Adaptive Linear and Nonlinear Filters

by

(Frank) Xiang Yang Gao

A thesis submitted in conformity with
the requirements for the degree of
Doctor of Philosophy

November 199 1

Department of Electrical Engineering
University of Toronto
Toronto, Ontario
CANADA

Abstract

The research work presented in this thesis advances the state-of-the-art of adaptive filtering by developing an efficient adaptive linear cascadeIIR filter, proposing four adaptive linearization schemes, introducing adaptive nonlinear recursive state-space (ANRSS) filters, and applying the algorithms to loudspeaker measurements.

Adaptive cascade IIR filters have the advantages of easy stability monitoring and good sensitivity performance. A novel technique of backpropagating the desired signal is proposed for a general cascade structure, which is then applied to a cascade IIR filter. The equation-error formulation is shown to be a special case of the backpropagation formulation.

Inevitable nonlinearities in systems intended to function linearly sometimes severely impair system performance. Three adaptive linearization schemes are devised to reducenonlinearities in these systems using adaptive FIR filters. They achieve linearization by canceling nonlinearity at the system output, post-distorting the signal, or pre-distorting the signal. The pre-distortion scheme is applied to linearize a loudspeaker model.

The adaptive nonlinear filters previously reported are almost all of FIR type. Although they have some nice properties. their computation requirements are impractical for those applications with long impulse responses. Hence, ANRSS filters are introduced as alternatives and efficient methods for gradient computation are developed to facilitate further their real-time application. The stability and the convergence of the filters are studied.

Measurements are performed on a loudspeaker system. Solutions of some problems arising from the practical data are proposed. Then. the algorithms developed in the thesis are applied to the measurement data.

Acknowledgements

I am very grateful to Dr. W. Martin Snelgrove, who has led me into and intelligently guided me in this exciting research area. I would also like to express my gratitude to Dr. David A. Johns for insightful advice and valuable discussions and Drs. Peter Schuck and Eric Verreault for performing loudspeaker measurements.

Thanks are due to the committee members of my Ph.D. examination, particularly Professors Kenneth Jenkins and Raymond Kwong, for their constructive suggestions.

My friends in the Snelly Zone have contributed greatly to the work presented in this thesis and to the thesis itself by reviewing my papers and thesis and creating a stimulating and friendly environment. They are Richard Schreier, Anees Munshi, Zhiquiang Gu, Ayal Shoval. Steve Jantzi, Guilin Zhang, Weinan Gao, Carl Sommerfeldt, Chris Ouslis, Duncan Elliott, and Di Stefano.

I would like to thank my family members in China for their support and Gail and Jim Collins for their friendship.

I am indebted to my wife. whose understanding, sacrifice, and love have inspired me.

Table of Contents

Chapter One

Introduction

1.1 Motivations and Contributions of the Thesis

Research work in this dissertation makes several contributions to the area of adaptive filtering. First, an efficient adaptive linear cascade IIR filter is developed on the basis of a novel backpropagation formulation. Next, four adaptive linearization schemes are developed for weakly nonlinear systems. Adaptive linearization of a loudspeaker system is proposed and is demonstrated successfully on an analytical loudspeaker model. Then, adaptive nonlinear recursive state-space (ANRSS) filters are introduced. Efficient gradient computation algorithms are presented for these nonlinear IIR filters, and the problems of their stability and convergence are studied. Finally, the algorithms proposed in the thesis, together with adaptive linear FIR, nonlinear FIR, equation-error, and linear state-space filters, are applied to measured data of a loudspeaker.

An adaptive filter is preferred to a fixed filter when an exact filtering requirement may be unknown and/or this requirement may be mildly non-stationary. While adaptive linear FIR filters are widely used [1], they have been found too computationally expensive for systems with long memory. The desire to search for efficient adaptive filters has triggered active research of adaptive IIR filters [2,3]. Adaptive linear IIR filters are often implemented using direct-form realizations which have poor sensitivity

performance and for which stability is hard to guarantee. Adaptive cascade IIR filters have an easy stability check and good sensitivity performance [4]. However, they have expensive gradient computation, usually quadratic in the filter order. An efficient adaptive cascade IIR filter is developed in this thesis to solve this problem. A novel technique is proposed for a cascade IIR filter, which suggests that the desired signal be backpropagated and the intermediate errors be generated. The intermediate errors are then minimized. In this filter, the poles are realized by cascading all-pole second-order sections, while the zeros are realized by one transversal section. The complexity of adaptation is only about the same as that of the filter itself. In the proposed filter, the transversal section and the inverse all-pole second-order sections, namely, the all-zero second-order sections, are adapted. It is shown that the equation-error formulation [2] is just a special case of the backpropagation formulation.

