TAP LEAKAGE APPLIED TO ECHO CANCELLATION

by

FRANCIS CHEUNG

Department of Electrical Engineering

McGill University, Montreal.

A project submitted in partial fulfillment

of the requirement for the degree of

Master of Electrical Engineering

ABSTRACT

This study presents a tap-leakage adjustment algorithm to control the tap drifting problem in an adaptive echo canceller. A nonrecursive transversal filter structure and stochastic gradient adaptation algorithm are first studied. On the basis of these studies, the effect of tap drift when the input spectrum does not cover the full band is presented. The tap-leakage algorithm, which has been used in fractional spaced equalizers and speech coding is introduced. In this thesis, the tap-leakage algorithm is applied to an echo canceller. In addition, a least-squares lattice filter is proposed to overcome slow convergence problems due to narrowband inputs. Finally, the simulation results of the stochastic gradient, tap-leakage and least-squares lattice algorithms are studied.

ACKNOWLEDGEMENTS

I wish to extend my sincere thanks to my supervisor, Prof. Peter Kabal who has given me continued guidance and encouragement throughout my graduate study. I also want to thank him for directing this work and for his many valuable suggestions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Most telephone connections generate echoes. In Fig.1.1a, a typical telephone connection is shown; although simplified, it is a good model to illustrate the echo problem. The connection contains two-wire portions on both ends (the local loops to the subscriber premises) which serve for bidirectional communication. The use of the same two wires for transmission and reception results in a saving in wire and local switching equipment. In the middle of the connection is the four-wire portion (carrier systems for medium to long-haul transmission). In this portion, a separate path is necessary for each direction of transmission. There are two reasons for using separate paths in long-distance communication. First, long circuits require repeaters for amplification, and amplifiers are unidirectional devices. Second, due to reasons of economy, most long-distance transmission are multiplexed. Multiplexing requires that signals in the two directions be sent in different slots.

A device that accomplishes the interfacing between the two-wire portion and the four-wire portion is called a hybrid coupler. Echoes are the result of impedance mismatches in the hybrid coupler. Since, a large number of different local loops can access to the hybrid coupler, it is not likely that a fixed balancing impedance can match all the impedances of the local loops. As a result, transmitted energy from A can leak through the coupler at B and transmits back to form an echo, as shown in Fig.1.1c. Another echo mechanism as shown in Fig.1.1d results in the receiver at B receiving a delayed echo of the message from A.

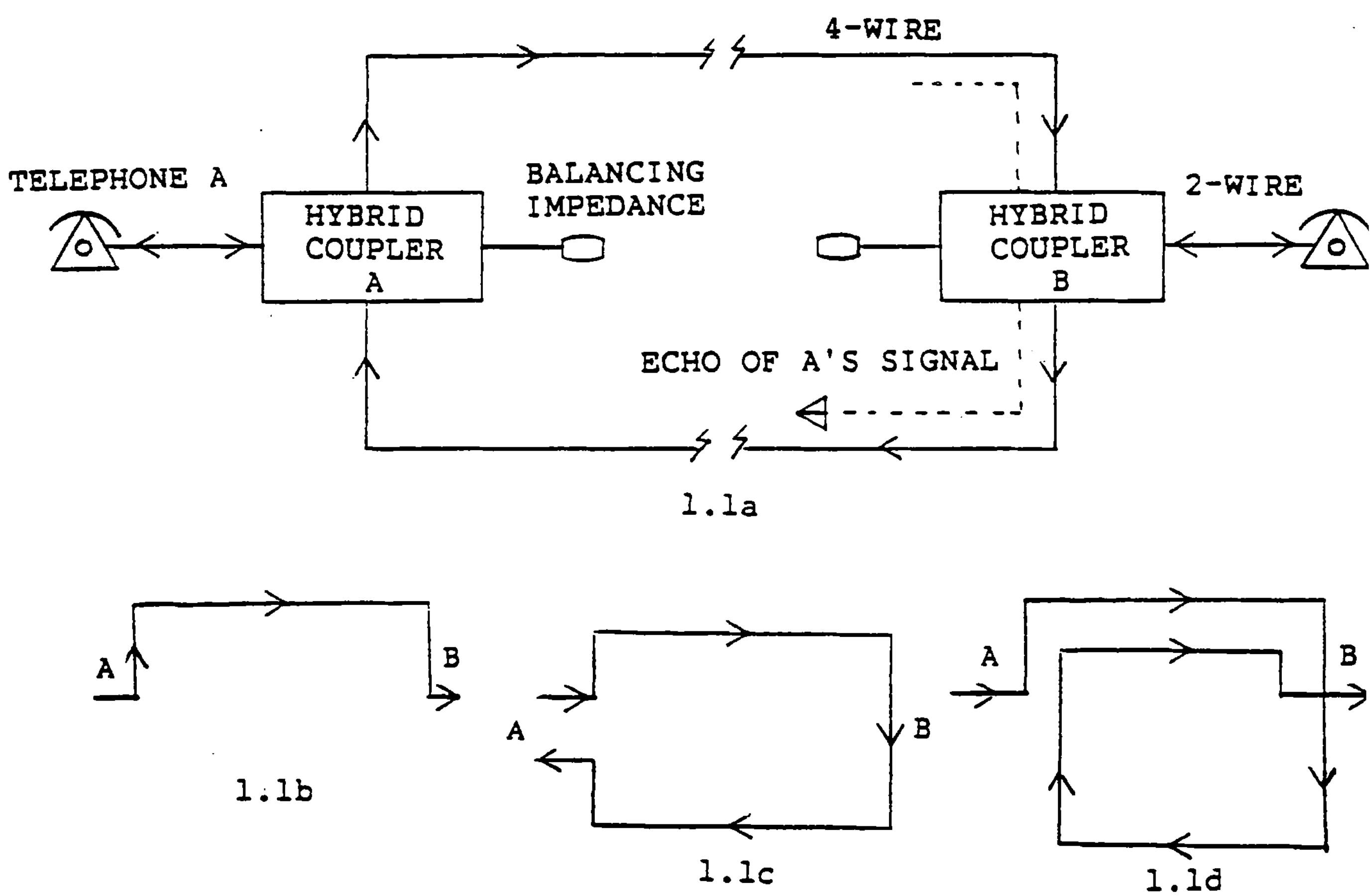


Fig. 1.1. Sources of echo in telephone network;
1.1a, model for long-distance circuits;
1.1b, talker speech path; 1.1c talker echo path;
1.1d listener echo path.

Echo suppressors have been accepted as the conventional methods to combat echoes since the early days of long-distance telephone [Sondhi, 1 and Mitchell, 2]. However, an echo suppressor introduces its own problem by chopping speech sounds and impairing conversational interruptions. Because of these distortion effects, researchers have proposed a new technique of echo cancelling recently and this has been proved to be subjectively superior to the conventional echo suppressor technique.

The idea of an echo canceller was first proposed by Kelly, Logan [3] and Sondhi [4]. An echo canceller is basically a combination of filter and subtractor as shown in Fig.1.2. The filter compensates the echoes by synthesizing a replica of the system response of the echo channel. If this synthesis is exact, then the inputs to the subtractor will be equal, and the residual echo from the subtractor will be completely null. Because the impulse response to the synthesized channel is initially unknown and may vary with time, the canceller must be an adaptive filter.

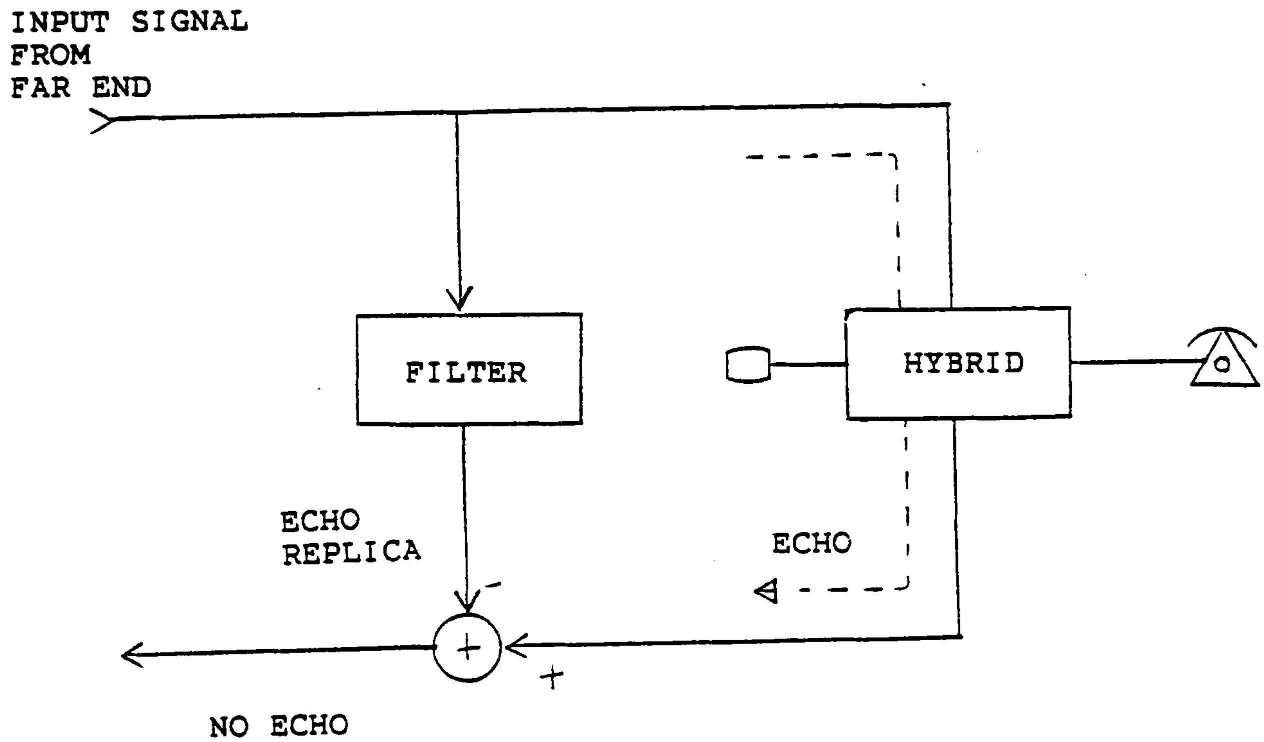


Fig. 1.2. General configuration of an echo canceller.

The adaptive filter used here in the illustration can be characterized in two ways. First, a filter has a finite number of internal parameters which can be used to control the transfer function of the filter. There are many ways to construct a filter with a transfer function which depends on a finite set of parameters. What is meant by structure of the filter is the particular configuration of realizing the filter. Second, an adaptation algorithm monitor the external changing environment and controls the filter transfer function by varying the aforementioned filter parameters. Two commonly used filter structures to be discussed in the future chapters are the transversal and lattice structures. Once a filter structure has been chosen, an adaptation algorithm must also be chosen. Two approaches to the adaptive filtering algorithms are the stochastic gradient (SG) and least-squares (LS) techniques. The adaptive filter structure which underlies almost all echo canceller designs is the nonrecursive transversal filter adapted according to the stochastic adjustment algorithms [Sondhi 1, Weinstein 5, Gritton 6, Messerschmitt 7]. The SG algorithm is largely due to the work of Bernard Widrow [8], who is one of the authors of the tutorial paper [Widrow et al 9, "Adaptive Noise Cancelling"].

1.1 Main Concerns

The main concerns in this project are the effects of adding input energy spectrum (e.g. broadband, such as speech, and narrowband, such as tones) to the echo cancellers that utilizes the transversal filter structure adapted according to the SG algorithm. When input energy covers only a narrowband, two problems will occur. First, the filter will not have a unique optimal solution set of filter parameters (or tap coefficients). Instead, there will be a continuous drifting effect away from the optimal solution. The parameters can reach very large values. When the input energy is changed to broadband suddenly after running the filter under the narrowband input for a long time, the drifting effect will cause the echo canceller a degradation in the performance (namely an increase in the cost function : the mean-square error). A tap-leakage adjustment algorithm can be used here in order to prevent the tap coefficients from reaching large values. This adjustment algorithm is being used in other two research areas : fractional spaced equalizers by Gitlin [10] and ADPCM speech coding by Atal [11]. In addition, the rate of convergence of the adaptation algorithm will be slowed down. An alternate echo canceller design using lattice structures adapted to the least-squares algorithm can be employed to speed up the rate [Honig, 12].

1.2 Organization

Chapter 2 will study the nonrecursive transversal filter structure of the echo canceller and the differences between the speech and data echo canceller designs. Chapter 3 will discuss the stochastic gradient (SG) algorithm in detail. In order to understand this algorithm, two other algorithms are studied as steping stones, namely : the minimum mean-square error (MMSE) and steepest descent algorithms. These are followed by discussions of the convergence properties of the SG algorithm. Chapter 4 will emphasize on the problems on the echo canceller when different input energy types are used. Chapter 5 will present a tap-leakage adjustment algorithm to deal with the problem that the tap coefficients can drift to large values. Chapter 6 will present the least-squares lattice canceller to cope with the second problem of the rate of convergence. The simulation results of the algorithms discussed in Chapter 3, 5, and 6 will be included in Chapter 7.

CHAPTER 2

ECHO CANCELLATION

2.1 Echo Cancellation in Speech Transmission

Echo cancellation is used to combat the echo for both speech and data transmission. The requirements for speech and data transmission are quite different, so this section will concentrate on speech and the next section will deal with data.

An echo canceller is basically a combination of an adaptive filter and subtractor. The adaptive filter gradually adapts its impulse response to the impulse response of the hybrid channel (or the echo channel). Thus, the filter synthesizes a replica of the echo and when subtraction has taken place between the echo and the echo replica, a very small residual echo will remain. In order to effectively cancel these echoes, it must be assumed that the echo channel is linear and time-invariant (or slowly varying), can be completely specified by its impulse response (or equivalently its transfer function). The typical filter structure in most of the echo canceller design is the transversal filter structure as shown in Fig.2.1. The filter contains n adjustable tap coefficients and (n-1) delay elements. The output from the echo channel, d(T) is a superposition of the near-end possible signal with the undesired echo signal. The reference input y(T) consists of the far-end signal alone.

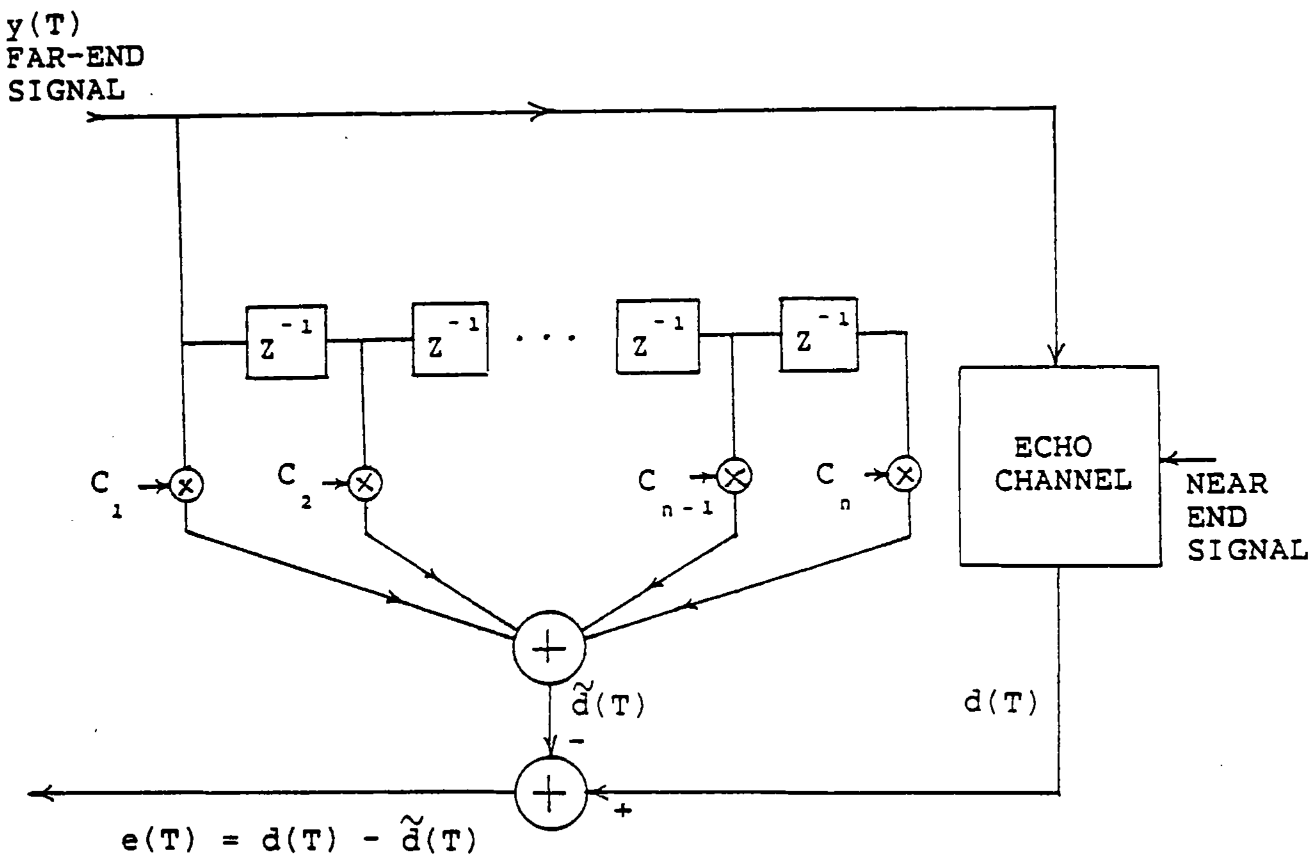


Fig. 2.1. Echo canceller for one direction of transmission.

The canceller replica $\tilde{d}(T)$ can be generated by a transversal filter with input y(T),

$$\tilde{d}(T) = \sum_{j=1}^{n} C_j \, y(T-j+1)$$

where C_j are the filter coefficients. Since the transfer function of the echo channel is not known in advance, it is necessary to adapt the filter coefficients C_j of the cancellation filter. The adaptation algorithms which try to minimize a cost function of the residual uncancelled echo, e(T), will be studied in detail in the next chapter.

If y(T) is bandlimited to the frequency range $|f| < f_{max}$, and τ is chosen to be the corresponding Nyquist interval ($\tau = 1/(2f_{max})$), a bandlimited filter will be obtained for which c_j's are Nyquist rate samples of the impulse response. Thus $n\tau$ must be chosen to be larger than the longest echo path that needs to be accommodated. Speech signals on the telephone network have little energy above 4000 Hz., so τ is approximately chosen as 125 μs. Thus 120 tap coefficients are required for a duration of 15 ms in the echo path.

In practice, it is necessary to cancel the echoes in both direction of a trunk. Thus, two adaptive transversal filters are necessary, as shown in Fig.2.2. It is desirable to position the cancellers in a split configuration, with the four-wire long delayed portion in the middle. The reason is that the number of tap coefficients for each of the transversal filters depends on the length of the impulse response of the echo channel, which is relatively short, and also on the transmission delay between the canceller and the hybrid channel. In the split configuration, each canceller can be as close to the hybrid as possible to minimize the transmission delay and thus minimize the number of coefficients used.

