19] Rissanen, J., Modelling by shortest data description, *Automatica*, 14(5), 465–471, Sept. 1978.

[20] Bershad, N.J., On the optimum gain parameter in LMS adaptation, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-35(7), 1065–1068, July 1987.

[21] Gardner, W.A., Nonstationary learning characteristics of the LMS algorithm, *IEEE Trans. Circ. Syst.*, 34(10), 1199–1207, Oct. 1987.

[22] Farden, D.C., Tracking properties of adaptive signal processing algorithms, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-29(3), 439–446, June 1981.

[23] Bitmead, R.R. and Anderson, B.D.O., Performance of adaptive estimation algorithms in dependent random environments, *IEEE Trans. Automatic Control*, AC-25(4), 788–794, Aug. 1980.

[24] Slock, D.T.M., On the convergence behavior of the LMS and the normalized LMS algorithms, *IEEE Trans. Signal Processing*, 41(9), 2811–2825, Sept. 1993.

[25] Douglas, S.C. and Meng, T.H.-Y., Normalized data nonlinearities for LMS adaptation, *IEEE Trans. Signal Processing*, 42(6), 1352–1365, June 1994.

[26] Mathews, V.J. and Xie, Z., A stochastic gradient adaptive filter with gradient adaptive step size, *IEEE Trans. Signal Processing*, 41(6), 2075–2087, June 1993.

[27] Ding, Z., Johnson, C.R., Jr., and Sethares, W.A., Frequency-dependent bursting in adaptive echo cancellation and its prevention using double-talk detectors, *Int. J. Adaptive Contr. Signal Processing*, 4(3), 219–216, May-June 1990.

[28] Harris, R.W., Chabries, D.M., and Bishop, F.A., A variable step (VS) adaptive filter algorithm, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34(2), 309–316, April 1986.

[29] Kushner, H.J. and Clark, D.S., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, New York, 1978.

[30] Bershad, N.J. and Qu, L.Z., On the probability density function of the LMS adaptive filter weights, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-37(1), 43–56, Jan. 1989.

[31] Rupp, M., Bursting in the LMS algorithm, *IEEE Trans. on Signal Processing*, 43(10), 2414–2417, Oct. 1995.

[32] Macchi, O. and Eweda, E., Second-order convergence analysis of stochastic adaptive linear filtering, *IEEE Trans. Automatic Control*, AC-28(1), 76–85, Jan. 1983.

[33] Benveniste, A., Métivier, M. and Priouret, P., *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, New York, 1990.

[34] Solo, V. and Kong, X., *Adaptive Signal Processing Algorithms: Stability and Performance*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[35] Duttweiler, D.L., Adaptive filter performance with nonlinearities in the correlation multiplier, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-30(4), 578–586, Aug. 1982.

[36] Bucklew, J.A., Kurtz, T.J., and Sethares, W.A., Weak convergence and local stability properties of fixed step size recursive algorithms, *IEEE Trans. Inform. Theory*, 39(3), 966–978, May 1993.

[37] Douglas, S.C. and Meng, T.H.-Y., Stochastic gradient adaptation under general error criteria, *IEEE Trans. Signal Processing*, 42(6), 1335–1351, June 1994.

[38] Cho, S.H. and Mathews, V.J., Tracking analysis of the sign algorithm in nonstationary environments, *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-38(12), 2046–2057, Dec. 1990.