In most adaptive signal processing applications, system linearity is assumed and adaptive linear filters are thus used. However, the performance of adaptive linear filters is not satisfactory in applications where nonlinearities are significant. For example, adaptive linear filters are normally used in channel equalization of data transmission. In high-speed data communication, channel nonlinearities greatly impair transmission quality and adaptive nonlinear filters are thus preferred to adaptive linear filters for equalization [5,6]. On the other hand, nonlinearities in systems intended to function linearly are not very strong in comparison with nonlinearities in systems intended to work nonlinearly. This thesis is mainly concerned with systems intended to be linear. The weakness of nonlinearities in such systems is taken advantage of in this thesis to develop efficient adaptive nonlinear filtering algorithms.

Reduction of excessive nonlinearities in a system intended to function linearly sometimes can not be successfully accomplished by conventional techniques. For example, a typical modern audio system consists of a high quality digital signal source, an electronic amplifier, a loudspeaker system, and A/D and D/A converters. Various design techniques have been used by designers to achieve linearity in each part. The loudspeaker system usually has the most significant nonlinearities among the parts of the audio system, hence, it is the limiting component. Linearization by feedback has difficulties combating nonlinearities in a loudspeaker due to a delay in the feedback signal which may cause instability. An adaptive linearization technique may be a solution.

The topic of adaptive linearization has not been well studied. Motivated by practical applications, three adaptive linearization schemes are presented in this thesis for weakly nonlinear systems using adaptive FIR filters. In the first scheme, linearization is performed by canceling nonlinearity at the output of a physical system. In the second scheme, a nonlinear post-processor is employed to post-distort signals, while in the third scheme, a pre-processor is used. The first scheme can achieve perfect linearization if an accurate estimate of the nonlinear signal is obtained. Other two schemes are able to reduce the nonlinearity substantially if the nonlinearity is weak. These schemes may be suitable for different applications. The scheme with a pre-processor is proposed to linearize a loudspeaker. Based on an analytical loudspeaker model, simulations of the proposed method have been performed. The results have shown that nonlinear distortions of a loudspeaker can be reduced significantly.

The reported adaptive nonlinear filters are almost all of FIR type. which have similar advantages and disadvantages to adaptive linear FIR filters. The computation

cost of an adaptive linear FIR filter increases linearly with the effective length of a system impulse response, while the computation of adaptive nonlinear FIR filters is super-linear with this length and thus is much more computationally demanding in the case of a long impulse response.

This thesis introduces a general class of adaptive nonlinear IIR filters, namely, ANRSS filters. The filters are recursive and thus generally have an infinite impulse response. They are expected to have many applications and are especially attractive for those with long memories where adaptive nonlinear FIR filters are too expensive to use. Efficient methods, which significantly reduce computation for gradients, are developed to facilitate their application in real-time signal processing. Guidelines are presented for maintaining the stability of an ANRSS filter and it is shown that an ANRSS filter can be approximated by a time-variant linear system whose stability can be more easily monitored. It is found that the convergence speed depends on the eigenvalue spread of the correlation matrix of the coefficient gradient signals. The theoretically predicted convergence rate agrees quite well with the actual value in simulation. Furthermore, an adaptive linearization scheme based on the filters is proposed for a class of nonlinear systems. The scheme is applied to linearize a loudspeaker model with nonlinearity in the suspension system. It is also proposed that the adaptive filters could be used to perform echo cancellation in a data communication channel with nonlinearity. Numerical experiments are performed on identification of a simple first-order system, identification and linearization of a loudspeaker model, and cancellation of echo. Although these filters are presented in the digital domain, they are also applicable in the continuous-time domain. This is another advantage of ANRSS filters and another

motivation for introducing them.

To see the performance of the algorithms in a practical situation, measurements are performed on a loudspeaker. Solutions to practical issues, such as inversion of a bandpass transfer function, are discussed. The algorithms, together with some existing techniques, are then applied to the measurement data.