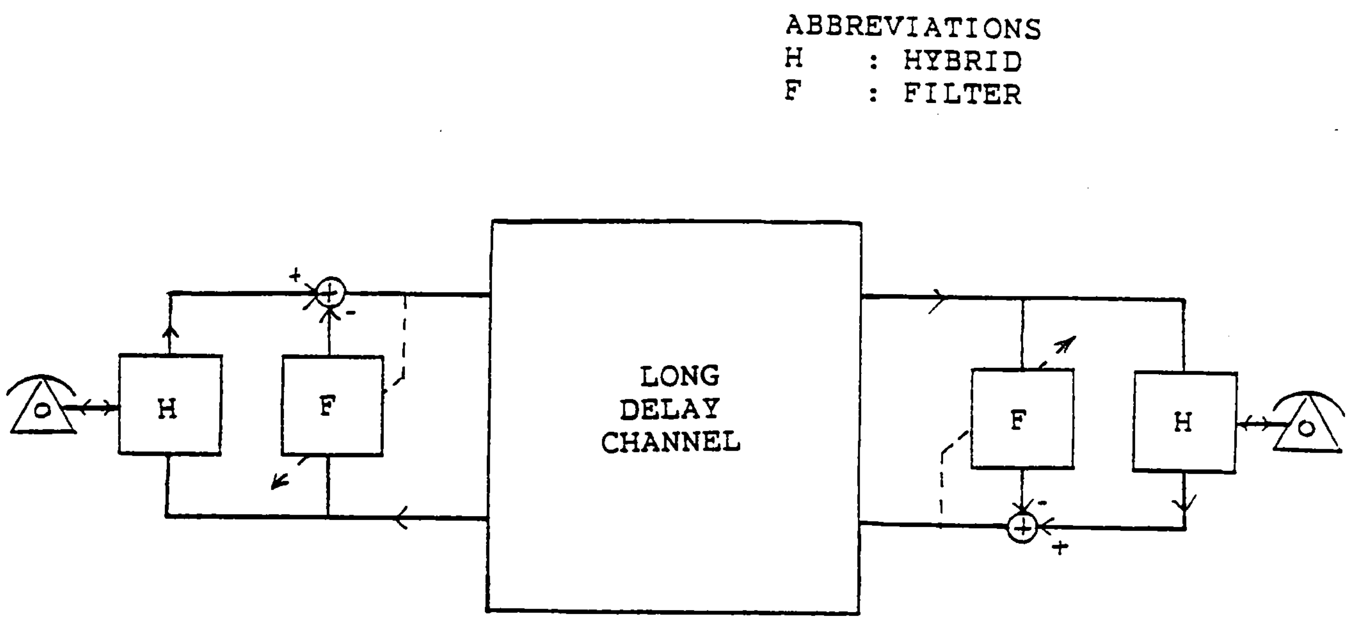
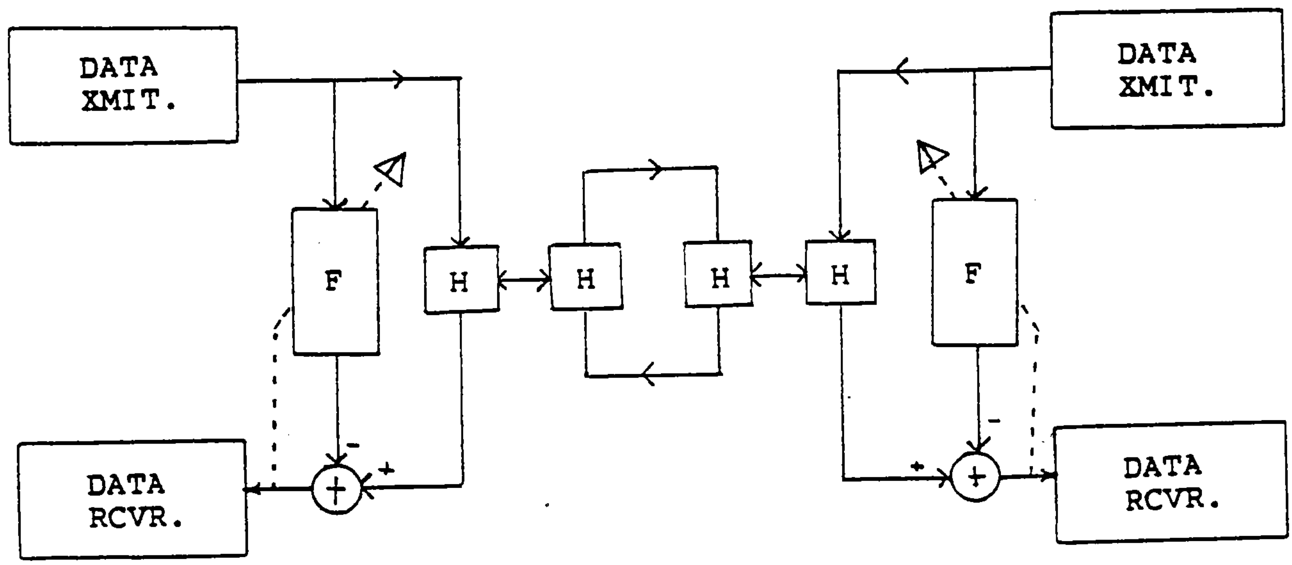


Fig. 2.2. Split echo canceller configuration for two directions.

2.2 Echo Cancellation in Data Transmission

Echo cancellation technique is also used in full duplex data transmission as shown in Fig.2.3, For half-duplex transmission, echoes create no problem. This is because transmission is in one direction only and there is no receiver on the current transmitting end to be affected by the echo. In full-duplex transmission, the data signals are transmitted in two directions simultaneously. Echoes from the data signals in one direction can interfere with the data signals in the opposite direction. An alternate method other than echo cancellation that can be used to separate the two directions of transmission is by using two non-overlapping frequency bands. However, this method has the disadvantage that the bandwidth is doubled relative to the bandwidth required in the echo cancellation method.

Although basically the data echo canceller is similar to the speech echo canceller (transversal filter with tap coefficients updated according to some adaptation algorithm to minimize a cost function of the error) there are several significant differences. The first difference is in the placement of the canceller. Unlike the voice echo cancellers which are considered as part of the telephone network, a data echo canceller would be located at the data set. This is because a local hybrid is needed at each data set in order to



ABBREVIATIONS
F : FILTER
H : HYBRID
XMIT : TRANSMITTER
RCVR : RECEIVER

Fig. 2.3. Echo canceller for full-duplex data transmission.

separate the transmit and receive directions. The echoes due to the local mismatch in speech communication can be classified as sidetones and are harmless. However, for data transmission, the early echoes are worse than the long delayed echoes that arrives after the round-trip delay in the four-wire circuit. The necessity of placing the echo canceller at the data set gives rise to the second difference between speech and data transmission. The time span over which echoes arrive is much larger for the latter because there exists a long-delayed echo. This problem is dealt with by splitting the transversal filter into two filters separated by a bulk delay. The third difference is that the data symbols have much simpler statistical properties to quantify. Sequences of data symbol (eg. ± 1 for binary sequences) may be assumed to be sequences of independent and identically distributed variables. The fourth difference is that data signals in both directions are present almost all the time, thus the near signal becomes an important issue during adaptation. Finally, there are significant problems of time recovery, synchronization and equalization exist in the data cancellation that do not exist in speech cancellation.

Full-duplex data transmission has arisen in two important applications. The first application is in the switched public telephone network and the second application is in the digital subscriber loop. The configuration for the former case is shown in Fig.2.3 . While the configuration for the latter is similiar to Fig.2.3 except the two-wire subscriber loop is present in between the two local hybrid instead of having the four-wire circuit. The subscriber loop can use baseband transmission, while the voiceband data set always uses passband transmission. For data transmission over a private leased line, echo cancellation is not necessary. This is because leasing a four-wire private line is only slightly more expensive than leasing a two-wire line.

A discussion of the details of full-duplex data transmission is beyond the scope of this report. The tutorial papers written by Sondhi [1] and Messerschmitt [7] can further the study on the subject. Three possible configurations for a data-driven echo canceller have been proposed by Mueller [13], Falconer [14] and Weinstein [15]. Recently, Werner [16] has written a paper to study the effects of channel impairments on the performance of a data-driven in-band echo canceller.

CHAPTER 3

APPLICATION OF STOCHASTIC GRADIENT ALGORITHM TO ECHO CANCELLATION

3.1 Minimum Mean-Square Error Solution

This chapter will present three adaptation algorithms for adapting the echo canceller constructed from a transversal filter structure. The starting point will be the discussion on the minimum mean-square error (MMSE) algorithm. This algorithm is built on the unrealistic assumptions that the input signals d(T) and y(T) are wide sense stationary discrete-time random processes with known autocorrelation and power spectrum. Although this algorithm is not practical, it can act as a starting point for the next two algorithms : the steepest descent algorithm and the stochactic gradient (SG) algorithm (also called the least-mean-square (LMS) algorithm), the former algorithm still preserves the unrealistic assumption while the latter assumes nonstationary and unknown statistics.

It will be useful to define a vector notation for the column vector of n tap coefficients for the echo canceller transversal filter,

$$\underline{C}' = [C_1 \; C_2 \; ...... \; C_n].$$

In addition define a vector of the present and (n-1) past input samples

$$\underline{Y}'(T) = [y(T) \; y(T-1) \; ...... \; y(T-n+1)].$$

Hence the error signal of the echo canceller can be written as

$$e(T) = d(T) - \underline{C}'\underline{Y}(T). \tag{3.1}$$

The MMSE approach to the problem attempts to minimize the cost function : the mean-square error(MSE) $E[e^2(T)]$.The mean-square error can be formulated from Eq.(3.1)

$$\begin{aligned} E[e^2(T)] &= E[\{d(T) - \underline{C}'\underline{Y}(T)\}^2] \\ &= E[d^2(T)] - 2\,\underline{C}'E[d(T)\underline{Y}(T)] + \\ &\quad \underline{C}'E[\underline{Y}(T)\underline{Y}'(T)]\underline{C} \\ &= E[d^2(T)] - 2\underline{C}'\underline{P} + \underline{C}'\underline{AC}, \end{aligned} \tag{3.2}$$

where

$$\underline{P} = E[d(T)\underline{Y}(T)], \tag{3.3}$$

$$\underline{A} = E[\underline{Y}(T)\underline{Y}'(T)]$$

$$= \begin{bmatrix} a_0 & a_1 & \cdot\cdot & a_{n-1} \\ a_1 & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ a_{n-1} & & \cdot\cdot & a_0 \end{bmatrix}. \tag{3.4}$$

Here $\underline{A}$ is called the autocorrelation matrix for the input process with the elements being the autocorrelation coefficients

$$a_i = E[y(T)y(T+i)]. \tag{3.5}$$

The autocorrelation matrix $\underline{A}$ has several important characteristics that make Eq.(3.2) become a quadratic form in the coefficient vector $\underline{C}$ and therefore there exists a unique minimum solution for the coefficient vector $\underline{C}$. At this point, several properties of the autocorrelation matrix are considering. [Haykin, 17]

<u>Property 1</u> The autocorrelation matrix $\underline{A}$ is symmetric, that is

$$\underline{A}' = \underline{A}. \tag{3.6}$$

Eq.(3.4) represets the matrix $\underline{A}$ and its transpose is the same as $\underline{A}$.

Property 2 The autocorrelation matrix $\underline{A}$ is Toeplitz, that is the i, j element is a function of (i-j).

The Toeplitz property of $\underline{A}$ is a direct consequence of the assumption that the input random process is stationary. Indeed, if the input is stationary and is sampled at T, the matrix $\underline{A}$ must be Toeplitz. Conversely, if the matrix $\underline{A}$ is Toeplitz, the input must be stationary.

Property 3 The autocorrelation matrix $\underline{A}$ is positive semi-definite.

Proof : Let $\underline{X}$ be an arbitrary n-by-1 vector. Define the scalar random variable :

$$b = \underline{X}'\underline{Y}(T) \tag{3.7a}$$

$$= \underline{Y}'(T)\underline{X}. \tag{3.7b}$$

The mean-square value of the random variable b equals the quadratic form $\underline{X}'\underline{A}\ \underline{X}$ as shown by

$$\begin{aligned} E[b^2] &= E[\underline{X}'\underline{Y}(T)\underline{Y}'(T)\underline{X}] \\ &= \underline{X}'E[\underline{Y}(T)\underline{Y}'(T)]\underline{X} \\ &= \underline{X}'\underline{A}\ \underline{X}. \end{aligned} \tag{3.8}$$

Since

$$E[b^2] \geq 0 \tag{3.9}$$

it follows that

$$\underline{X}'\underline{A}\ \underline{X} \geq 0. \tag{3.10}$$

Accordingly, the autocorrelation matrix is positive semi-definite. In practice, there will always be some noise added to the matrix $\underline{A}$ at all frequencies. Assume that the noise is white with noise variance σ^2, then the noise matrix is $\sigma^2\underline{I}$ and is positive definite. Hence when the noise is added to the input matrix $\underline{A}$, the quadratic form satisfies the condition that

$$\underline{X}'\underline{A}\ \underline{X} > 0, \qquad \forall\underline{X} \tag{3.11}$$

then the matrix $\underline{A}$ will be positive definite.

It remains to find the optimal coefficient vector, $\underline{C}_{opt}$. Completing the square in Eq.(3.2)

$$E[e^2(T)] = E[d^2(T)] - \underline{P}'\underline{A}^{-1}\underline{P} + (\underline{A}^{-1}\underline{P} - \underline{C})'\underline{A}(\underline{A}^{-1}\underline{P} - \underline{C}). \tag{3.12}$$

As $\underline{A}$ is positive definite, the third term is non-negative and can be minimized by choosing

$$\underline{C}_{opt} = \underline{A}^{-1}\underline{P}. \tag{3.13}$$

With this choice, the resultant minimum mean-square error becomes

$$E[e^2(T)]_{min} = E[d^2(T)] - \underline{P}'\underline{A}^{-1}\underline{P}$$

$$= E[d^2(T)] - \underline{P}'\underline{C}_{opt} . \tag{3.14}$$

In order to find $\underline{C}_{opt}$, it requires the solution of a system of linear equations. However, since the autocorrelation matrix is Toeplitz, the Levinson-Durbin algorithm can be used to solve these equations. A good description on this algorithm can be found in Chapter 2 of Proakis [18].

In the context of the echo canceller for voice (described in Chapter 2), the matrix $\underline{P}$ can be determined and the optimal coefficient vector $\underline{C}_{opt}$ can be further simplified as follows : Assume the discrete-time channel characteristics is h_i and write the desired input to the canceller in the form

$$d(T) = \sum_{i=1}^{\infty} h_i y(T-i+1) + z(T), \qquad (3.15)$$

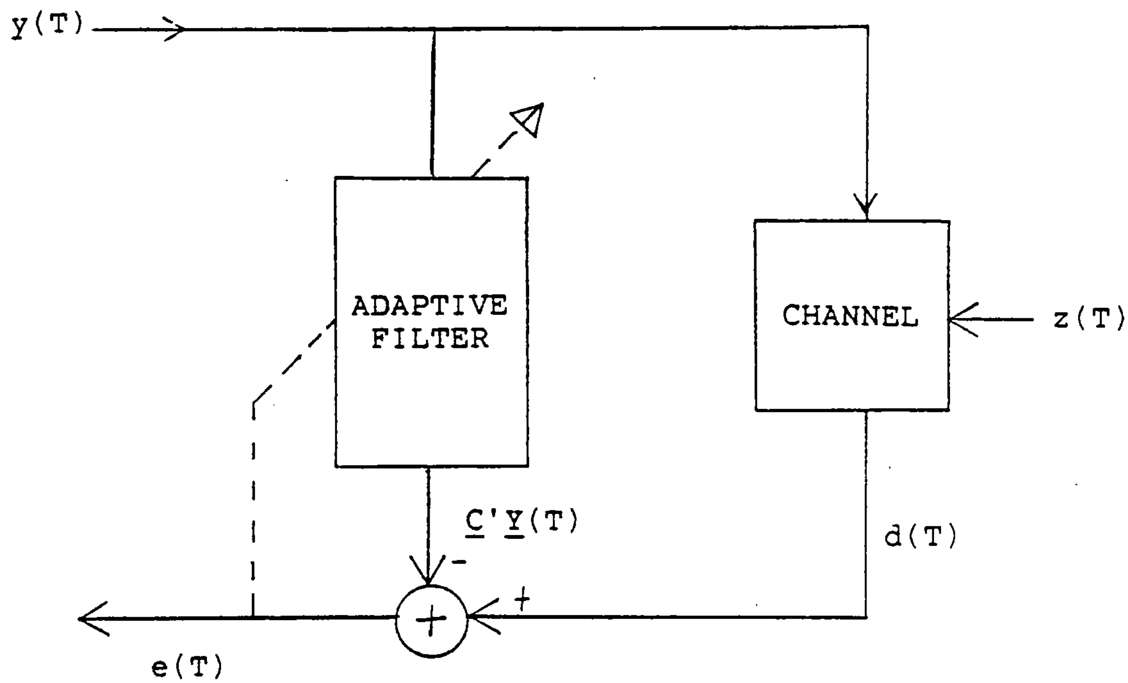


Fig. 3.1. Echo canceller with near-end possible signal plus echo and noise.

where z(T) is the near-end possible signal plus echo and noise. In addition, define the vector for the echo impulse response within the range that can be cancelled by the n-coefficient transversal filter

$$\underline{H}' = [h_1 \; h_2 \; ..... \; h_n]. \tag{3.16}$$

Let

$$v(T) = \sum_{i=n+1}^{\infty} h_i y(T-i+1) + z(T)$$

be the uncancelable portion of the desired input; that is, the echo response with delays too large for the n-coefficient echo canceller plus the near-end signal and noise. Then the desired input can be written as

$$d(T) = \underline{H}'\underline{Y}(T) + v(T). \tag{3.17}$$

Substituting Eq.(3.17) into Eq.(3.3), then

$$\underline{P} = E[\underline{Y}(T)\underline{Y}'(T)]\underline{H} + E[v(T)\underline{Y}(T)]$$

$$= \underline{AH} + E[v(T)\underline{Y}(T)], \tag{3.18}$$

and from Eq.(3.13), the optimum n-coefficient echo canceller becomes

$$\underline{C}_{opt} = \underline{H} + \underline{A}^{-1}E[v(T)\underline{Y}(T)]. \tag{3.19}$$

For the special case when $v(T)$ and $\underline{Y}(T)$ are uncorrelated then

$$\underline{C}_{opt} = \underline{H}, \tag{3.20}$$

i.e., the echo canceller just replicates the first n-coefficients of the discrete-time channel characteristics.