1.2 Tour Map of the Thesis

This dissertation has four core chapters: Chapters Three, Four, Five, and Six, where adaptive filtering algorithms are developed and tested. And it has three supporting chapters, Chapters One, Two, and Seven, which present principles, conduct surveys, and draw conclusions.

Chapter One discusses motivations and contributions of this thesis and gives a thesis outline.

Chapter Two first discusses the need for an adaptive filter. Next, it presents adaptation laws, principles of adaptive linear FIR filters, and principles of adaptive IIR filters. Then, it conducts a survey of adaptive nonlinear filters and a survey of applications of adaptive nonlinear filters. This chapter furnishes the reader with the necessary background theory and information on the state-of-the-art.

Chapter Three presents results on an adaptive linear cascade IIR filter. An idea of backpropagating the desired signal is proposed for a cascade IIR filter. Then, stability monitoring of an IIR second-order section and the convergence of the filter are discussed. Finally, simulation results are presented.

Chapter Four presents the results on adaptive linearization using adaptive FIR filters. Three adaptive linearization schemes are proposed. One of the schemes is also proposed to linearize a loudspeaker and simulations demonstrate it is able to achieve a significant reduction in distortion.

Chapter Five introduces ANRSS filters. After discussing motivations for studying an adaptive nonlinear IIR filter, a nonlinear recursive state-space structure for adaptive nonlinear IIR filters is introduced and efficient methods are developed for gradient computation. Then, the issues of stability and convergence are addressed and potential applications of ANRSS filters are proposed. Finally, numerical results are presented.

Chapter Six describes measurements of a loudspeaker system and discusses some issues associated with the data and their solutions. Then results are presented on application of the preposed algorithms to the data.

The last chapter, **Chapter Seven** draws conclusions and suggests what should be carried out further to improve or continue the work.

References

[1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, New Jersey: Prentice-Hall, 1985.

[2] J.J. Shynk, "Adaptive IIR Filtering," *IEEE ASSP Magazine, pp.4 - 21,* April 1989.

[3] C.R. Johnson, Jr., "Adaptive IIR Filtering: Current Results and Open Issues," *IEEE Trans. on Information Theory,* vol.IT-30, pp.237-250, March 1984.

[4] T. Kwan and K.W. Martin. "Adaptive Detection and Enhancement of Multiple Sinusoids Using a Cascade IIR Filter," *IEEE Trans. on Circuits and Systems,* vol. **36,** pp.937-947, July 1989.

[5] D.D. Falconer, "Adaptive Equalization of Channel Nonlinearities in QAM Data Transmission Systems," *The Bell System Technical Journal,* vol.57, *pp.2589 2611,* Sept. 1978.

[6] E. Biglieri, A. Gersho, R.D. Gitlin, and T.L. Lim, "Adaptive Cancellation of Nonlinear Intersymbol Interference for Voiceband Data Transmission," *IEEE J. Selected Areas in Communications,* vol.SAC-2, *pp.765777,* Sept. 1984.

Chapter Two

Principles and A Survey

2.1 Introduction

This chapter presents some principles of adaptive linear and nonlinear filters and conducts a concise survey of the research in the area. Active research on adaptive filters has been carried out for about three decades. Hence, many algorithms and structures have been developed and a rich body of literature has been formed. This chapter focuses on those concepts, algorithms, and structures related to this thesis. The adaptation laws are first outlined with the emphasis on the least mean square (LMS) algorithm. Then the adaptive linear FIR and IIR filters are discussed. Finally, adaptive nonlinear filters and their applications are presented.

2.2 The Need for an Adaptive Filter

A conventional fixed filter, which is used to extract information from an input time sequence, is linear and time invariant. An adaptive filter is a filter which automatically adjusts its coefficients to optimize an objective function. A conceptual adaptive filter is shown in Fig.2.1, where the filter minimizes the objective function of mean square error by modifying itself and is thus a time varying system. An adaptive filter is useful when an exact filtering operation may be unknown and/or this operation may be mildly nonstationary.

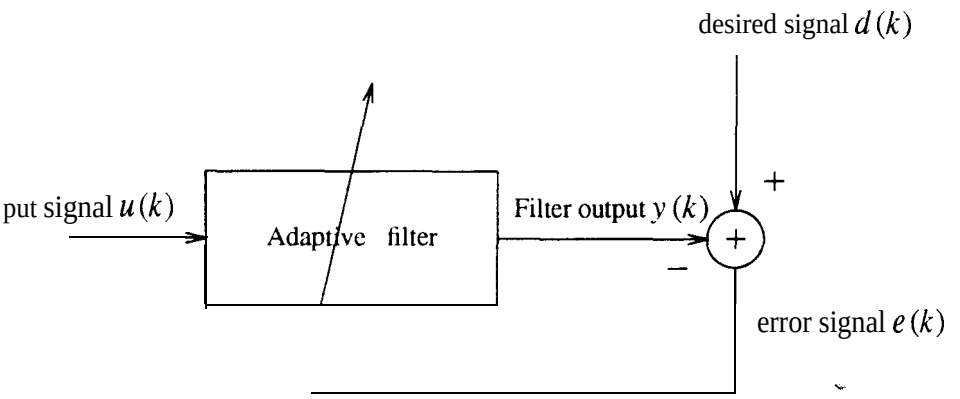


Fig.2.1 An adaptive filter.

Adaptive filters have found applications in many areas such as speech processing, data communications, image processing, and sonar processing. Two adaptive signal processing applications will be discussed in this section to help illustrate the need for an adaptive filter. One application is equalization of a data transmission channel [1] and another is noise cancellation [2].

2.2.1 Equalization of a Data Transmission Channel

The rapidly increasing need for computer communications has been met primarily by higher speed data transmission over the widespread telephone network. Binary data are converted to voice-frequency signals, transmitted, and converted back. The frequency response of a telephone line with nominal passband 300 Hz to 3000 Hz deviates from the ideal of constant amplitude and constant delay and thus time dispersion results. In pulse amplitude modulation (PAM), each signal is a pulse whose amplitude

level is determined by a symbol. The effect of each symbol transmitted over a time-dispersive channel extends beyond the time interval used to represent that symbol.

Assuming that the channel is linear, the sampled data symbol at the receiver can be represented as a convolution of the channel impulse response h_i with the transmitted data symbols $u(k)$,

$$y(k) = \sum_{i=0}^{\infty} h_i u(k-i) + n_0(k) \tag{2.1}$$

where n_0 is a noise signal. The sampled data symbol can also be expressed as

$$y(k) = h_\delta u(k-\delta) + \sum_{i \neq \delta} h_i u(k-i) + n_0(k) \tag{2.2}$$

where δ is the effective delay of the channel. The first term is the attenuated and delayed data symbol and the second term is the intersymbol interference among symbols due to the dispersion of the channel. An adaptive filter can be used to remove the intersymbol interference by inverting the channel. The need for adaptive filtering arises from a lack of prior knowledge of the impulse response *h* and from the time variance of the channel.

A typical receiver is shown in Fig.2.2 [l]. A pre-filter suppresses the out-of-band noise. A timing recovery device detects the data symbol rate so that the sampler can work at this rate. After sampling, an adaptive equalizer, often an adaptive transversal filter in the case of PAM data transmission, inverts the channel and removes the interference. At the beginning, a training sequence is generated and is used to train the adaptive filter. At the output of the filter, a slicer is used to detect the symbols transmitted. After the training period, the detected symbols are used to adapt the filter.