Before going onto the next section, it is valuable to study the principle of orthogonality. In order to find the optimal solution for the mean-square error in Eq.(3.2), it can be differentiated and set to zero.

$$\begin{aligned} 0 &= \frac{d}{dC_j} E[e^2(T)] \\ &= 2E[e(T)\frac{d}{dC_j}e(T)] \\ &= 2E[e(T)\frac{d}{dC_j}\{d(T) - \underline{C}'\underline{Y}(T)\}] \\ &= -2E[e(T)y(T-j+1)], \qquad \forall\ 1 \leq j \leq n. \end{aligned} \tag{3.21}$$

It is legitimate to interchange the order of expectation and differentiation as both operations are linear. Eq.(3.21) states that for the optimum filter, the error signal and any of the inputs are uncorrelated (orthogonal). The result is known as the principle of orthogonality. As the MSE is a quadratic function of the coefficients, there exists a unique optimal coefficient vector and Eq.(3.21) is a necessary and sufficient condition on the optimality of the coefficients.

3.2 Stochastic Gradient Algorithm

In the preceding section, in order to solve for the optimum MMSE coefficient vector, it is necessary to solve for the system of linear equations in Eq.(3.19). If the autocorrelation matrix is Toeplitz, one possible solution to solve for it is by the use of Levinson-Durbin algorithm. However, the most widely used practical algorithm for adaptation of an echo canceller is the stochastic gradient (SG) algorithm.

As a starting point, the ensemble statistics of the reference input y(T) is assumed to be known again and a steepest descent algorithm is just derived before the study of the stochastic gradient algorithm. The steepest descent approach is to develop a recursive procedure whereby appropriate corrections are applied to the filter coefficients which guarantee convergence to the optimum coefficient vector (since the output mean-square error, MSE, given by Eq.(3.2) is a quadratic form in the coefficient vector $\underline{C}$, there exists a unique minimum). The recursive procedure starts from an arbitrary point on the MSE surface and continuous adjusting the coefficient vectors iteratively until the minimum point of the MSE surface has been reached. Given the present coefficient $\underline{C}(T)$, with index T corresponds to the iteration number for the iterative algorithm for solving a system of linear equation, successive corrections to $\underline{C}(T)$ are made by subtracting off a term proportional to the error gradient $\nabla_c \{E[e^2(T)]\}$. Thus the resultant tap vector should be close to $\underline{C}_{opt}$. Explicitly, the approach is

$$\underline{C}(T+1) = \underline{C}(T) - \beta\nabla_c \{E[e^2(T)]\}, \tag{3.22}$$

where β is the proportionality constant or step size. The reason why the approach works is that error gradient is a vector in the direction of the maximum increase of the error. Moving a short distance (i.e. with small step size) in the negative direction of the gradient should then reduce the error. However, moving too far in that direction might result in overshoot away from the optimum and divergence will occur.

At the minimum point of the MSE surface, the components of the error gradient vector

$$\nabla_c \{E[e^2(T)]\} = -2E[e(T)\underline{Y}(T)] \tag{3.23}$$

are simultaneously zero. Accordingly, the principle of orthogonality has stated that for the minimum MSE, the error signal and the input signal are uncorrelated.

Fig.3.2 shows the illustration of the steepest descent algorithm for the second order case. The contours in the plane of filter coefficients of constant MSE have an elliptical shape. The negative of the gradient points in the

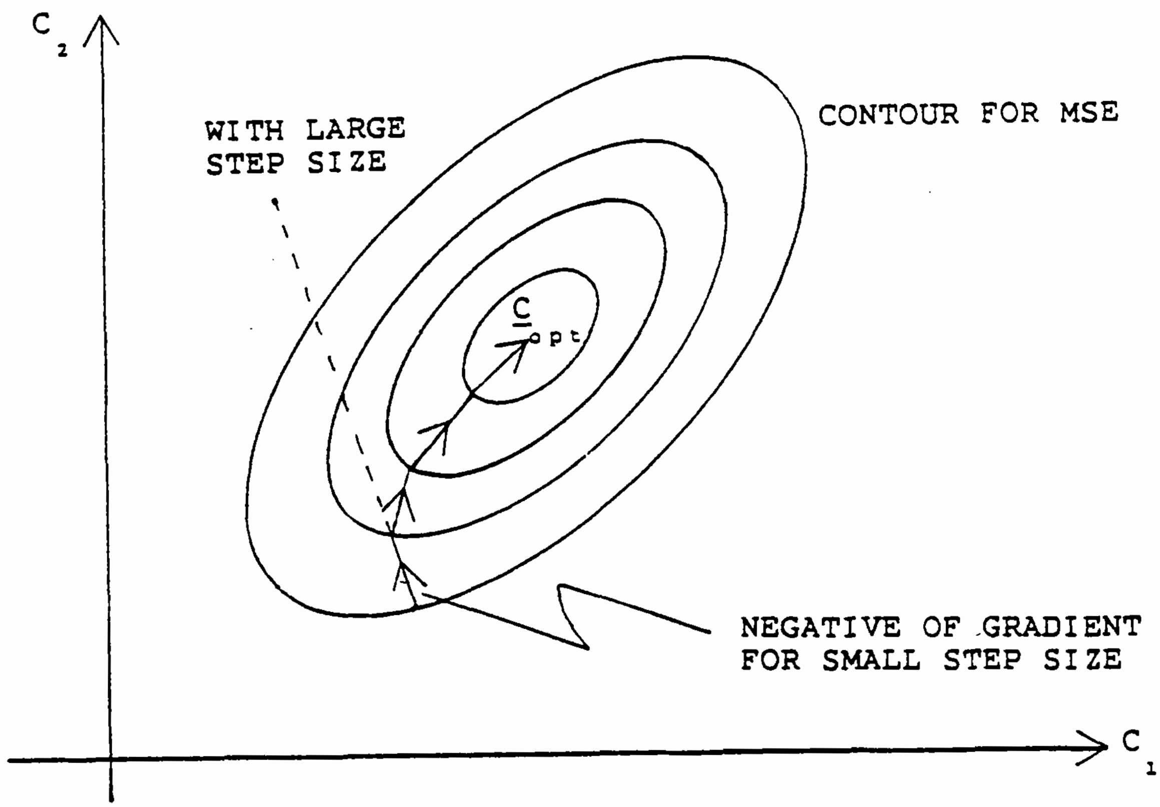


Fig. 3.2. Illustration for steepest descent algorithm.

maximum decrease of the MSE. With a small step size, the MSE is reduced incrementally. With a large step size, overshoot may occur to a point where the MSE will become larger and instability will occur.

Now concerning the case when the underlying statistics are not known in advance, then the expectation operator in Eq.(3.22) will become troublesome part. The principle behind the stochastic gradient (SG) algorithm is to neglect the expectation operator. As a result, the deterministic error gradient vector is replaced by a random stochastic estimate vector

$$\tilde{\nabla}_c [e^2(T)] = -2e(T)\underline{Y}(T). \tag{3.24}$$

Note that this estimate is unbiased because its expected value is exactly the same as the actual gradient vector of Eq.(3.23).

The stochastic gradient algorithm changes the filter coefficient vector along the direction of the gradient vector estimate as follows

$$\begin{aligned} \underline{C}(T+1) &= \underline{C}(T) - 0.5\beta\tilde{\nabla}_c [e^2(T)] \\ &= \underline{C}(T) + \beta e(T)\underline{Y}(T) \end{aligned} \tag{3.25}$$

where

T = iteration number that corresponds to the sample number (or time index) of the given input data.

$\underline{C}(T)$ = filter coefficient vector before estimation.

$\underline{C}(T+1)$ = filter coefficient vector after estimation.

β = step size parameter.

$e(T)$ = error signal at the T-th iteration.

$\underline{Y}(T)$ = tap input vector at the T-th iteration.

The difference between the stochastic gradient algorithm in Eq.(3.25) and the steepest descent algorithm in Eq.(3.22) is that in the latter case, the coefficient vector follows a predictable trajectory, while in the former case, the trajectory is noisy and stochastic. This random fluctuation is the cost of using the time average in place of ensemble average. Hence, due to the fact that a noisy estimate of the error gradient is used to adapt the filter, the coefficients will have some asymptotic variance that causes the asymptotic MSE to be greator than the minimum MSE obtained from an optimum fixed coefficient filter.

3.3 Convergence Properties of the SG Algorithm

The convergence of the stochastic gradient (SG) algorithm can be analyzed in two ways. The first way is to determine analytically how the average of the coefficient vector trajectories converges to the optimum. The second way is to determine analytically how the ensemble average of the mean-square error trajectories converges to the optimum. Notice that in both cases, the coefficients and the mean-square errors are fluctuating due to the presence of the stochastic terms, thus only the average trajectories are determined in this section. In other words, the analysis here does not represent the particular coefficient vector trajectory or mean-square error trajectory that converge to the optimum.

3.3a The average trajectory of the coefficient vector

In order to analyse the convergence behavior of a given data adaptive algorithm, the given data is often assumed to be wide-sense stationary random process with known autocorrelation function statistics (though this is unrealistic). Hence in Eq.(3.25), the expectation of the coefficient vector should be taken, i.e.

$$E[\underline{C}(T+1)] = E[\underline{C}(T)] + \beta E[e(T)\underline{Y}(T)]$$

$$= E[\underline{C}(T)] + \beta E[\{d(T) - \underline{C}'(T)\underline{Y}(T)\}\underline{Y}(T)]$$

$$= E[\{\underline{I} - \beta\underline{Y}(T)\underline{Y}'(T)\}\underline{C}(T)] + \beta E[d(T)\underline{Y}(T)], \qquad (3.26)$$

where $\underline{I}$ is the identity matrix. To facilitate the analysis, an approximation can be made that the coefficient vector is independent of the input data samples. This approximation is valid when β is small, which gives rise to the slow trajectory of $\underline{C}(T)$ and makes the coefficient vector and the input data samples approximately uncorrelated. In this way, Eq.(3.26) becomes

$$E[\underline{C}(T+1)] \doteq (\underline{I} - \beta\underline{A})E[\underline{C}(T)] + \beta\underline{P}. \qquad (3.27)$$

If this algorithm is simply iterated for some arbitrary initial guess $\underline{C}(1)$, it will converge to $\underline{C}_{opt}$ as in Eq.(3.13).

If a mean coefficient-error vector is defined as

$$E[\underline{Z}(T)] = E[\underline{C}(T) - \underline{C}_{opt}]$$

$$= E[\underline{C}(T)] - \underline{C}_{opt}, \qquad (3.28)$$

then subtract $\underline{C}_{opt}$ from both sides of Eq.(3.27), it gives

$$E[\underline{Z}(T+1)] \doteq (\underline{I} - \beta\underline{A})E[\underline{Z}(T)] - \beta\underline{AC}_{opt} + \beta\underline{P}. \qquad (3.29)$$

As the optimum coefficient vector is given by Eq.(3.13) (repeated here for convenience)

$$\underline{C}_{opt} = \underline{A}^{-1}\underline{P},$$

then Eq.(3.29) can be further simplified as

$$E[\underline{Z}(T+1)] \doteq (\underline{I} - \beta\underline{A})E[\underline{Z}(T)]. \qquad (3.30)$$

Iterating Eq.(3.30), then one can get

$$E[\underline{Z}(T+1)] \doteq (\underline{I} - \beta\underline{A})^{T}E[\underline{z}(1)]. \qquad (3.31)$$

In determining whether this mean coefficient-error vector approaches zero, several important properties of the matrix $\underline{A}$ are again investigated. The matrix is symmetric, Toeplitz and positive definite. The last property implies that the matrix has positive real eigenvalues and is invertible. $\underline{A}$ can be written as

$$\underline{A} = \underline{MVM}', \tag{3.32}$$

where the diagonal matrix $\underline{V}$ consists of the eigenvalues of $\underline{A}$,

$$\underline{V} = \mathrm{diag}[\ \lambda_1\ .....\ \lambda_n\ ], \tag{3.33}$$

and $\underline{M}$ is an orthonormal matrix with

$$\underline{MM}' = \underline{I}, \tag{3.34}$$

where the columns of $\underline{M}$ ($\underline{M}_1$, $\underline{M}_2$, ..., $\underline{M}_n$) denote the eigenvectors of the eigenvalues λ_1, ..., λ_n respectively. Then premultiply both sides of Eq.(3.31) by $\underline{M}'$ gives

$$\begin{aligned} E[\underline{M}'\underline{Z}(T+1)] &= \underline{M}'(\underline{I} - \beta\underline{A})E[\underline{Z}(T)] \\ &= E[\underline{M}'\underline{Z}(T)] - \beta\underline{M}'\underline{A}E[\underline{Z}(T)] \\ &= E[\underline{M}'\underline{Z}(T)] - \beta\underline{M}'\underline{AMM}'E[\underline{Z}(T)] \\ &= E[\underline{M}'\underline{Z}(T)] - \beta\underline{V}E[\underline{M}'\underline{Z}(T)] \\ &= (\underline{I} - \beta\underline{V})E[\underline{M}'\underline{Z}(T)]. \end{aligned} \tag{3.35}$$

Define the transformed coefficient as

$$\underline{U}(T) = \underline{M}'\underline{Z}(T). \tag{3.36}$$

Accordingly, rewrite Eq.(3.35) as

$$E[\underline{U}(T+1)] \doteq (\underline{I} - \beta\underline{V})E[U(T)]. \qquad (3.37)$$

Since $\underline{V}$ is diagonal, Eq.(3.37) can be written in component form as

$$E[u_j(T+1)] \doteq (1 - \beta\lambda_j)^T E[u_j(1)], \qquad (3.38)$$

where u_j is the j component of $\underline{U}$. The mean coefficient-error vector,

$$E[\underline{Z}(T)] = \underline{M}E[\underline{U}(T)],$$

will decay exponentially to zero as long as the condition

$$| 1 - \beta\lambda_i | < 1 \qquad \forall = 1, \ldots, n$$

is satisfied. Since the eigenvalues are distinct and can be ordered from the smallest eigenvalue, λ_{min} to the largest eigenvalue, λ_{max}, the condition becomes

$$0 < \beta < 2/\lambda_{max} . \qquad (3.39)$$

In other words, the stochastic gradient algorithm converges in the mean, i.e. $E[\underline{C}(T)]$ approaches $\underline{C}_{opt}$ as the number of iteration, T, approaches infinity when the step size satisfies the condition (3.39).

Notice that with the choice of β [Gersho, 19] to be

$$\beta_{opt} = \frac{2}{\lambda_{max} + \lambda_{min}}$$

eigenvalues converge at the same fastest rate i.e. proportional to

$$\left\{ \frac{\lambda_{max}/\lambda_{min} - 1}{\lambda_{max}/\lambda_{min} + 1} \right\}^T$$

This rate will be increased when $\lambda_{max}/\lambda_{min}$ is decreased. Thus the parameter $\lambda_{max}/\lambda_{min}$, is seen to be of fundamental importance; it is called the eigenvalue spread. The eigenvalue spread has a minimum value of one, and can be arbitrary large. The larger the eigenvalue spread, the slower the convergence of the gradient algorithm.

It is instructive to relate the eigenvalue spread to the power spectrum of the reference random process. From a classical result of Toeplitz theory that

$$\min_w S(w) < \lambda_j < \max_w S(w) \tag{3.40}$$

where S(w) is the power spectrum. The eigenvalues depend on the order of the matrix n, as $n \rightarrow \infty$, then

$$\lambda_{max} \rightarrow \max_w S(w) \tag{3.41a}$$

$$\lambda_{min} \rightarrow \min_w S(w). \tag{4.41b}$$

This follows that when the amplitude of the spectrum is almost flat (close to white noise), then the eigenvalue spread is close to one and results in fast convergence.

One final point that can be studied from the derivation is the time constant. From Eq.(3.38), the time constant associated with the j-th normal mode (time for $E[\underline{Z}(T)]$ to decay to $\underline{Z}(1)/e$) is

$$\tau_j = \frac{-1}{\ln(1 - \beta\lambda_j)} \doteq \frac{1}{\beta\lambda_j} . \tag{3.42}$$

This shows that the convergence of the mean coefficient vector is limited by the smallest eigenvalue λ_{min}, which produces the largest time constant τ_{min}.

3.3b The ensemble average trajectory of the MSE

In this section, the output mean-square error for the deterministic steepest descent algorithm will be studied first. Subsequently, the average mean-square error of the stochastic gradient algorithm will be touched slightly in the later section.

In section 3.1, it has already shown that with optimum tap coefficient vector, $\underline{C}_{opt}$, the minimum mean-square error ϵ_{min} is obtained as Eq.(3.14), (repeated here for convenience)

$$\epsilon_{min} = E[e^2(T)]_{min} = E[d^2(T)] - \underline{P}'\underline{A}^{-1}\underline{P}.$$

Hence the expression for the mean-square error in Eq.(3.2) can be rewritten as

$$\begin{aligned} \epsilon(T) &= E[e^2(T)] \\ &= \epsilon_{min} + [\underline{C}(T) - \underline{C}_{opt}]'\underline{A}[\underline{C}(T) - \underline{C}_{opt}]. \qquad (3.43) \end{aligned}$$

Although the quadratic form on the right hand side of Eq.(3.43) is quite informative, it is desirable to diagonalize the autocorrelation matrix $\underline{A}$ and thus the new expression for the mean-square error becomes

$$\epsilon(T) = \epsilon_{min} + [\underline{C}(T) - \underline{C}_{opt}]'\underline{MVM}'[\underline{C}(T) - \underline{C}_{opt}]. \quad (3.44)$$

Now let the M-by-1 vector $\underline{U}$ denote the transformed coefficient error vector.

$$\underline{U}(T) = \underline{M}'[\underline{C}(T) - \underline{C}_{opt}]. \quad (3.45)$$

Substitute Eq.(3.45) into Eq.(3.44), the MSE becomes

$$\epsilon(T) = \epsilon_{min} + \underline{U}(T)'\underline{V}\,\underline{U}(T)$$

$$= \epsilon_{min} + \sum_{i=1}^{n} \lambda_i U_i^2(T) \quad (3.46)$$

where $U_i(T)$ is the i-th component of the transformed coefficient vector. To determine the MSE for a coefficient vector $\underline{C}$, Eq.(3.46) says that the components of this vector in the direction of the eigenvector are found, square the components and multiply by the corresponding eigenvalue. This implies that the MSE increases most rapidly in the direction of the eigenvector corresponding to λ_{max} and most slowly in the direction corresponding to λ_{min}. This can be understood with the help of Fig.3.3. Eq.(4.46) illustrates that the contours of equal MSE are elliptical in shape with the axes in the direction of the eigenvectors. The eccentricity of these contours is directly related to the eigenvalue spread. Fig.3.3 depicts the case of a second order coefficient vector with $\lambda_2 > \lambda_1$, then the major axis of the ellipse is in the direction of $\underline{M}_1$ and the minor axis is in the direction of $\underline{M}_2$. The MSE increases more slowly in the direction of $\underline{M}_1$.