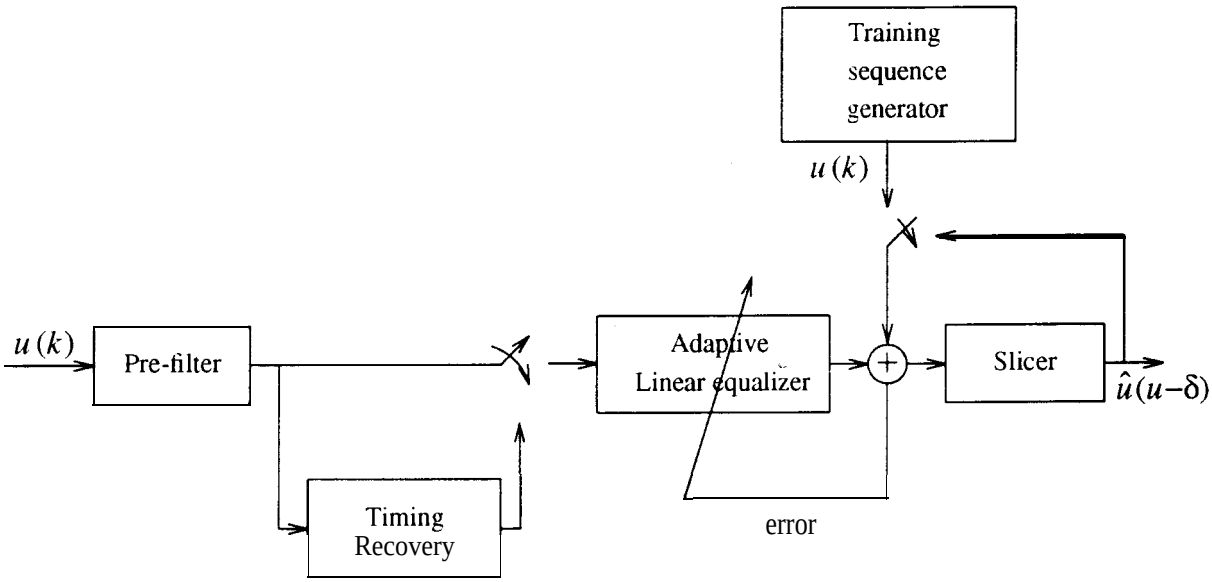


Fig.22 A receiver utilizing adaptive equalization.

2.2.2 **Noise Cancellation**

A signal corrupted by additive noise can be estimated by passing it through a filter, such as the pre-filter mentioned above, that tends to suppress the noise while leaving the signal relatively unchanged. Prior knowledge of the characteristics of both the signal and the noise is required for the design of fixed filters. Adaptive filters are sometimes preferred since little or no prior knowledge of the signal or noise characteristics is required for their design.

Adaptive noise cancellation is illustrated in Fig.2.3 [2]. The first sensor receives a signal s plus an uncorrelated noise n_1. A second sensor picks up the noise n_2 from the

noise source, which is independent of the signal s and correlated in some way with the primary noise n_1. An adaptive filter provides an estimate of the noise n_1 using the measured original noise n_2. The estimate of the noise n_1 is then subtracted from the primary signal $s+n_1$ to cancel the primary noise n_1. As explained in the following, the adaptive filter achieves this by minimizing the power of the system output z, which is the difference between the primary signal and the filter output.

Taking account of the assumption that s is uncorrelated with n_1 and n_2, it can be shown [2] that

$$\min E(z^2) = E(s^2) + \min E((n_1 - y)^2) \tag{2.3}$$

where E indicates the expectation operator. Hence, when the filter adjusts its coefficients so that $E(z^2)$ is minimized, $E((n_1-y)^2)$ is minimized. The filter output y is then a least square estimate of the primary noise n_1. Moreover, considering

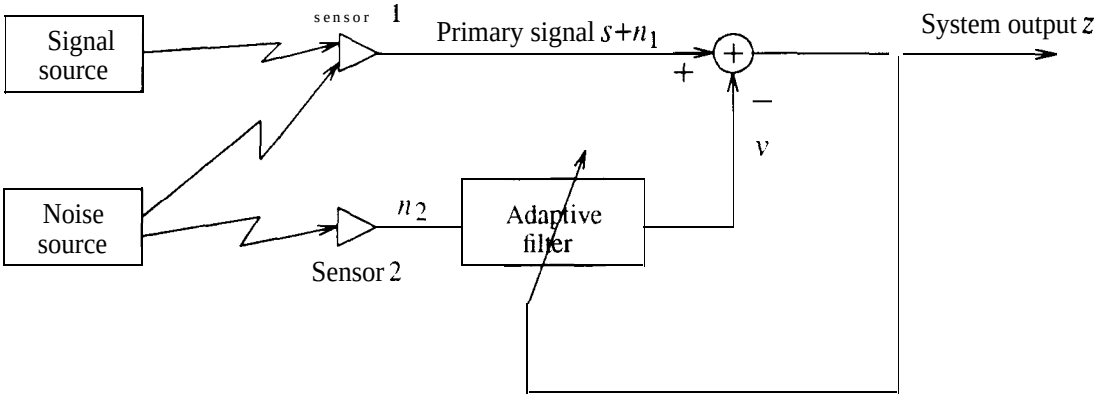


Fig.2.3 Adaptive noise cancellation.

$$z - s = n_1 - y \tag{2.4}$$

it is clear that minimizing the power of the system output by an adaptive filter minimizes the output noise power. This adaptive noise cancellation technique, however, is not universal; for instance, it is not very applicable for removing the additive channel noise in data transmission discussed above since the noise source is unknown.