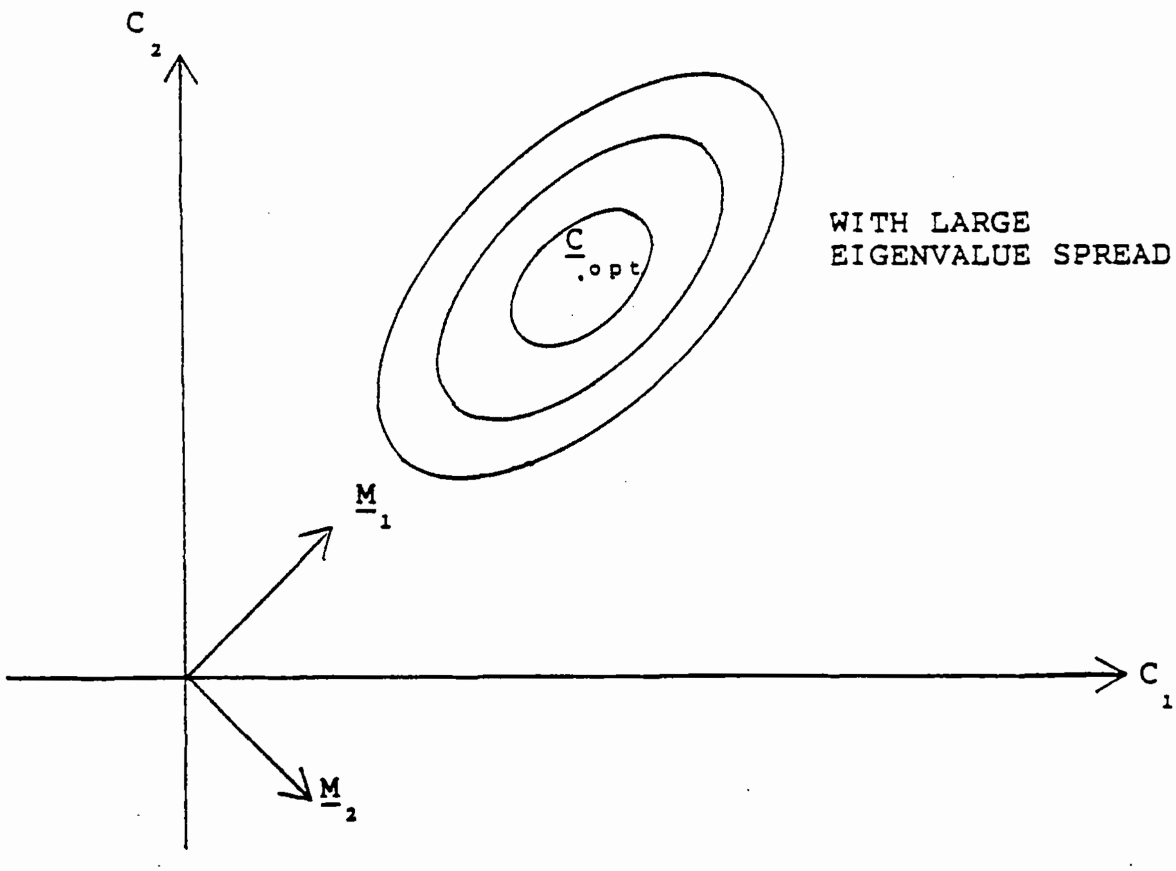


Fig. 3.3. Contours of equal MSE.

The above derivation of the MSE is for the deterministic steepest descent algorithm. In the case of the stochastic gradient, it is necessary to ensemble average over the filter coefficients. Thus

$$\epsilon(T) = \epsilon_{min} + \sum_{i=1}^{n} \lambda_i E[U_i^2(T)] . \qquad (3.47)$$

Ideally speaking, the SG algorithm is convergent, i.e. irrespective of the initial conditions

$$\lim_{T\to\infty} \epsilon(T) = \epsilon_{min} .$$

The curve obtained by plotting the MSE, $\epsilon(T)$ versus the number of iterations, T, is called a learning curve. However, in the practical case for the SG algorithm, it relies on a noisy estimate for the gradient vector, with the result that the coefficient vector executes small fluctuation about $\underline{C}_{opt}$. Hence, even after a large number of iterations, the actual value of $\epsilon(\infty)$ is greater than ϵ_{min} which is called excess mean-square error. In addition, the learning curve consists of very noisy decaying exponentials, and only the ensemble average of the noisy learning curves can be computed. Thus the average MSE denoted by $E[\epsilon(T)]$ is used to describe the dynamic behavior of the SG algorithm.

The mathematical evaluations of the excess MSE and the average MSE are outside the scope of this paper. Further study can be found in Ungerboeck [20], Gitlin [21] and Honig [22].

CHAPTER 4

EFFECTS OF USING DIFFERENT INPUT SPECTRUM ON THE ECHO CANCELLER

4.1 Conditions for Unique Tap Values

This section will determine the conditions under which the filter coefficients of an echo canceller are unique. In section 3.1 and 3.2, the existence of unique filter coefficients to the MMSE problem and the argument for the convergence of the steepest descent and stochastic algorithm depended of the nonsingularity of the autocorrelation matrix $\underline{A}$. There are times when the autocorrelation matrix $\underline{A}$ beomes singular and results in nonunique tap setting. Since $\underline{A}$ is always positive semi-definite, $\underline{A}$ can only be singular when one or more of its eigenvalues are zero. In section 3.3, Eq.(3.41) has shown that when the spectrum of the reference input vanishes at some frequencies, there would possible be one or more zero eigenvalues.

A condition for $\underline{A}$ to be singular is the vanishing of the quadratic form $\underline{X}'\underline{A}\ \underline{X}$ for any nonzero test vector $\underline{X}$, i.e. $\underline{A}$ is singular when

$$\underline{X}'\underline{A}\ \underline{X} = 0, \tag{4.1}$$

as this condition implies that $\underline{A}$ is positive semi-definite but not positive definite. Eq.(4.1) can be expressed in summation format and taking into consideration that the components of $\underline{A}$ are given by Eq.(3.5) (repeated here for convenience)

$$a_i = E[y(T)y(T+i)]. \qquad \forall\ 1 \leq i \leq n$$

Thus Eq.(4.1) becomes

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{i-j} = 0. \tag{4.2}$$

Defining the power spectrum of the reference random process as the Fourier transform of the autocorrelation function

$$S(e^{jw}) = \sum_{T=-\infty}^{\infty} a_T e^{-jwT}, \tag{4.3a}$$

$$a_T = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(e^{jw})\, e^{jwT} dw. \tag{4.3b}$$

Substituting (4.3b) into (4.2), and get

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} S(e^{jw})\,|x(e^{jw})|^2\,dw = 0, \qquad (4.4)$$

where

$$x(z) = \sum_{i=0}^{n-1} x_{i+1} z^{-i}$$

is the Z-transform of the vector $\underline{X}$. Since the integrand of Eq.(4.4) is nonnegative, the integral becomes zero when the integrand is identically zero.

$$S(e^{jw})|x(e^{jw})|^2 = 0, \qquad 0 \leq w \leq \pi. \qquad (4.5)$$

Hence, this is the condition for $\underline{A}$ to be singular. Since x(z) is an (n-1)-th order polynomial in z, it has at most (n-1) zeros. Thus (4.5) cannot be satisfied (i.e. $\underline{A}$ is nonsingular) as long as the spectrum $S(e^{jw})$ is nonzero at n or more frequencies. For example, if n is equal to three and if the input signal consists of a sinusoid, here the spectrum of the sinusoid is only nonzero for two frequencies. Thus, Eq.(4.5) is satisfied, and $\underline{A}$ will be singular and the MMSE solution will not be unique. This can be understood intuitively, as a sinusoid has only 2 parameters, the amplitude and phase of a sinusoid, and transversal filter has 3 degrees of freedom (filter coefficients). There is thus one degree of freedom left over after the filter transfer function is adjusted at the two frequencies. This happens in the echo canceller context, as the transversal filter has 3 or more taps usually and when the input is a periodic sine wave, (i.e. sending a tone as a training sequence in the voice channel) the filter coefficients may become nonunique and the transfer function can assume any values at the frequencies other than the particular frequencies of the input sinusoid. This nonuniqueness situation will cause some problems and will be discussed in the next section.

4.2 Problems Created By a Singular Input Matrix

An input signal that does not cover the full bandwidth is a legitimate concern in echo cancellation. This will give rise to an ill-conditioning or singularity problem of the input autocorrelation matrix $\underline{A}$. In practice, there will always be some noise added to the matrix $\underline{A}$ at all frequencies, and thus $\underline{A}$ will become positive definite too. As a symmetric positive definite matrix implies nonsingularity, it can be considered that where there is noise present, the input matrix $\underline{A}$ is still nonsingular. However, for vanishing small noise case, the input matrix $\underline{A}$ can become singular and thus the study of how a singular matrix $\underline{A}$ affect the performance of the adaptive filter is still significant.

Two important problems will arise as a result of this singularity. First, as the tap coefficients become nonunique and owing to the random component in the algorithm's correction term, there will be a continuous change of the filter coefficients about the optimum solutions. This phenomenon is called tap drifting problem. When the taps have drifted for a while, and at this moment when the input energy is changed to white noise, the errors in the tap coefficients will cause the echo canceller an increase in the MSE suddenly. In addition to this degraded performance, in most analog and digital implementation, there is always a maximum value for each tap coefficient register. Sometimes, tap drifting may result in overflow in the tap coefficient register. These tap drifting effects will be studied in detail in the following paragraphs. Second, the distribution of the eigenvalues of $\underline{A}$ will affect the rate of convergence of the adaptive filter to the optimum setting. Since some small eigenvalues in the matrix will increase the eigenvalue spread, the larger the eigenvalue spread of the autocorrelation matrix, the slower the convergence of the filter coefficients. The problem of the rate of convergence will be dealt with in Chapter 6 by using a new kind of adaptive structure and adaptation algorithm.

The random tap drifting problem can be understood with the help of the illustration in Fig.4.1. The two axes correspond to the amplitudes of the coefficients of a second order filter and the shaded area represents the region where the coefficient registers have been saturated, i.e. either one

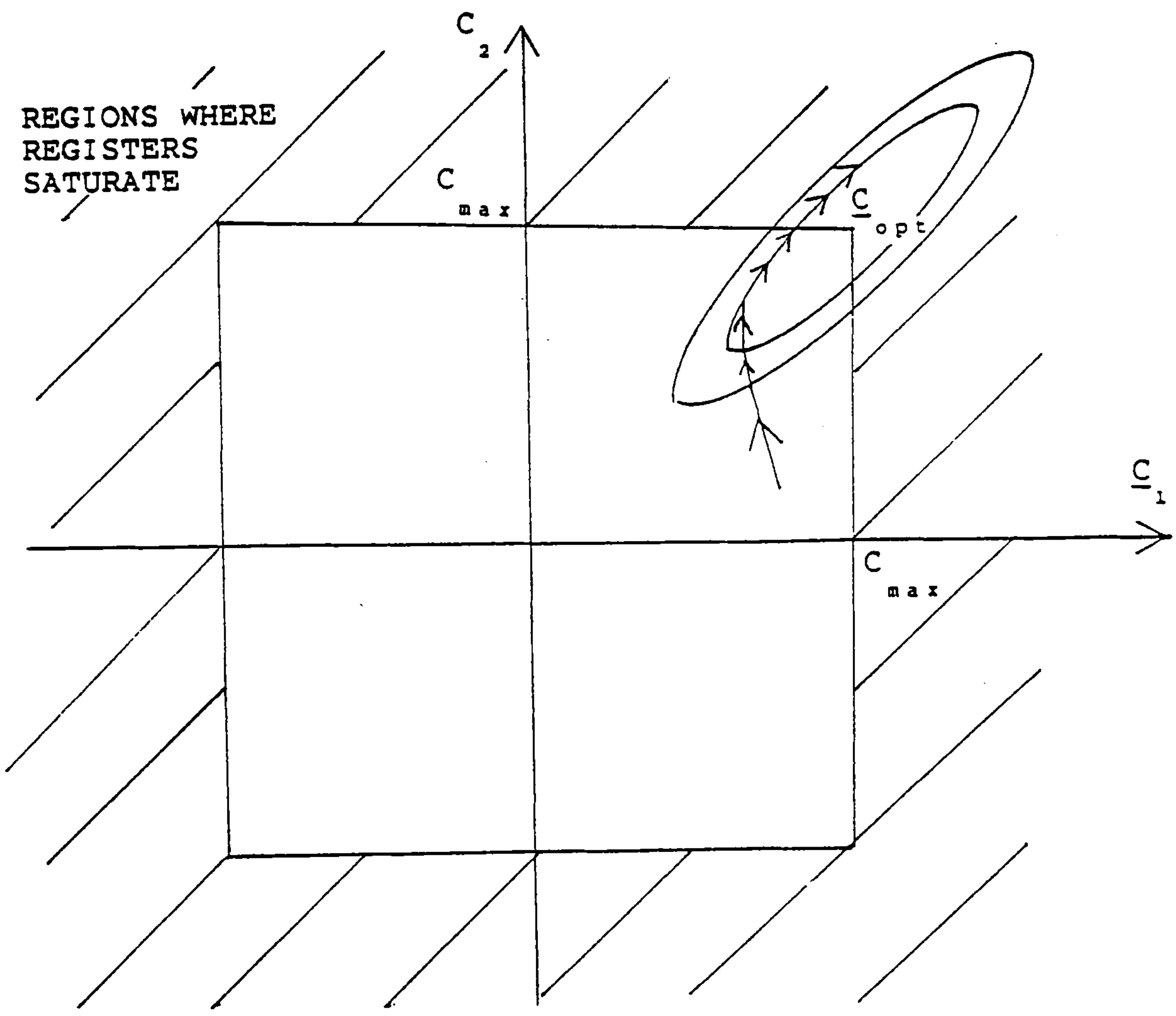


Fig. 4.1. Contours of equal MSE and tap convergence.

of the coefficient values has exceeded $\underline{C}_{max}$. During adaptation, the tap will shift from $\underline{C}_1$ initially and converge to the optimal coefficient vector $\underline{C}_{opt}$.

In practice, due to the random component in the algorithm's update term, the fluctuation of the filter coefficients about the optimum will continue. The fluctuation of the coefficients in the direction of the eigenvector corresponding to the minimum eigenvalue will tend to be larger (as shown in Section 3.3b). Eq.(3.42) of Section 3.3a has shown that the settling time of the stochastic gradient algorithm is limited by the smallest eigenvalue. Since the time constant depends on $(1/\beta\lambda)$, as the eigenvalue get vanishing small, the time constant will become very large.

As the tap coefficients keep fluctuating, there is a probability that the taps will become so large that one or more registers will saturate. The main reason why the saturation occurs is the presence of a bias component in a digital implementation that can drive the tap to a large value. A typical mechanism for such a bias is the two's complement type of quantizing characteristic. To quantify the discussion, the bias is denoted by a time invariant vector $\underline{K}$, and the adjustment algorithm becomes

$$E[\underline{C}(T+1)] = E[\underline{C}(T)] + \beta E[e(T)\underline{Y}(T)] + \beta\underline{K}, \qquad (4.6)$$

again define the coefficient error vector as

$$\underline{Z}(T) = \underline{C}(T) - \underline{C}_{opt} . \tag{4.7}$$

Subtracting $\underline{C}_{opt}$ from both sides of (4.6), then

$$E[\underline{Z}(T+1)] = E[\underline{Z}(T)] + \beta E[e(T)\underline{Y}(T)] + \beta \underline{K}$$

$$= (\underline{I} - \beta \underline{A})E[\underline{Z}(T)] + \beta \underline{K} , \tag{4.8}$$

and thus the steady-state mean tap error satisfies

$$E[\underline{Z}(\infty)] = \underline{A}^{-1}\underline{K}. \tag{4.9}$$

If λ_i and $\underline{M}_i$ denote the i-th eigenvalue and eigenvector of $\underline{A}$, then

$$E[\underline{Z}(\infty)] = \sum_{i=1}^{n} (\underline{M}_i \underline{K} \underline{M}'_i)/\lambda_i .$$

Clearly, if there is a small eigenvalue whose eigenvector is not orthogonal to $\underline{K}$, then the steady-state tap error can be quite large.

Now assume the echo canceller has been trained under a narrowband input for a long time, there has been a significant increase in the tap coefficient error already. The tap coefficients registers can be considered as not being overflowed yet. If suddenly, the input has been changed to white noise, then the echo canceller will have an increase in the MSE suddenly. This can be explained with the help of the illustration in Fig.4.2. The contour maps for the white noise

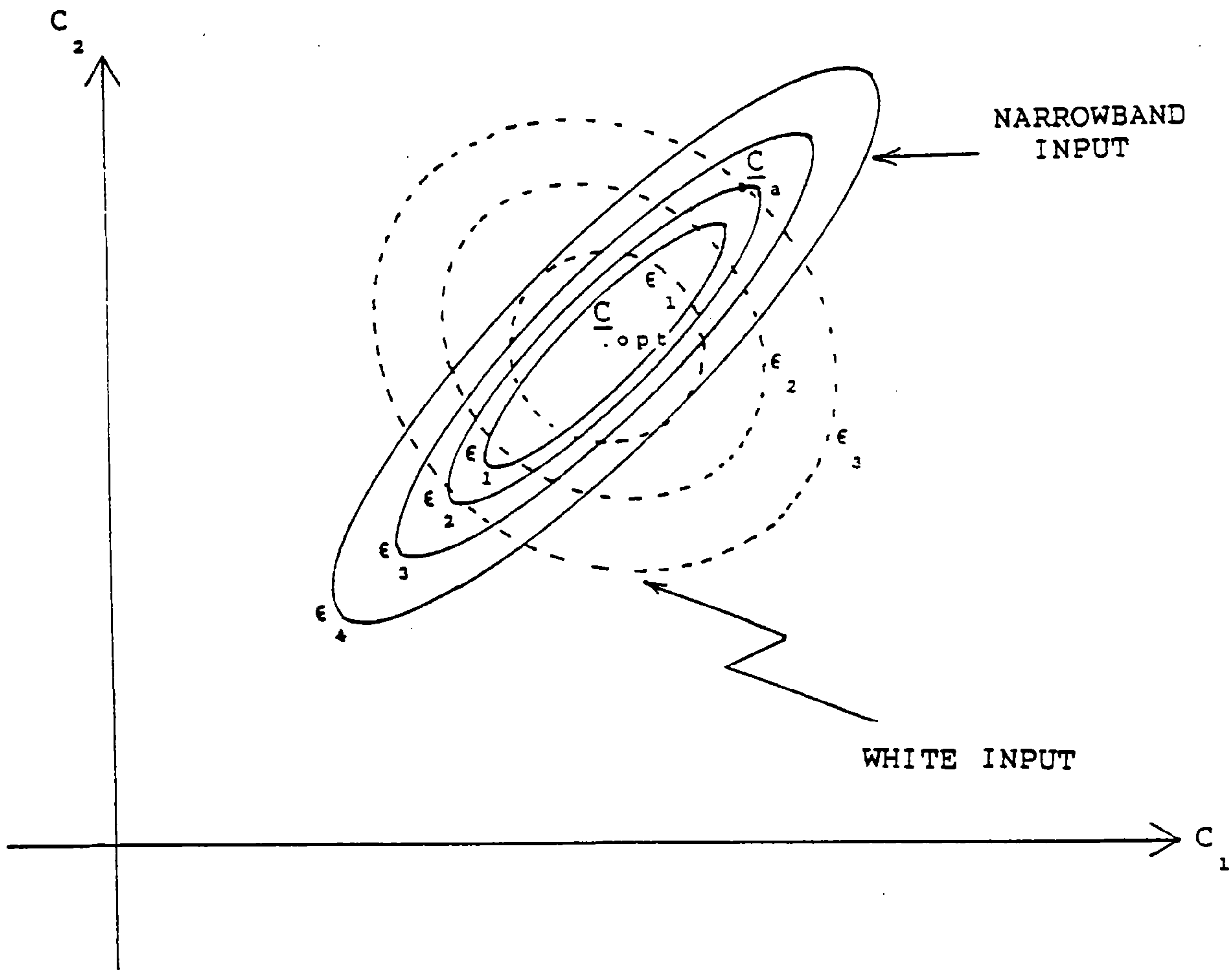


Fig. 4.2. Contours of equal MSE for narrowband input and white input.

and narrowband input are drawn. They both have the same optimum coefficient vector, $\underline{C}_{opt}$. Assume that the MSE contours are arranged in the order $\epsilon_1 < \epsilon_2 < \epsilon_3 < \epsilon_4$. The point $\underline{C}_a$ corresponds to the coefficients that have been reached after running the narrowband input for a long time, the MSE is just ϵ_2. However, when white noise is input, the MSE corresponds to this coefficient vector at $\underline{C}_a$ has increased immediately to ϵ_3.