These two applications clearly demonstrate the need for adaptive filters. Although a fixed filter could be used to replace the adaptive filter in the data transmission receiver or in the noise canceler, it would not be as effective as an adaptive filter since the characteristics of the data transmission channel and the noise channel are usually unknown and change slowly with time. The properties of the noise to be canceled are also often unknown to the designer. All these make an adaptive filter preferred or necessary.

2.3 Adaptation Laws

As discussed in the previous section, an adaptive filter adapts, by some means, its coefficients to achieve a prescribed objective. A widely applied objective is minimizing the mean square of the output error, which is defined as the difference between the desired signal and the filter output. This is called the output-error formulation which is the basis of the majority of the algorithms proposed in this thesis. All the adaptive filters reviewed in this chapter are based on this formulation. Another popular formulation, the equation-error formulation, will be introduced in Chapter Three for comparison with the backpropagation formulation developed in this thesis. One class of

adaptation laws for the output-error formulation is gradient based, which has the following general expression for coefficient adjustments

$$\mathbf{p}^{k+1} = \mathbf{p}^{k} - \mu \mathbf{R}^{-1} \left(\frac{\partial E\left(e^{2}(k)\right)}{\partial \mathbf{p}} \right) \tag{2.5}$$

where $\mathbf{p}$ is a vector of parameters, k is the iteration number or number of samples, μ is a diagonal matrix of step sizes, R is a matrix chosen to improve the convergence rate, $E(e^{2}(k))$ indicates the mean squared error (MSE), the error signal $e(k)$ is defined as $e(k) = d(k) - y(k)$, the signal $d(k)$ is the desired signal, and the signal $y(k)$ is the filter output. If the matrix is chosen to be the correlation matrix of the gradient signals, the dependence of the filter convergence on the eigenvalue spread of the gradient signals becomes substantially reduced. In this adaptation algorithm, filter coefficients are updated in the opposite direction of the gradient vector so that the adaptation goes downhill on the MSE surface.

For real-time signal processing, the computation load should be reduced to a minimum. If the mean squared error $\boldsymbol{E}(e^{2}(k))$ is approximated by the instantaneous square error $e^{2}(k)$, the gradient in the above adaptation law can be replaced by its corresponding estimate which is noisy, but unbiased. Furthermore, if the matrix R is replaced by the unit matrix, the adaptation law becomes the well-known and most widely used real-time adaptation law - the LMS algorithm [1-3]

$$\mathrm{p}^{k+1} = \mathbf{p}^{k} + 2\mu e(k) \frac{\partial y(k)}{\partial \mathbf{p}} \tag{2.6}$$

where y is the filter output.

The step sizes of an adaptive filter control the convergence speed. Smaller step sizes result in a slower convergence and a lower residual MSE, while larger step sizes

cause a faster convergence and a higher residual MSE. Step sizes which are too large make the filter unstable. Choice of step sizes depends on the filter structure, the adaptation algorithm, and the properties of the input signal. How to choose a step size is well understood for adaptive FIR filters, but not for adaptive IIR filters. All the adaptive filters discussed in this dissertation are based on the LMS algorithm.

2.4 Adaptive Linear FIR Filters

There are two popular kinds of adaptive linear FIR filters, transversal filters and FIR lattice filters. We discuss only adaptive linear transversal filters since knowledge of adaptive lattice filters is not essential for discussing the algorithms presented in this thesis. Adaptive linear transversal filters are popular because of such nice properties as guaranteed stability and global convergence. An adaptive linear transversal filter, shown in Fig.2.4, has the following form

$$y(k) = \sum_{i=0}^{n} h_i u(k-i) \tag{2.7}$$

where n is the filter order. u is the input signal, and h is the impulse response of the filter.

It has been shown [4] that assuming the coefficients of an adaptive filter change slowly, we have

$$\frac{\partial y(k)}{\partial p} = Z^{-1}\left[\frac{\partial Y(z)}{\partial p}\right] \tag{2.8}$$

where p is a filter coefficient to be adapted and Z^{-1} indicates the inverse z-transform