One way to deal with this tap drifting problem is to add some noise to the input signal and thus making the input matrix $\underline{A}$ nonsingular with large eigenvalues. Usually, the noise cannot be depended alone and alternate ways have been proposed. There are two methods to cope with the case that the tap coefficients have been drifted beyond the maximum value of the coefficient registers. One is the use of the saturation mathematics which maintains the coefficients on the boundary of the allowed region. A second method, that would be studied in detail in the next chapter, is to put a leakage term in the adjustment algorithm that tends to force the coefficients toward the origin, thus keeping the coefficients smaller.

CHAPTER 5

THE TAP-LEAKAGE COEFFICIENT ADJUSTMENT ALGORITHM

The reason to employ the tap-leakage adjustment algorithm is to prevent the tap coefficients from drifting to large values. The idea of the algorithm is to use a leakage term to decrease the magnitudes of the tap coefficients. Now this chapter will help to study the algorithm in detail.

5.1 Tap-Leakage Algorithm

The tap-leakage adjustment algorithm is being used in several important applications of adaptive filters in communication such as ADPCM coder system for telephone transmission [Evci et al 23, Nishitani et al 24] and fractional spaced equalizers (FSE) to enhance modem performance [Giltin 10]. In the former case, the algorithm is added to the conventional ADPCM encoding scheme in order to improve the stability and robustness to transmission error. In the latter case, the fractional spaced equalizers will generally have many sets of tap values even though each set gives nearly equal values of mean-square error and the algorithm helps to prevent the biases that cause register overflows. However, in both applications, the adjustment algorithms result in some small degradation in the performance of the adaptive filters.

Instead of minimizing the mean-square error, the algorithm control large-tap build up by minimizing the augmented cost function

$$J = E[e^2(T)] + \mu \underline{C}'(T)\underline{C}(T), \qquad (5.1)$$

where $E[e^2(T)]$ is the MSE and μ is a suitably chosen small constant. The function J ascribes a quadratic penalty to the magnitude of the tap vector. The stochastic gradient algorithm corresponding to minimizing (5.1) is

$$\begin{aligned} \underline{C}(T+1) &= \underline{C}(T) - 0.5\beta\nabla_C\{E[e^2(T)] + \mu\underline{C}'(T)\underline{C}(T)\} \\ &= \underline{C}(T) - 0.5\beta\{-2E[e(T)\underline{Y}(T)] + 2\mu\underline{C}(T)\} \\ &= (1 - \beta\mu)\underline{C}(T) + \beta E[e(T)\underline{Y}(T)]. \qquad (5.2) \end{aligned}$$

Since the taps are adjusted adaptively, μ cannot be interpreted as a Lagrange multiplier. The use of Lagrange multipler would be appropriate if J is minimized in a deterministic manner by using the true gradient. If the gradient of $E[e^2(T)]$ with respect to $\underline{C}(T)$ is not available, μ must be chosen beforehand by using some knowledge of the system parameters. Similar to the cases of speech coding and FSE, when $\underline{C}(T)$ is chosen to minimize the augmented cost function J via the SG algorithm, there is a degradation in the minimum attainable steady state MSE. In other words, there will be an increase in the steady state MSE as a result of controlling the tap drifting problem. Apply the expectation operator to both sides of Eq.(5.2) and again assume that the input data sample is independent of the coefficient vector, then

$$\begin{aligned} E[\underline{C}(T+1)] &= (1 - \beta\mu)E[\underline{C}(T)] + \beta E[e(T)\underline{Y}(T)] \\ &= (1 - \beta\mu)E[\underline{C}(T)] \\ &\quad + \beta E[\{d(T) - \underline{C}'(T)\underline{Y}(T)\}\underline{Y}(T)] \\ &= \{(1 - \beta\mu)\underline{I} - \beta\underline{Y}(T)\underline{Y}'(T)\}E[\underline{C}(T)] \\ &\quad + \beta E[d(T)\underline{Y}(T)], \end{aligned} \tag{5.3}$$

with

$$\underline{A} = E[\underline{Y}(T)\underline{Y}'(T)] \qquad \underline{P} = E[d(T)\underline{Y}(T)]$$

Eq.(5.3) becomes

$$E[\underline{C}(T+1)] = \{\underline{I} - \mathrm{diag}(\beta\mu) - \beta\underline{A}\}E[\underline{C}(T)] + \beta\underline{A}, \tag{5.4}$$

where diag(c) is a diagonal matrix with the elements c. For stationary signals, the long term coefficient estimation is given by

$$\underline{C}(\infty) = E[\underline{C}(\infty)] = [\mathrm{diag}(\beta\mu) + \beta\underline{A}]^{-1}\beta\underline{P}$$

$$= [\mathrm{diag}(\mu) + \underline{A}]^{-1}\underline{P}, \qquad (5.5)$$

but

$$\underline{C}_{opt} = \underline{A}^{-1}\underline{P}$$

thus

$$\underline{C}(\infty) = [\mathrm{diag}(\mu) + \underline{A}]^{-1}\underline{A}\underline{C}_{opt} \; . \qquad (5.6)$$

The difference between $\underline{C}_{opt}$ and $\underline{C}(\infty)$ can be further illustrated by the decomposition of the matrix $\underline{A}$ in terms of its n positive eigenvalues :

$$\underline{A} = \underline{M}\mathrm{diag}(\lambda_i)\underline{M}'$$

where $\underline{M}$ is an orthonormal modal (n×n) square matrix with

$$\underline{M}\underline{M}' = \underline{I}$$

Eq.(5.6) can be written as

$$\underline{C}(\infty) = \underline{M}\text{diag}\left(\frac{\lambda_i}{\lambda_i + \mu}\right)\underline{M}'\underline{C}_{opt} \quad . \tag{5.7}$$

If the leakage term is sufficiently small with respect to λ_i and if the matrix $\underline{A}$ is not too ill-conditioned, i.e. if the dynamic range of the input signal spectrum is not large, then take the approximation that

$$\frac{\lambda_i}{\lambda_i + \mu} \doteq 1 - (\mu/\lambda_i) \; ,$$

then substitute this into Eq.(5.7)

$$\begin{aligned} \underline{C}(\infty) &\doteq \underline{M}\text{diag}(1 - \mu/\lambda_i)\underline{M}'\underline{C}_{opt} \\ &= \underline{C}_{opt} - \underline{M}\text{diag}(\mu/\lambda_i)\underline{M}'\underline{C}_{opt} \\ &= (\underline{I} - \mu\underline{A}^{-1})\,\underline{C}_{opt} \quad . \end{aligned} \tag{5.8}$$

Recall that with the optimal set of coefficient, this leads to the minimum MSE, ϵ_{min}. While the introduction of the leakage term gives rise to an increase in the MSE. The increase in magnitude is given by

$$\epsilon - \epsilon_{min} = [\underline{C}_{opt} - \underline{C}(\infty)]'\underline{A}[\underline{C}_{opt} - \underline{C}(\infty)]. \qquad (5.9)$$

Eq.(5.8) can be rewritten as

$$\underline{C}_{opt} - \underline{C}(\infty) \doteq \mu\underline{A}^{-1}\underline{C}_{opt} \;,$$

substitute this into Eq.(5.9)

$$\epsilon - \epsilon_{min} \doteq \mu^2\underline{C}'_{opt}\,\underline{A}^{-1}\,\underline{C}_{opt} \;. \qquad (5.10)$$

Now the increased MSE grows only as the square of the leakage term, μ, while the eigenvalue distribution and the range that the taps can wander can be favorably altered, by employing a small value of μ.

5.2 High Frequency Compensation

In 1979, Atal and Schroeder [11] wrote a paper to discuss methods for reducing the subjective distortion in linear predictive coders for speech signals. While they were discussing about prediction based on spectral envelope, they found similar problem of having nonunique solution of the predictive coefficients and thus causing high power gain to the predictor. The reason was that the convariance matrix of the speech signal had missing components rejected by a low-pass filter. The missing high frequency components produced some low eigenvalues of the matrix and thus made the matrix nearly singular. In order to avoid this ill-conditioning of the matrix, they added another matrix proportional to the convariance matrix of high-pass filtered white noise, i.e.

$$\hat{\phi}_{ij} = \phi_{ij} + \lambda \epsilon_{min} \rho_{i-j} , \qquad (5.11)$$

where λ is a small constant (in the range 0.01-0.1), ϵ_{min} is the minimum value of the mean-square prediction error, ρ_i is the autocorrelation of the high-pass filtered white noise. Ideally the high frequency compensation should be the filter complementary to the low-pass filter used in the sampling process.

This paper is actually very similar to the tap-leakage algorithm. In fact if the additional matrix is just a matrix of white noise, the method used by Atal et al is just the same as the tap-leakage algorithm. This can be shown by taking the new autocorrelation matrix as

$$\underline{\tilde{A}} = \underline{A} + \mu\underline{I} \tag{5.12}$$

where μ is a noise variance and it resembles the product $\lambda\epsilon_{min}$. The identity matrix represents the autocorrelation matrix corresponding to white noise. Now concerning about the equation

$$\underline{C}(T+1) = \underline{C}(T) + \beta E[e(T)\underline{Y}(T)],$$

the stochastic component can be written as

$$\begin{aligned} E[e(T)\underline{Y}(T)] &= E[\{d(T) - \underline{C}'(T)\underline{Y}(T)\}\underline{Y}(T)] \\ &= \underline{P} - \underline{AC}(T). \end{aligned} \tag{5.13}$$

Now instead of taking the right hand side of (5.13) as the stochastic component, this expression is replaced by using a new autocorrelation matrix $\underline{\tilde{A}}$, thus

$$\underline{P} - \underline{\tilde{A}}\underline{C}(T) = \underline{P} - [\underline{A} + \mu\underline{I}]\underline{C}(T)$$

$$= \underline{P} - \underline{AC}(T) - \mu\underline{IC}(T)$$

$$= E[e(T)\underline{Y}(T)] - \mu\underline{IC}(T). \quad (5.14)$$

Then the new stochastic component becomes $E[e(T)\underline{Y}(T)] - \mu\underline{IC}(T)$ and in this way, the tap coefficients are updated under a new equation,

$$\underline{C}(T+1) = \underline{C}(T) + \beta\{E[e(T)\underline{Y}(T)] - \mu\underline{IC}(T)\}$$

$$= (1 - \beta\mu)\underline{C}(T) + \beta E[e(T)\underline{Y}(T)]. \quad (5.15)$$

As a result, Eq.(5.15) is the tap leakage formula with leakage term μ.

So far it has been proved that the method used by Atal et al is very similar to the tap-leakage algorithm and they are the same in fact if the additional matrix is just a matrix of white noise. The remaining work is to determine under which situations should an additional matrix of white noise or high frequency compensation be used. Two situations are mentioned as follows :-

Case (1) Constraining the magnitude of the tap vector

Concerning about this objective, the cost function is given already in the previous section, i.e.

$$J = E[e^2(T)] + \mu\, \underline{C}'(T)\underline{C}(T).$$

Now let us minimize J in a deterministic manner by using the true gradient, then μ can be interpreted as a Lagrange multiplier. Hence, taking the derivative of the cost function with respect to $\underline{C}(T)$ and set the expression to zero. It becomes

$$2E[\underline{Y}(T)\underline{Y}'(T)]\underline{C}(T) - 2E[d(T)\underline{Y}(T)] + 2\mu\underline{C}(T) = 0. \quad (5.16)$$

Simplifying Eq.(5.16),

$$2\underline{AC}(T) - 2\underline{P} + 2\mu\underline{C}(T) = 0$$

and hence

$$(\underline{A} + \mu\underline{I})\underline{C}(T) = \underline{P}. \quad (5.17)$$

Eq.(5.17) has shown that an additional matrix of white noise with diagonal elements μ should be used in this situation. μ is chosen to set $\underline{C}'(T)\underline{C}(T)$ to any desired value.

Case (2) To minimize the eigenvalue spread

Recall that in the previous case, the matrix $\underline{A}$ has been perturbed to $\underline{A} + \mu\underline{I}$. The corresponding eigenvalue spread has become $(\lambda_{max} + \mu) / (\lambda_{min} + \mu)$. Though this has improved the eigenvalue spread from the original $\lambda_{max} / \lambda_{min}$, it could be further improved to $\lambda_{max} / (\lambda_{min} + \mu)$. While the leakage term is just used to prevent the vanishing small λ_{min} in the denominator, it does not increase the λ_{max} in the numerator. This can be achieved by using an additional matrix which is proportional to the complementary matrix $\underline{A}$ (i.e. for a low-pass filtered white noise matrix, the additional matrix would then be a high frequnecy compensation) Recall in the previous section, with a fixed leakage μ, the long-term tap coefficient is given by Eq.(5.5), i.e.

$$\underline{C}(\infty) = [\text{diag}(\mu) + \underline{A}]^{-1} \underline{A}\underline{C}_{opt} \quad .$$

If $\underline{A}$ is being diagonalized as

$$\underline{A} = \underline{M}\text{diag}(\lambda_i)\underline{M}' \quad ,$$

then $\underline{C}(\infty)$ can be written as

$$\underline{C}(\infty) = \underline{M}\text{diag}\left(\frac{\lambda_i}{\lambda_i + \mu}\right)\underline{M}'\underline{C}_{opt} \quad . \tag{5.18}$$

The increased magnitude in the MSE would be

$$\epsilon - \epsilon_{min} = [\underline{C}_{opt} - \underline{C}(\infty)]'\underline{A}[\underline{C}_{opt} - \underline{C}(\infty)]. \tag{5.19}$$

Hence μ should be chosen to make $\underline{C}(\infty)$ as close to $\underline{C}_{opt}$ as possible in order to minimize the increase in the MSE in Eq.(5.19) and μ is used in Eq.(5.18) to prevent small eigenvalues in $\underline{A}$ that will cause the problem of singularity. Now instead of using just a diagonal matrix with all elements μ, a matrix $\underline{N}$ is used, i.e.

$$\tilde{\hat{\underline{A}}} = \underline{A} + k\underline{N}$$

then the tap coeficients are updated under a new equation,

$$\underline{C}(T+1) = (1 - \beta k\underline{N})\underline{C}(T) + \beta E[e(T)\underline{Y}(T)]$$

where k is a suitably chosen constant, and

$$\underline{C}(\infty) = [k\underline{N} + \underline{A}]^{-1}\underline{A}\underline{C}_{opt} \quad . \tag{5.20}$$

Assume $\underline{N}$ can be diagonalized as

$$\underline{N} = \underline{M}\mathrm{diag}(\rho_i)\underline{M}' \quad . \tag{5.21}$$

This assumption can be valid when the order of the matrix $\underline{A}$ is high, the matrix $\underline{A}$ which is Toeplitz would then be asymptotically equivalent to a circulant matrix, and all circulant matrices have the same set of eigenvectors [Gray 25]. In this way $\underline{M}$ can be known beforehand and ρ_i are just arbitrary constants to be determined. Now with $\underline{N}$ being diagonalized as in Eq.(5.21), then

$$\underline{C}(\infty) = \underline{M}\mathrm{diag}\left(\frac{\lambda_i}{\lambda_i + \rho_i}\right)\underline{M}'\underline{C}_{opt} \quad . \tag{5.22}$$

If ρ_i are chosen to be zero when λ_i are large and ρ_i are chosen to be a small positive number μ when λ_i are small, then $\underline{C}(\infty)$ is closer to $\underline{C}_{opt}$ and the increase in the MSE in Eq.(5.19) is smaller. In this case, the perturbed matrix $\underline{A}$ would then have the largest eigenvalue λ_{max} and the smallest eigenvalue $(\lambda_{min} + \mu)$ and the additional matrix $\underline{N}$ is also complementary to the autocorrelation matrix $\underline{A}$.

CHAPTER 6

APPLICATION OF LEAST-SQUARES LATTICE ALGORITHM TO ECHO CANCELLER

6.1 Idea of Least-Squares Approach

Although the computational simplicity of the SG algorithm make it the predominant adaptation algorithm in the echo cancellers, it has paid the price of being slow in converging to the optimum solution, especially when its input matrix have a large eigenvalue spread. The situation is even worse when some eigenvalues are vanishing small as mentioned before. The slow convergence is mainly due to the fact that the gradient algorithm has only a single adjustable parameter for controlling the convergence rate, i.e., the step size β. In order to obtain faster convergence, it is necessary to design some complex algorithms which involve more adjustable parameters. In particular, if the matrix A is (n×n) and has n eigenvalues, it is much better to use an algorithm that contains n parameters, one for each of the eigenvalues. The optimum selection of these parameters to obtain faster convergence is the main concen of this chapter.

In deriving faster convegence, a recursive least-squares (RLS) approach is adopted. This appoach is called least-squares since it minimizes the sum of squares of an error signal rather than the mean-square error as in the MMSE algorithm. That is, the performance index in this algorithm is expressed in terms of a time average instead of an ensemble average. This ovecomes the problem of not knowing the statistics of the input signals. In addition, in this approach, the computation of new coefficient estimate utilizes all the past information available as compared with the SG algorithm which utilizes only the instantaneous sample value. Because the LS approach makes better use of the past information, its start up is expected to be faster [Mueller, 26]

The time-average weighted square error to be minimized will be

$$\epsilon(T) = \sum_{j=0}^{T} w_{T-j} \left[d(j) - \underline{C}'(T)\underline{Y}(T) \right]^2 \tag{6.1}$$

where w_l is a casual window function. This window function is intended to ensure data far in the past receive a lot less attention than recent input samples so that the filter can be operated in a nonstationary environment. One such form of window that is commonly used in practice is the exponential window defined by

$$w_l = \gamma^l , \qquad l \geq 0$$

where γ is a positive constant less than unity.

Now define a time-average autocorrelation matrix at sample T as

$$\underline{B}(T) = \sum_{j=0}^{T} w_{T-j} \underline{Y}(j)\underline{Y}'(j). \qquad (6.2)$$

Notice that this matrix is now non-Toeplitz. Define a cross correlation vector estimate as

$$\underline{Q}(T) = \sum_{j=0}^{T} w_{T-j} d(j)\underline{Y}(j). \qquad (6.3)$$

In terms of these estimates, it is easy to show that Eq.(6.1) can be minimized (eg. differentiating (6.1) with respect to $\underline{C}(T)$ and set it to zero) by the choice

$$\underline{C}(T) = \underline{B}^{-1}(T)\underline{Q}(T) \qquad (6.4)$$

As long as the window function can be written recursively, both the autocorrelation and crosscorrelation estimates can be written respectively as

$$\underline{B}(T) = \gamma\underline{B}(T-1) + \underline{Y}(T)\underline{Y}'(T) \tag{6.5}$$

$$\underline{Q}(T) = \gamma\underline{Q}(T-1) + d(T)\underline{Y}(T). \tag{6.6}$$

Consider the error at time T being defined as

$$e(T) = d(T) - \underline{C}'(T-1)\underline{Y}(T). \tag{6.7}$$

Substituting Eq.(6.4) into Eq.(6.6)

$$\begin{aligned} \underline{Q}(T) &= \underline{B}(T)\underline{C}(T) \\ &= \gamma\underline{B}(T-1)\underline{C}(T-1) + d(T)\underline{Y}(T). \end{aligned} \tag{6.8}$$

With Eq.(6.5), Eq.(6.8) becomes

$$\begin{aligned} \underline{B}(T)\underline{C}(T) = \{&\underline{B}(T) - \underline{Y}(T)\underline{Y}'(T)\}\underline{C}(T-1) \\ &+ d(T)\underline{Y}(T) \end{aligned} \tag{6.9}$$

Eq.(6.9) can be further simplified with the help of Eq.(6.7) as

$$\underline{C}(T) = \underline{C}(T-1) + \underline{B}^{-1}(T)e(T)\underline{Y}(T) \tag{6.10}$$

Eq.(6.10) is almost identical to the coefficient updating formula of the stochastic algorithm except that the constant scalar step size β is replaced by the inverse of the autocorrelation matrix estimate. This modification is the main source of the improved convergence performance of the RLS algorithm. Since the inverse of the correlation matrix has the effect of decorrelating the successive tap inputs, it makes the algorithm self-orthogonalizing. Because of this property, the algorithm is essentially independent of the eigenvalue spread of the input matrix. This algorithm is known as the Kalman/Godard algorithm. Despite its superior in tracking performance, the algorithm has two disadvantages. One of the disadvantages is its computational complexity, as $O(n^2)$ operations are necessary to compute the inverse of $\underline{B}(T)$. The second is its sensitivity to roundoff noise that accumulates due to the recursive computations. The latter may cause instabilities in the algorithm. One way to improve the computational efficiency is by using the fast Kalman algorithm [Falconer and Ljung 27].

6.2 Least-Squares Lattice Filter

Another alternative to solve the problems of computational complexity and instability is by applying the RLS algorithm on the lattice filter structure. The lattice filter is order-recursive and as a consequence, the number of sections it contains can be readily increased or decreased without affecting the parameters of the other sections. In contrast, the coefficients of a transversal filter obtained on the basis of the MSE or LS criterion are interdependent. Thus the Kalman algorithm discussed before is recursive but not recursive in order. However, the least-squares lattice algorithm is recursive both in order and in time.

Notice in Fig.6.1a that the lattice filter can be partitioned into two stages : stage 1 produces a set of uncorrelated backward prediction error $b_m(T)$, which are then normalized by their respective power and correlated with $e_m(T)$ in stage 2. The forward and backward errors $f_m(T)$ and $b_m(T)$ are usually called the residuals. The decoupling among stages of the lattice is due to the orthogonality properties of the residuals [Proakis 18].

The mathematical derivation of the LS lattice algorithm is quite lengthy. A readable account of the derivation is given by Satorious [28], Mueller [26] and Schicor [29].

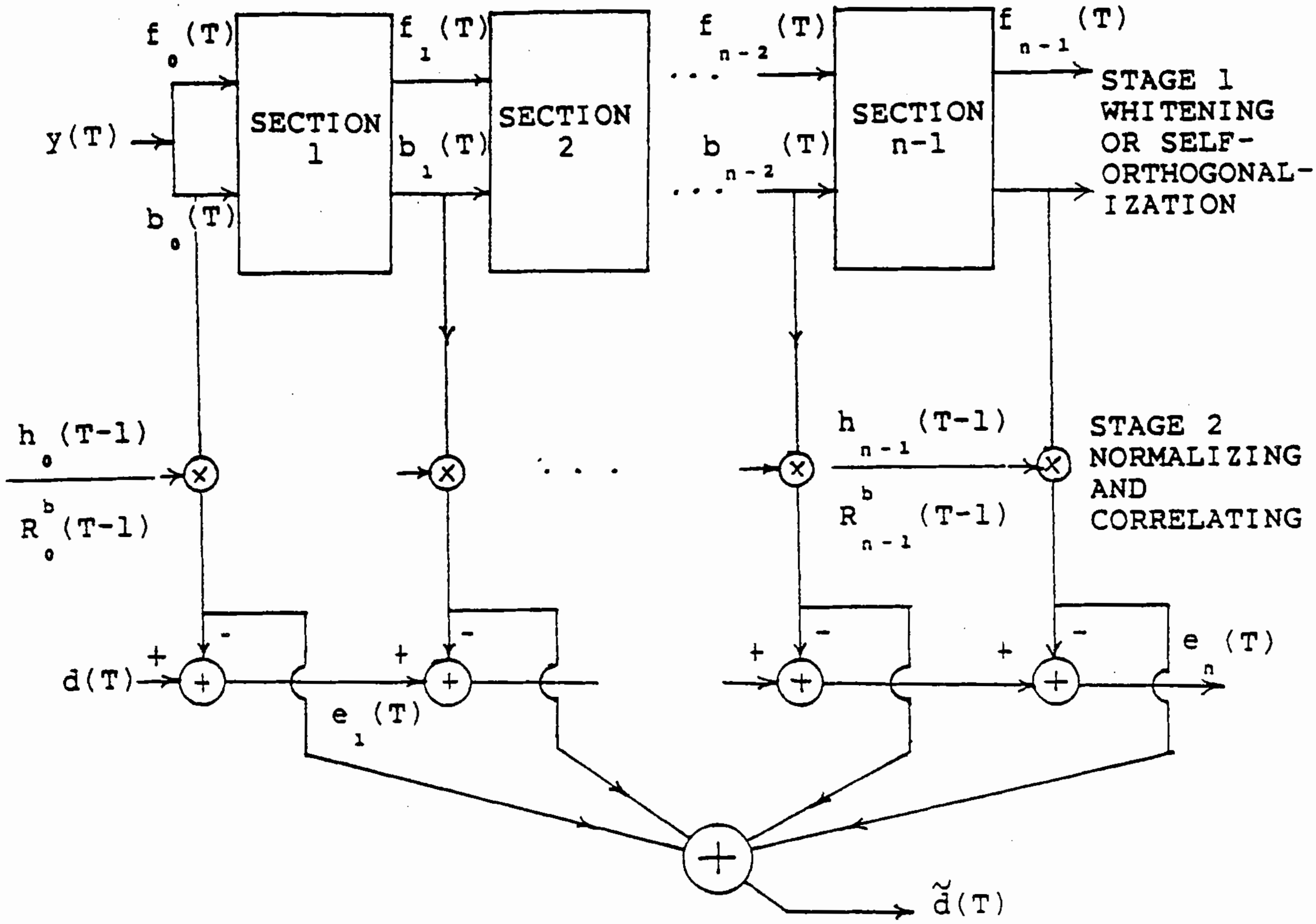


Fig. 6.1a. Least-Squares lattice filter

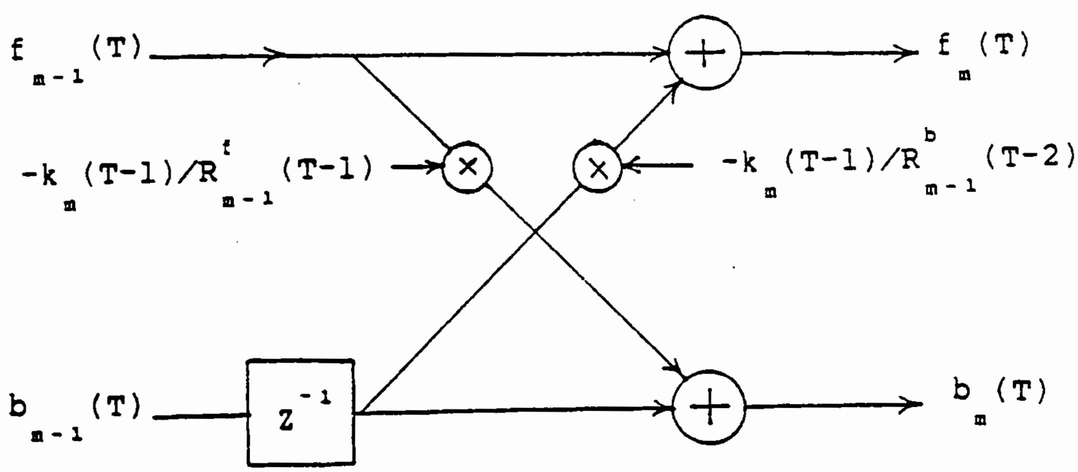


Fig. 6.1b. Detail description of the self-orthogonalization section.

A summary of the LS lattice algorithm is given below.

(1) Start with the initial condition

$$f_m(-1) = b_m(-1) = k_m(-1) = h_m(-1) = 0 \qquad m = 0,..,n-1$$

$$\alpha_m(-1) = \alpha_{-1}(-1) = 1$$

$$R^f_m(-1) = R^b_m(-1) = R^b_m(-2) = \epsilon > 0,$$

where ϵ is an a priori estimate of the prediction-error variance.

(2) For each instant of time, T = 1, 2,...., compute the various zeroth-order (m = 0) variables

$$f_0(T) = b_0(T) = Y(T)$$

$$R^f_0(T) = R^b_0(T) = \gamma R^f_0(T-1) + y^2(T)$$

$$\alpha_{-1}(T) = 1$$

$$e_0(T) = d(T).$$

(3) Compute the various order updates in the following sequence

$$k_{m+1}(T) = \gamma k_{m+1}(T-1) + \alpha_m(T-1) f_m(T) b_m(T-1) \qquad m = 0,.,n-2$$

$$f_{m+1}(T) = f_m(T) - k_{m+1}(T-1) b_m(T-1)/R^b_m(T-2)$$

$$b_{m+1}(T) = b_m(T-1) - k_{m+1}(T-1) f_m(T)/R^f_m(T-1)$$

$$R^f_{m+1}(T) = R^f_m(T) - k^2_{m+1}(T)/R^b_m(T-1)$$

$$R^b_{m+1}(T) = R^b_m(T-1) - k^2_{m+1}(T)/R^f_m(T)$$

$$\alpha_{m+1}(T) = \alpha_m(T) - \alpha^2_m(T)\ b^2_m(T)/R^b_m(T)$$

$$h_m(T) = \gamma h_m(T-1) + \alpha_m(T) b_m(T) e_m(T) \qquad m = 0,...,n-1$$

$$e_{m+1}(T) = e_m(T) - h_m(T-1) b_m(T)/R^b_m(T-1).$$

CHAPTER 7

COMPUTER SIMULATIONS

The operation of the SG, tap-leakage and LS lattice algorithms described in the previous chapters were simulated on a digital computer. Fig.7.1 represents the system model used. The equivalent discrete-time channel characteristic used to generate the convergence results was selected from [Proakis 18] with coefficients {0.304, 0.903, 0.304}. The channel had an eigenvalue spread (ratio of largest-to-smallest eigenvalues of channel correlation matrix) of 21. A small amount of uncorrelated Gaussian noise (noise variance = 0.001, the SNR was equivalent to 30 dB) was added to the output of the channel. Two different input bipolar sequences y(T) were used, namely the highly correlated sequence (eg. +1 -1 +1 -1 +1 -1 ...) and random sequence of +1 and -1 to represent the periodic and white input data respectively. In all simulations, 11-tap filters were used for both the transversal and the lattice filter structures. The tap values are set to zero initially.

For the stochastic gradient and the tap-leakage algorithms, the tap coefficients were updated with the following parameters. The step size, β was chosen to be 0.02 in the transversal filter as this choice of β resulted in approximately the same MSE as that of the lattice filter. In the tap-leakage algorithm, μ was chosen to be 0.1. This is reasonable, as the suitable values of the leakage terms suggested by Atal [11] were in the range of 0.01 - 0.1 in order to limit the degradation in the MSE.

In chapter 5, another tap-leakage algorithm has been proposed to minimize the eigenvalue spread. A high frequency compensation of the autocorrelation matrix $\underline{A}$ was used. Concerning this high frequency compensation, it should ideally be the filter complementary to the autocorrelation matrix $\underline{A}$. However, satisfactory results can be obtained with the high-pass filter $[0.5(1 - 1/z)]^2$. For this filter, the autocorrelatons are $\mu_0 = 3/8$, $\mu_1 = -1/4$, $\mu_2 = 1/16$ and $\mu_i = 0$ for $i > 2$. The constant k in the tap updated equation could be chosen in order to match the MSE for the tap-leakage algorithm with μ equal to 0.1.

In the lattice filter simulation, ϵ was set to 0.001 (as additional simulations revealed that the start up performance of the lattice filter is highly insensitive to the choice of ϵ), and γ was set close to one as the channel used is time invariant. In the transversal and lattice filter simulations, the minimum MSE was obtained by setting the optimal coefficient vector to be identical to the impulse response of the time-invariant channel.

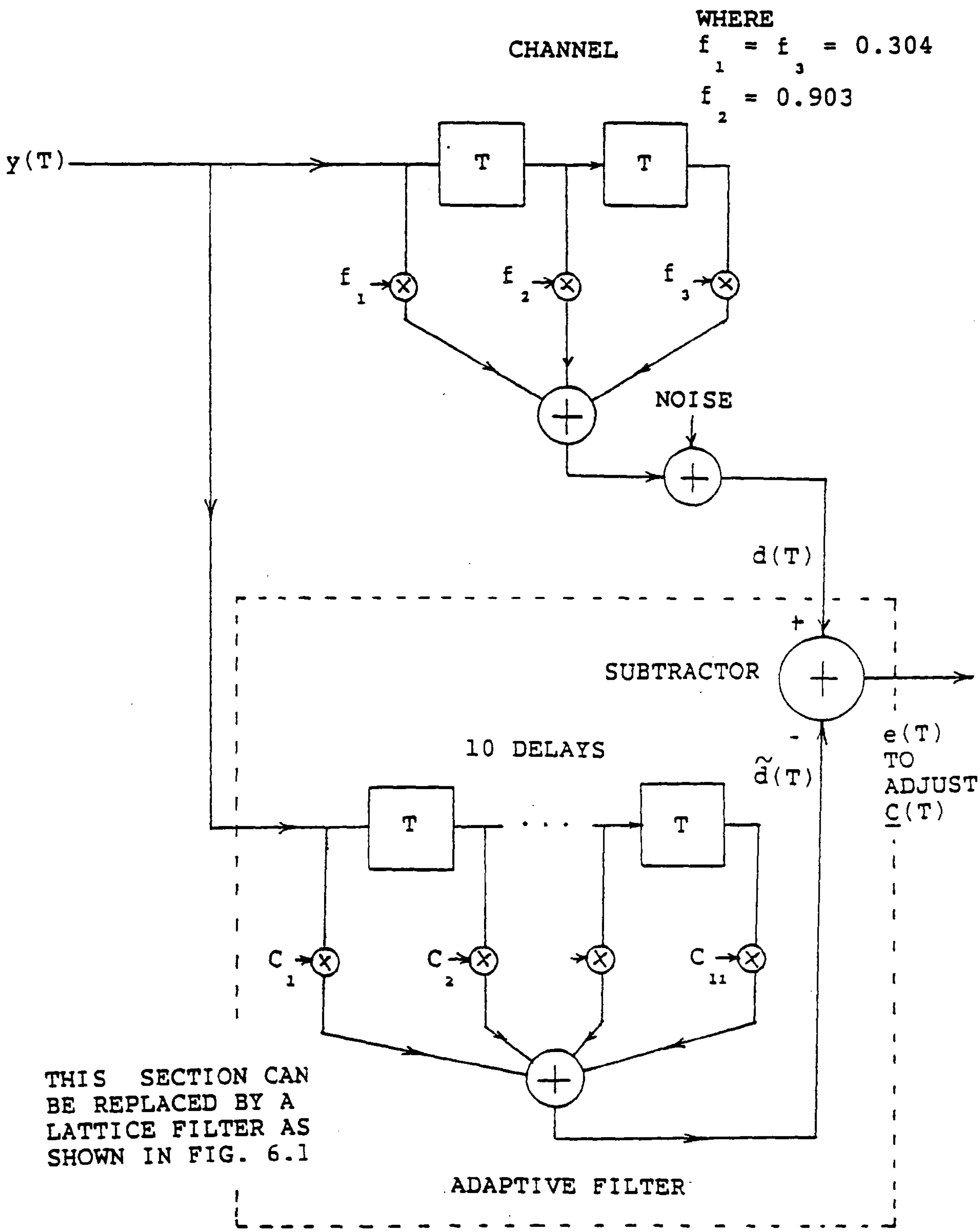


Fig. 7.1. System model for computer simulations.

Results

(1) Convergence properties of the transversal filter

Fig.7.2 shows the learning curve for the 11-tap transversal filter. The first half of the input sequence consists of periodic data and the second half consists of random data. It can be observed that there is a sudden increase in the MSE during the transition period. This has justified the symptoms predicted in chapter 4 as a result of tap drifting effect. Note also that the rates of convergence are not the same for these two input signals. It takes more iterations for the filter to converge to near the optimum MSE for the periodic input. This shows that the rate of convergence of the SG algorithm is dependent on the eigenvalue spread of the input. The periodic data has higher eigenvalue ratio and thus causes this slow convergent rate. Simulation results in Fig.7.3 shows the amplitudes of the tap coefficient after running the filter under both periodic and white inputs for 200 iterations. The tap coefficients that have been trained with white data are very close to the optimum solutions (the impulse response of the channel). While the alternative set of tap coefficient vector is a poor match to the unit impulse response of the echo channel.

It remains to examine if there is an increase in the MSE when the input data is changed from white to periodic. This can be tested by using an alternate input sequence of periodic and white data, and the result is illustrated in Fig.7.4. This figure shows that there are only increases in the MSE in the first and third transitions (while the input sequence changes from periodic to white) but there is no increase in the second transition (when the input sequence changes from white to periodic) This shows that there is no drifting effect away from the optimum tap coefficients when the data covers the full spectrum.