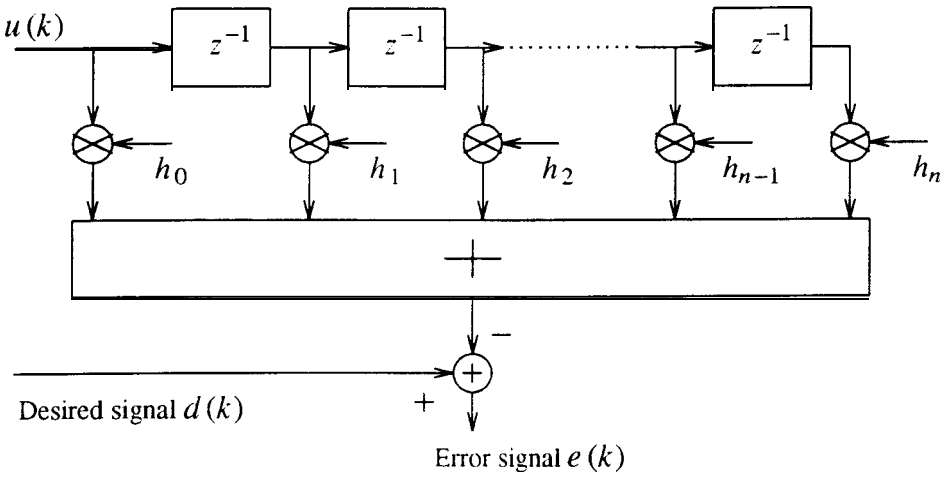


Fig.2.4 Adaptive linear transversal filter.

and $Y(z)$ [1] is the z-transform of the time domain variable $y(k)$. The relationship in Equation (2.8) permits us to carry out the derivation of gradient evaluation formulas in both the time and z-domains. Deriving gradient formulas in the z-domain is often very convenient for an adaptive linear filter, as we shall see in the following sections and the chapter on linear cascade IIR filters. Obviously, the gradient vector of the transversal filter coefficients is

$$\frac{\partial Y(z)}{\partial \mathbf{h}} = (1\ z^{-1}\ z^{-2} \cdots\ z^{-n}\ )^T U(z) \tag{2.9}$$

where $\mathbf{h} = (h_0\ h_1 \cdots h_n)^T$. Using the LMS algorithm in Equation (2.6), we can update the coefficients according to

[1] In this thesis. a time domain variable is in lower case and has an index k, *e.g., y(k)* , and its z-transform counterpart is in upper case and has an index z , *e.g. Y(z)*

$$\mathbf{h}^{k+1} = \mathbf{h}^{k} + 2\mu e(k)\mathbf{u}(k) \tag{2.10}$$

where u(k) $= (u(k)\,u(k-1)\cdots u(k-n))^{T}$.

To simplify the statistical analysis of the LMS algorithm for a transversal filter, it is often assumed [38] that the current input signal vector u(k) of the transversal filter is uncon-elated with its previous values $\mathbf{u}(k-1), \mathbf{u}(k-2), \ldots . \mathbf{u}(0)$. Although the assumption is often violated in practice since the input signal is colored, experiences have shown that the results obtained are quite useful.

Considering the adaptation formula in Equation (2.10) and the assumption made above, we can write

$$E(\mathbf{e}_h(k+1)) = (\mathbf{I} - 2\mu\mathbf{R})E(\mathbf{e}_h(k)) \tag{2.11}$$

where ***E*** indicates the expectation operator, the vector $\mathbf{e}_h(k+1)$ is the difference of the coefficient vector $\mathbf{h}^{k+1}$ and the Wiener solution $\mathbf{h}_0$, R is the correlation matrix of the input signal u(k), and $\mathbf{I}$ is the identity matrix. It has been shown that the mean of the coefficient error $\mathbf{e}_h$ goes asymptotically to zero if the step size μ satisfies [38]

$$0 < \mu < \frac{1}{\lambda_{\max}} \tag{2.12}$$

where $\mu_{\max}$ is the maximum eigenvalue of the correlation matrix $\mathbf{R}$. For a chosen step size $\mu = \alpha/\mu_{\max}$, the convergence time constant τ is

$$\tau = \frac{1}{\lambda_{\min}\mu} = \frac{\lambda_{\max}}{\lambda_{\min\,CY.}} \tag{2.13}$$

where cx is a constant between 0 and 1 and $\lambda_{\min}$ is the minimum eigenvalue of the matrix R.