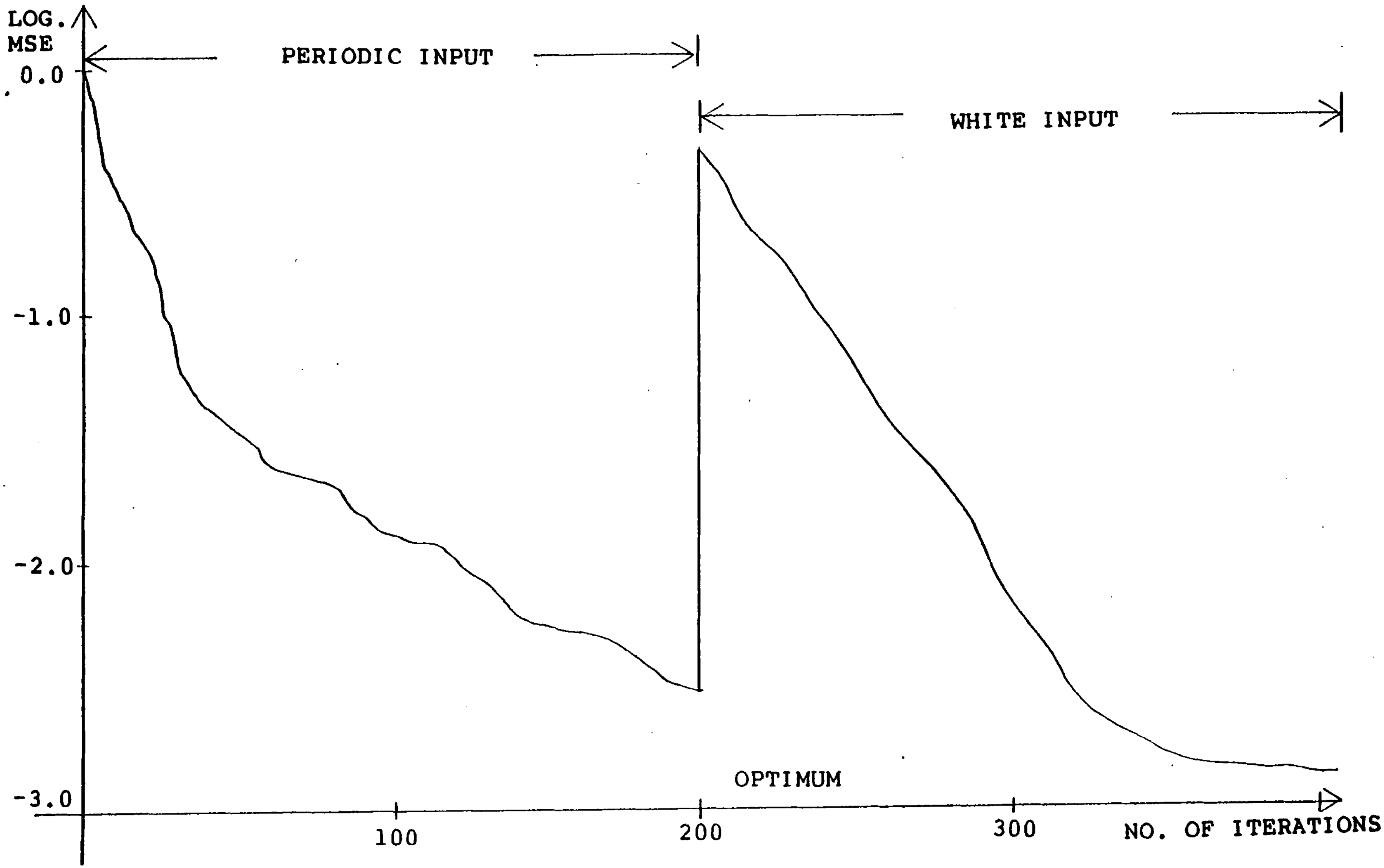


Fig. 7.2 Convergence properties of the stochastic gradient algorithm

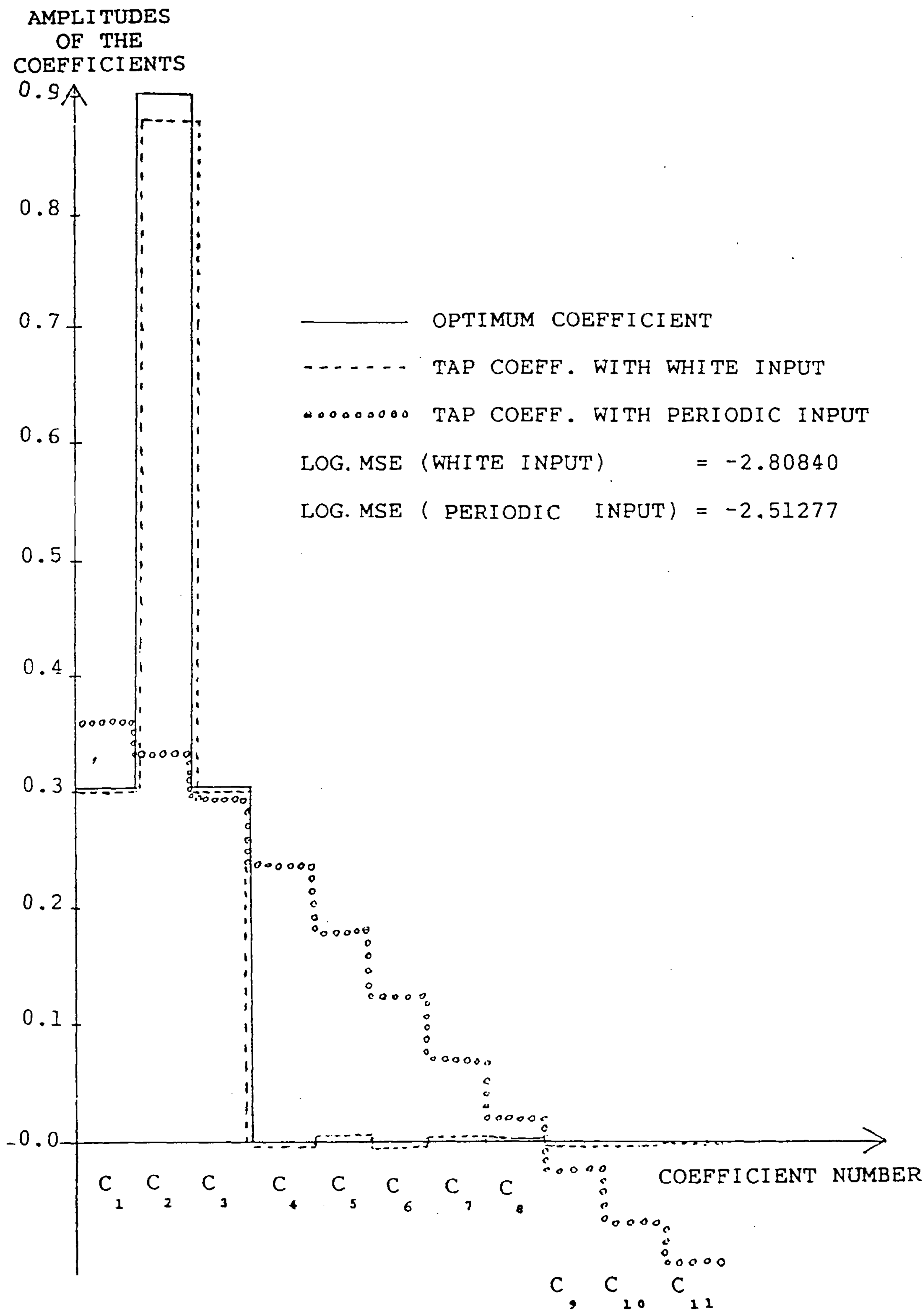


Fig. 7.3. Amplitudes of the coefficients after running the input data for 200 iterations.

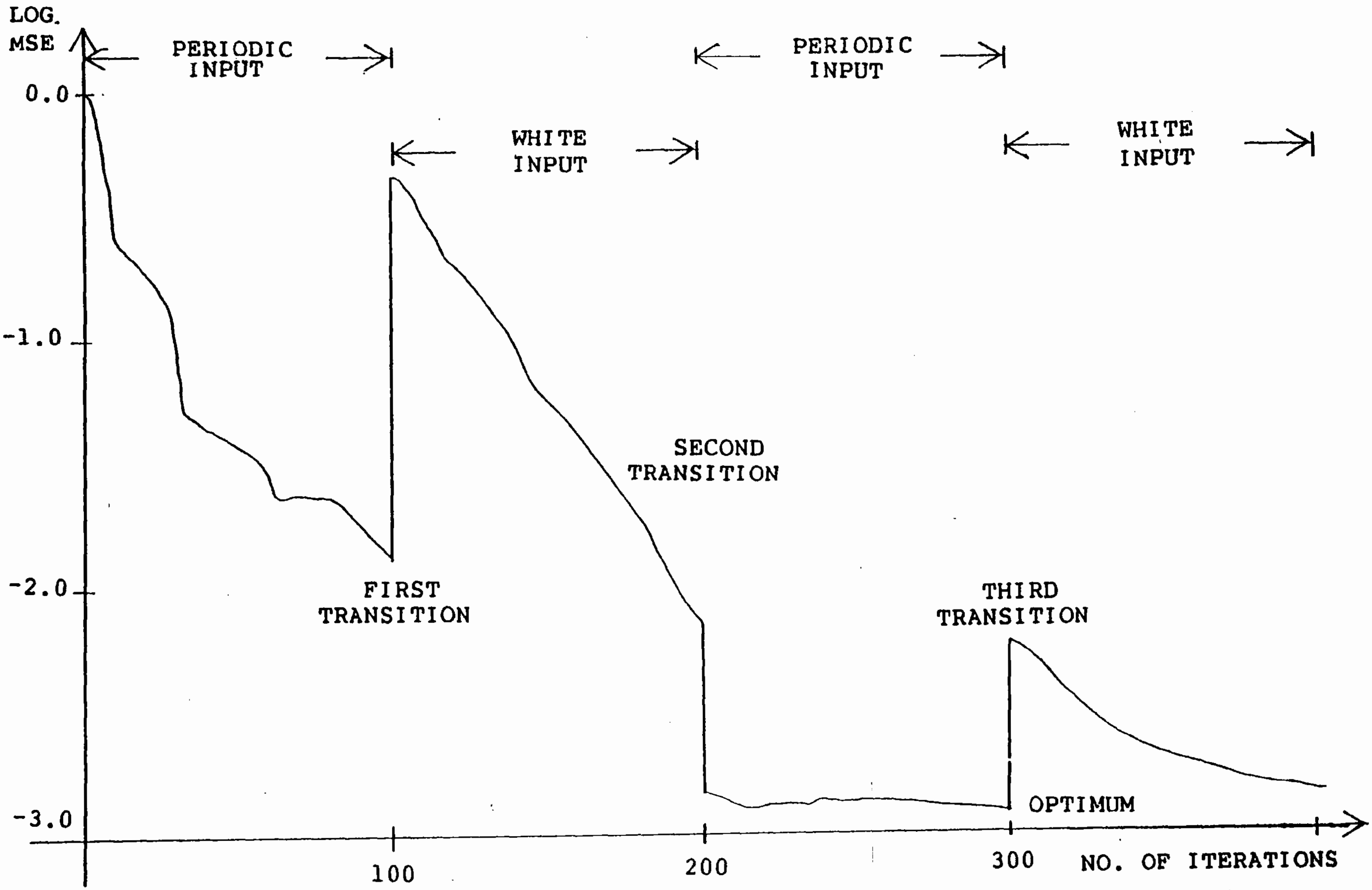


Fig. 7.4. Convergence properties of the stochastic gradient algorithm with alternate input sequence.

(2) Adjustment properties of the tap-leakage algorithm

Fig.7.5 depicts the magnitudes of the tap coefficients after running the transversal filter under the periodic input for 200 iterations. Two cases, with or without tap-leakage algorithm, can both be observed in this figure. The magnitude of each tap and the sum of squares of the tap coefficients have been reduced with the tap-leakage adjustment algorithm. However, there is an increase in the MSE as a result.

	μ = 0.0	with fixed leakage μ = 0.1	with high freq. compensation k = 4.667
$\underline{C}'\underline{C}$	0.4603474	0.375460	0.417334
MSE	0.0030706	0.014774	0.014774

Table 7.1 Comparison of the tap-leakage algorithms

Table 7.1 gives the comparison of the two tap-leakage algorithms (the one with fixed leakage term and the one with high frequency compensation). It shows that when k is adjusted to set the MSE equal in these tap-leakage algorithms, the sum of squares of the coefficients are reduced in both cases and the fixed leakage term has reduced the sum to the minimum value for a given excess MSE as predicted in Chapter 5. Moreover, the eigenvalue spreads are compared numerically between $\underline{A}$, $\underline{A} + \mu\underline{I}$ and $\underline{A} + k\underline{N}$, with μ equal to 0.1 and k equal to 4.667. Simulations results show that the eigenvalue spread of $\underline{A}$ is 21.73, the fixed leagkage term has reduced the spread to 11.49 and the high frequency compensation matrix has further reduced the spread to 2.57. This has confirmed that the additional high frequency compensation can help to minimize the eigenvalue spread of the autocorrelation matrix.

The tap-leakage algorithm has been proved to be superior in preventing the tap coefficients from drifting to very large values. However, when this algorithm is used again for the alternate input sequence of periodic and white data, Fig.7.6 shows that the increase in the MSE in the third transition is even higher. The reason is that as drifting has caused errors in the coefficients, the tap coefficients will have a chance to be either too large or too small in value. However, in this example, the random fluctuation has brought a decrease in the magnitude instead of an increase. Hence, the tap-leakage algorithm, which is originally designed to force the coefficients further towards the origin has thus caused a further increase in the MSE.

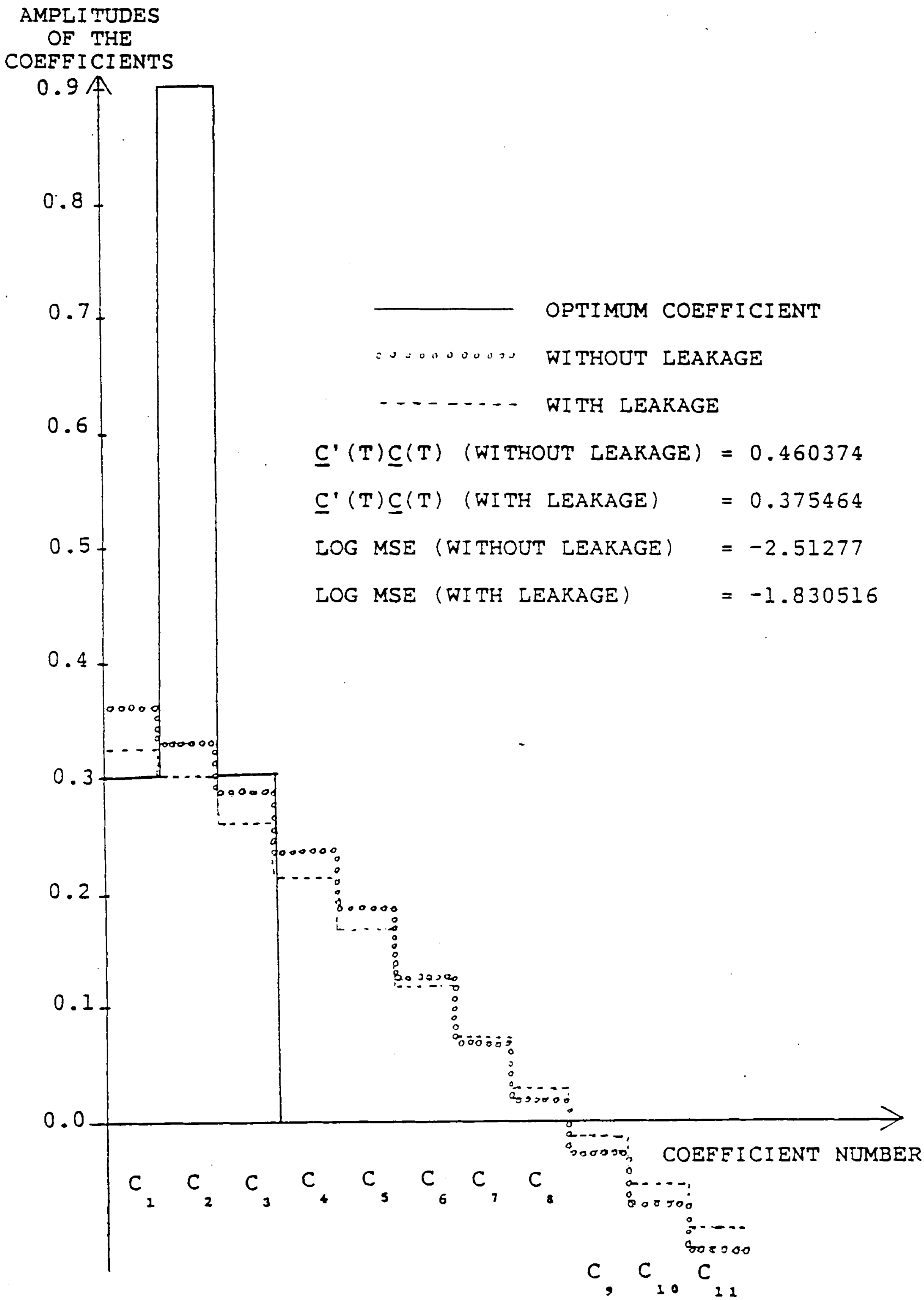


Fig. 7.5. Amplitudes of the coefficients when a fixed leakage term is added.

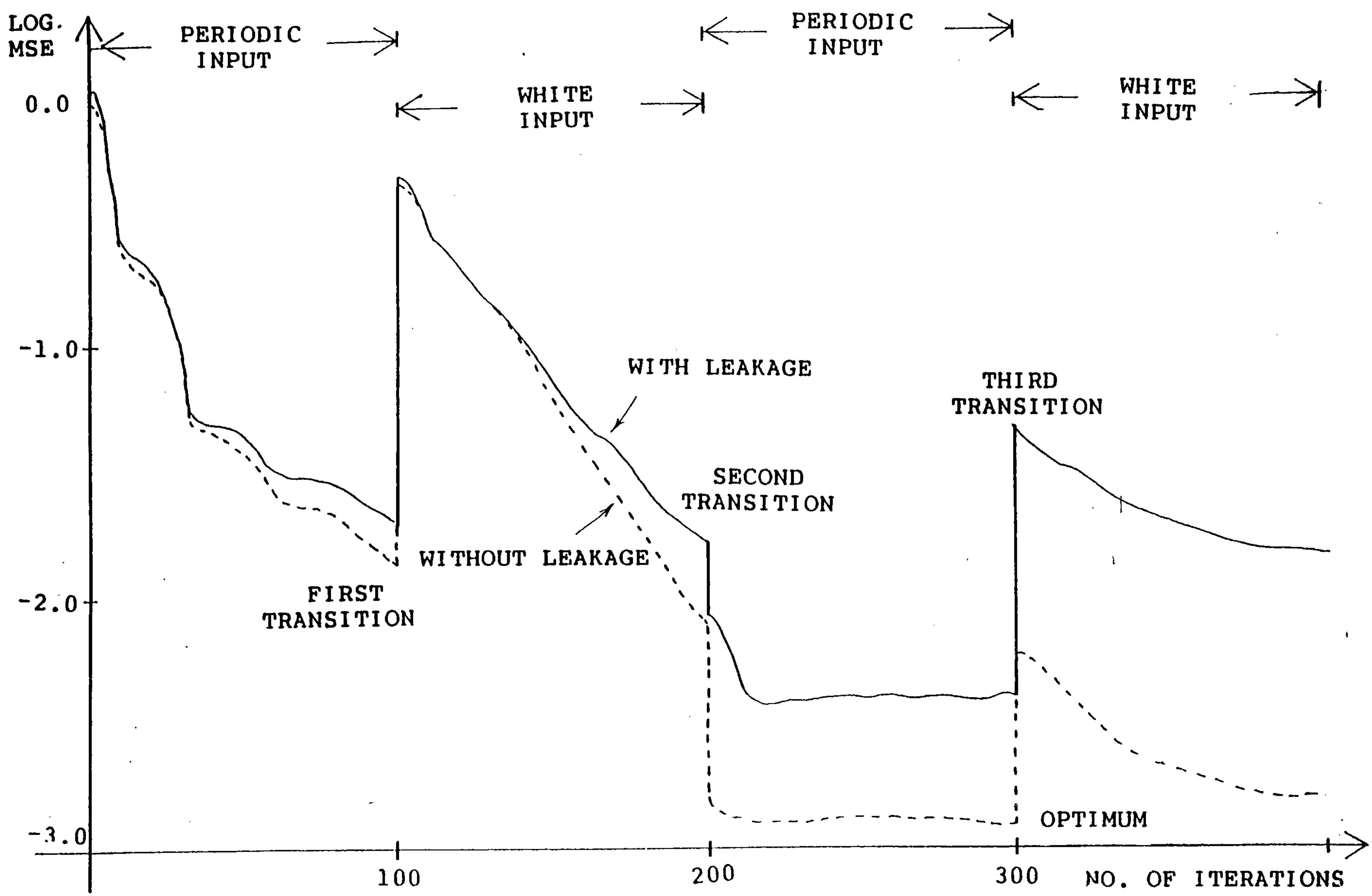


Fig. 7.6. Convergence properties of the SG and tap-leakage algorithms.

(3) Convergence properties of the LS lattice algorithm

Fig.7.7 reveals the convergence properties of the least-squares lattice filter when the first half of the input sequence is periodic and the second half is random data. The lattice filter is much superior in the rate of convergence as compared with the transversal filter. The result of the simulation shows that the lattice filter converges in about 10 to 20 iterations, independent of the eigenvalue spread. In addition, the learning curve exhibits a similar increase in the MSE during the transition. However, due to this fast convergent rate, the filter can bring the MSE down to the optimum again in a short time after the transition.

In Fig.7.8, an alternate input sequence of periodic and white data is used again to show that there is no increase in the MSE when the input energy changes from broadband to narrowband. The result of this simulation has justified this because there are only increases in the first and third transitions but none in the second one.

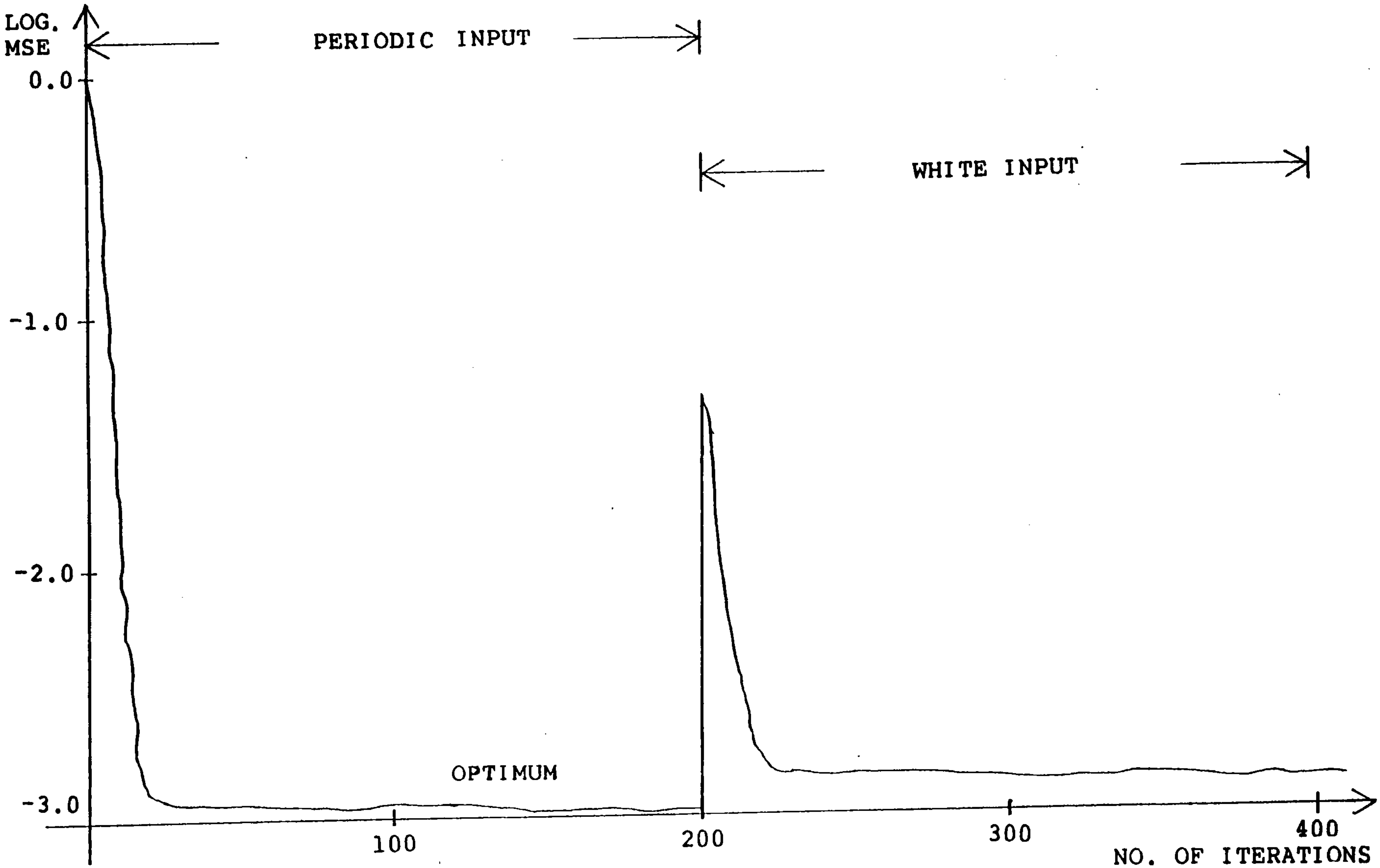


Fig. 7.7. Convergence properties of the least-squares lattice algorithm.

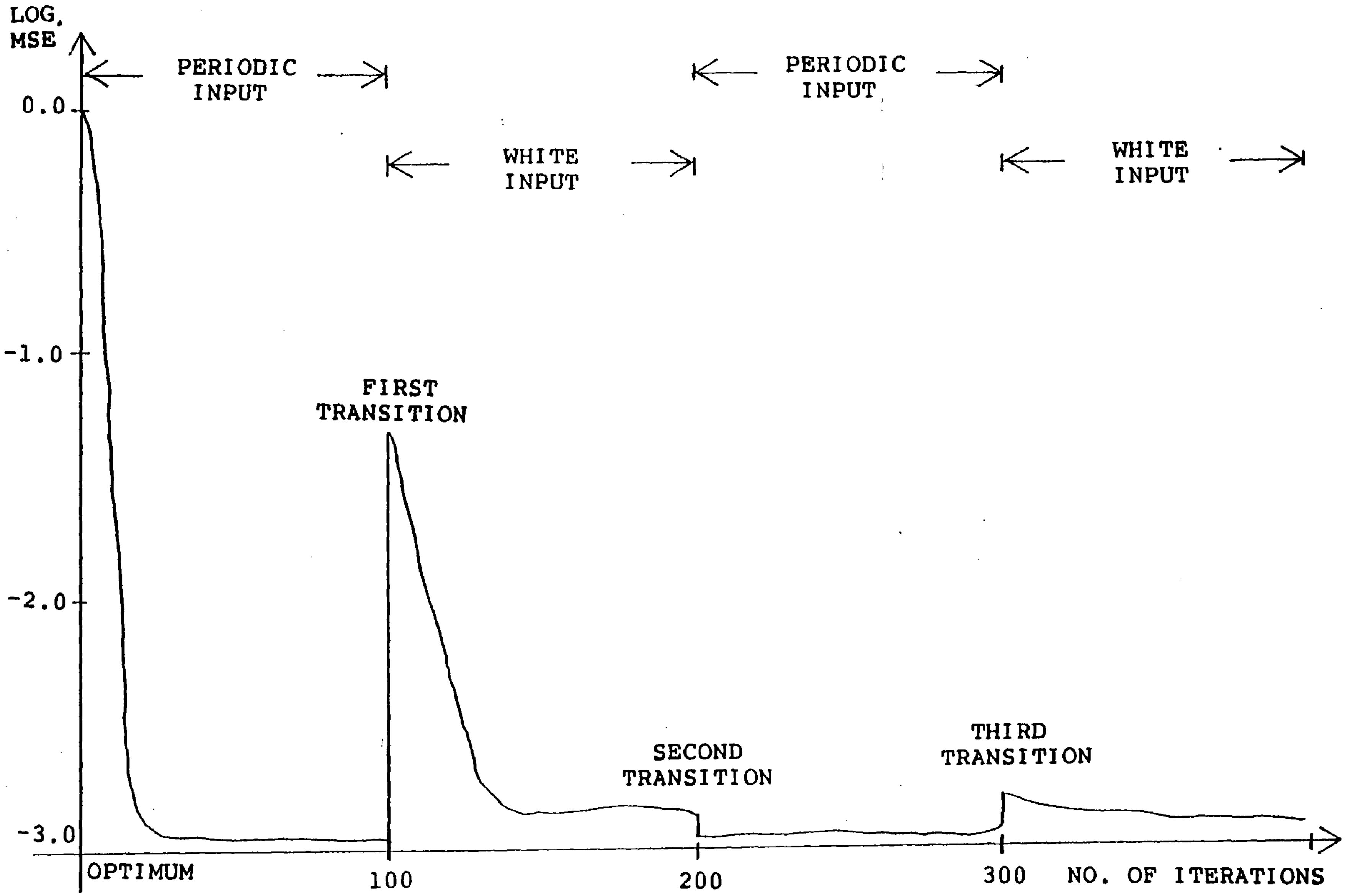


Fig. 7.8. Convegence properties of the LS lattice algorithm with alternate input sequence.

CHAPTER 8

CONCLUSIONS

This study has presented various adaptation algorithms for adapting the echo canceller constructed from both transversal and lattice filter structures. The stochastic gradient algorithm constructed from transversal filter is the main concern here. The SG algorithm is based on gradient search techniques using a stochastic approximation to the MMSE gradient, whereas the MMSE criterion applies only to the stationary case and the SG algorithm can be applied in more realistic nonstationary environment. The existence of a unique solution to the MMSE problem and the convergence of the SG algorithm depended on the nonsingularity of the input autocorrelation matrix. When the autocorerelation matrix is singular, that is, when the input signal is just a narrowband signal, two problems will occur. First, the tap coefficients can assume any values other than the optimum coefficients, and the errors in the coefficients will grow with time. When the input is changed back to wideband data, the errors in the coefficients (either being too large or too small) will cause the echo canceller an increase in the MSE suddenly. Sometimes the magnitudes of the coefficients become too large that one or more registers will saturate. Second, the matrix will have high eigenvalue ratio that will slow down the rate of convergence of the SG algorithm.

In order to prevent the saturation in the first problem, a tap-leakage algorithm is applied onto the tap coefficients. Leakage can be applied to the coefficients in two ways. When a small constant leakage term is added to every diagonal elements of the autocorrelation matrix, the sum of squares of the coefficients will be decreased. When the input autocorrelation matrix is slightly increased by adding a matrix complemenatry to the input matrix, the sum of squares of the coefficients will be decreased (though the sum will not be as small as the previous method) and the eigenvalue spread will be decreased more compared with the first method. In both methods, the bias introduced on the coefficeints leads to an increase of the error signal power, and thus in order to maintain an acceptable level of performance, the leakage term should be kept small. It has been found that even though the tap-leakage algorithm can help to prevent the tap coefficients from saturation, it cannot help to prevent the sudden increase in the MSE when the input energy has changed from narrowband to wideband. This is because the drifting might causes a decrease of magnitudes of the coefficients from the optimum $\underline{C}_{opt}$. The leakage term introduced has furthered the decrease of the magnitudes and thus caused a larger increase in the MSE.

The rate of convergence problem occured in the SG algorithm can be solved by adopting the LS algorithm. The LS algorithm minimizes a time average criterion instead of ensemble average criterion as in the SG algorithm. The LS algorithm has been proved that it can outperform the SG algorithm in convergence as the LS algorithm converge independent of the eigenvalue spread. Though the LS algorithm based on transversal and lattice filter structures are similar in convergence performance, the lattice filter structure will offer significant advantage of simplifying the computational complexity that occurs in the alternate structure.

In the simulation results, the superior convergence properties of the LS lattice filter can help to lessen the problem of having an increase in the MSE when the input data changes from sinusoidal to white. This sudden increase can be brought down to near the optimum value again in a short time.

REFERENCES

[1] Sondhi M.M. and Berkley D.A., " Silencing Echoes on the Telephone Network", Proc. IEEE, Vol.68, No.8, pp.948-963, Aug. 1980.

[2] Mitchell O.M.M. and Berkley D.A., "A Full-Duplex Echo Suppressor Using Center-Clipper", Bell Syst. Tech. J., Vol.50, No.5, pp.1619-1630, May-June 1971.

[3] Kelly, Jr, J.L. and Logan, Jr, B.F., "Self-Adaptive Echo Canceller", U.S. Patent, No.3,500,000, March 10, 1970.

[4] Sondhi M.M., "An Adaptive Echo Canceller", Bell Syst. Tech. J., Vol.46, No.3, pp.497-511, March 1967.

[5] Weinstein S.B., "Echo Cancellation in the Telephone Network", IEEE Commun. Mag., pp.9-15, Jan. 1977.

[6] Gritton C.W.K and Lin D.W., "Echo Cancellation Algorithms", IEEE ASSP MAG., pp.30-37, April 1984.

[7] Messerschmitt D.G., "Echo Cancellation in Speech and Data Transmission", IEEE J.Select. Areas Commun., Vol SAC-2, No.2, PP.283-297, Mar. 1984.

[8] Widrow B., "Adaptive Filters I : Fundamentals", (Tech Rept. 6764-6), Dec. 1966.

[9] Widrow B et al, "Adaptive Noise Cancelling : Principles and Applications", IEEE Proc., Vol.63, No.12, PP.1692-1716, Dec. 1975.

[10] Gitlin R.D. and Meadors H.C., "The Tap-Leakage Algorithm : an Algorithm for the Stable operation of a Digital Implemented FSE", Bell Syst. Tech. J., Vol.61, No.8, pp.1817-1839, Oct. 1982.

[11] Atal B.S. and Schroeder M.F., "Predictive Coding of Speech Signals and Subjective Error Criteria", IEEE Trans Acoust., Speech, Signal Processing, Vol. ASSp-27. No.13, pp.247-254, June 1977.

[12] Honig M.L., "Echo Cancellation of Voice Band Data Signals Using Recursive Least-Squares and Stochastic Gradient Algorithms", IEEE Trans. Commun., Vol.Com-33, NO.1, PP.65-73 Jan. 1985.

[13] Mueller K.H., "A New Digital Echo Canceller for Two-Wire Full Duplex Data Transmission", IEEE Trans. Commun., Vol.com-24, No.9, pp.956-967, Sept. 1976.

[14] Falconer D.D., Mueller K.H. and Weinstein S.B. "Echo Cancellation Techniques for Full Duplex Data Transmission on Two-Wire Lines", Proc. NTC, Dallas Sept. 1976.

[15] Weinstein S.B., "A Passband Data-Driven Echo Canceller for Full Duplex Transmission on Two-Wire Circuits", IEEE Trans. Commun., pp.654-666, July 1977.

[16] Werner J.J., "Effects of Channel Impairements on the Performance of an In-Band Data-Driven Echo Canceller", At&T Technical J., Vol.64, No.1, Jan. 1985.

[17] Haykin S., "Introduction to Adaptive Filters", Macmillan Publishing Company, New York, 1984.

[18] Proakis J.G., "Digital Communications", McGraw-Hill Book Company, 1983.

[19] Gersho A., "Adaptive Equalization of Highly Dispersive Channels for Data Transmission", Bell Syst. Tech. J., Vol.48., pp.55-70, Jan. 1969.

[20] Ungerboeck G., "Theory on the speed of Convergence in Adaptive Equalizers for Digital Communication", IBM. J. Res Develop., pp.546-555, Nov. 1972.

[21] Gitlin R.D. and Weinstein S.B., "On the Required Tap Weight Precision for Digitally Implemented, Adaptive, Meas-Square Equalizers", Bell Syst. Tech. J., Vol.58, No.2. PP.301-321, Feb.1979.

[22] Honig M.L. and Messerschmitt D.G., "Adaptive Filters : Structures, Algorithms and Applications", Kluwer Academic Publ., 1984.

[23] Evci C.C. and Bellanger M.G., "Characteristics of Adaptive Filters with Leakage", Int. Conf. Commun., 1984.

[24] Nishitani T., Aikoh A., Araseki T., Ozawa K. and Marinta R., "A 32-Kbit/s Toll Quality ADPCM Codec Using a Single Chip Signal Processor", Proc. ICASSP, 82 Paris, pp.960-963, May 1982.

[25] Gray R.M., "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices", IEEE Trans. Inform. Theory, Vol.It-18, No.6, pp.725-730, Jan. 1972.

[26] Mueller M.S, "Least-Squares Algorithms for Adaptive Equalizers", Bell Syst. tech. J., Vol.60. No.8, pp.1905-1925, Oct. 1981.

[27] Falconer D.D. and Ljung L., "Application of Fast Kalman Estimation to Adaptive Equalization", IEEE Trans. Commun., Vol.Com-26, pp.1439-1446, 1978.

[28] Satorious S.H. and Pack J.K., "Application of Least-Squares Lattice algorithms to Adaptive Equalization", IEEE Trans. Commun., Vol.Com-29, pp.136-142, Feb. 1981.

[29] Shichor E., "Fast Recursive Estimation Using the Lattice Structure", Bell Syst. Tech. J., Vol.61, pp.97-115, Jan. 1982.