In a practical application or in a simulation, step sizes should be chosen smaller than the theoretical upper bounds obtained above because of the noise in gradient estimates. The analysis presented above focuses on the necessary conditions on which the mean coefficient error vector $\mathbf{e}_h$ of an FIR section converges to its Wiener solution. However, these conditions do not guarantee a finite variance for the coefficient error vector nor a finite mean square output error. A smaller upper bound for the step size was obtained for an adaptive LMS transversal filter [39,40], when both the necessary and sufficient conditions were considered. For a transversal filter having a step size μ, an input correlation matrix R with eigenvalues λ_i, it was shown [39,40] that the convergence is ensured if

$$0 < \mu \leq \frac{1}{\lambda_{max}} \tag{2.14}$$

and

$$\sum_{i=0}^{n} \frac{\mu\lambda_i}{1-2\mu\lambda_i} < 1 \tag{2.15}$$

where λ_{max} is the maximum eigenvalue. A criterion, which is more conservative and easier to use, is

$$0 < \mu \leq \frac{1}{3tr(\mathbf{R})} \tag{2.16}$$

The convergence speed of the LMS algorithm depends on the eigenvalue spread of the correlation matrix of the input signal. To speed up convergence. one can choose the correlation matrix of the input signal as the matrix R in Equation (2.5) or perform transform to orthogonalize the input signal [37].

2.5 Adaptive Linear IIR Filters

Although adaptive FIR filters have nice properties, they are found to be expensive for some applications, such as echo cancellation in acoustical systems, where system impulse responses are long. Adaptive IIR filters may be computationally more efficient for these applications. This has sparked active research on adaptive IIR filters. Several IIR structures have been investigated, which include direct form [3,9,15,32], lattice form [11,13], recursive state-space form [7,8], parallel form [12,14], and cascade form [5,10,33,34].

2.5.1 Adaptive Direct-Form Filters

Adaptive direct-form filters are very popular in the literature and they can be described as

$$y(k) = \sum_{i=1}^{n} a_i y(k-i) + \sum_{i=0}^{n} h_i u(k-i), \tag{2.17}$$

where a_i and h_i are the feedback and the feedforward coefficients, respectively. The filter output can be written in the z-domain

$$Y(z) = \frac{H(z)}{C(z)} U(z) \tag{2.18}$$

where

$$H(z) = \sum_{i=0}^{n} h_i z^{-i}$$

and

$$C(z) = 1 - \sum_{i=1}^{n} a_i z^{-i}$$

The filter described in the time domain in Equation (2.17) or in the z-domain in Equation (2.18) can be rearranged as a cascade of an IIR section l/C (z) followed by a transversal FIR section $H(z)$. The filter output can be rewritten as

$$Y(z) = H(z) Y_{iir}(z) \tag{2.19}$$

where Y_{iir} is the output of the IIR section and is equal to

$$Y_{iir}(z) = \frac{1}{C(z)} U(z) \tag{2.20}$$

Hence, the gradient vector for h is obviously

$$\frac{\partial Y(z)}{\partial \mathbf{h}} = (\, 1 \; z^{-1} \cdots z^{-n} \,)^T Y_{iir}(z) \tag{2.21}$$

Differentiating both sides of Equation (2.18) with the coefficients a_i results in

$$\frac{\partial Y(z)}{\partial \mathbf{a}} = (\, z^{-1} \; z^{-2} \cdots z^{-n} \,)^T \left(\frac{1}{C(z)}\right) Y(z) \tag{2.22}$$

where $\mathbf{a} = (\, a_1 \; a_2 \cdots a_n \,)^T$.

The filter structure and the implementation of the gradient computation is depicted in Fig.2.5. The IIR section of the input side and the FIR section form the filter. The gradient signals for the coefficients of the FIR section are the states of the section. The gradient signals of the filter's IIR section are obtained by passing the filter output through another IIR section. With the filter structure in this figure instead of the one suggested in Equation (2.17) the output of the IIR section $y_{iir}(k)$ is computed when computing the filter output $y(k)$. Hence, evaluation of gradients of the feedforward coefficients h_i according to Equation (2.21) involves no further computation. Equation (2.22) shows that evaluation of gradients of the feedback coefficients a_i needs only half of the computation required in computing the adaptive filter output. This method of computing gradients for the Output-error Direct-form Filter (ODF) is very efficient. These results

were presented in [15] and similar results were obtained for the recently proposed linear recursive state-space structure [7,8].

2.5.2 **Adaptive Cascade IIR Filters**

An adaptive filter may update its coefficients into an unstable region. It may be important to prevent this from happening by some means, for example. a stability

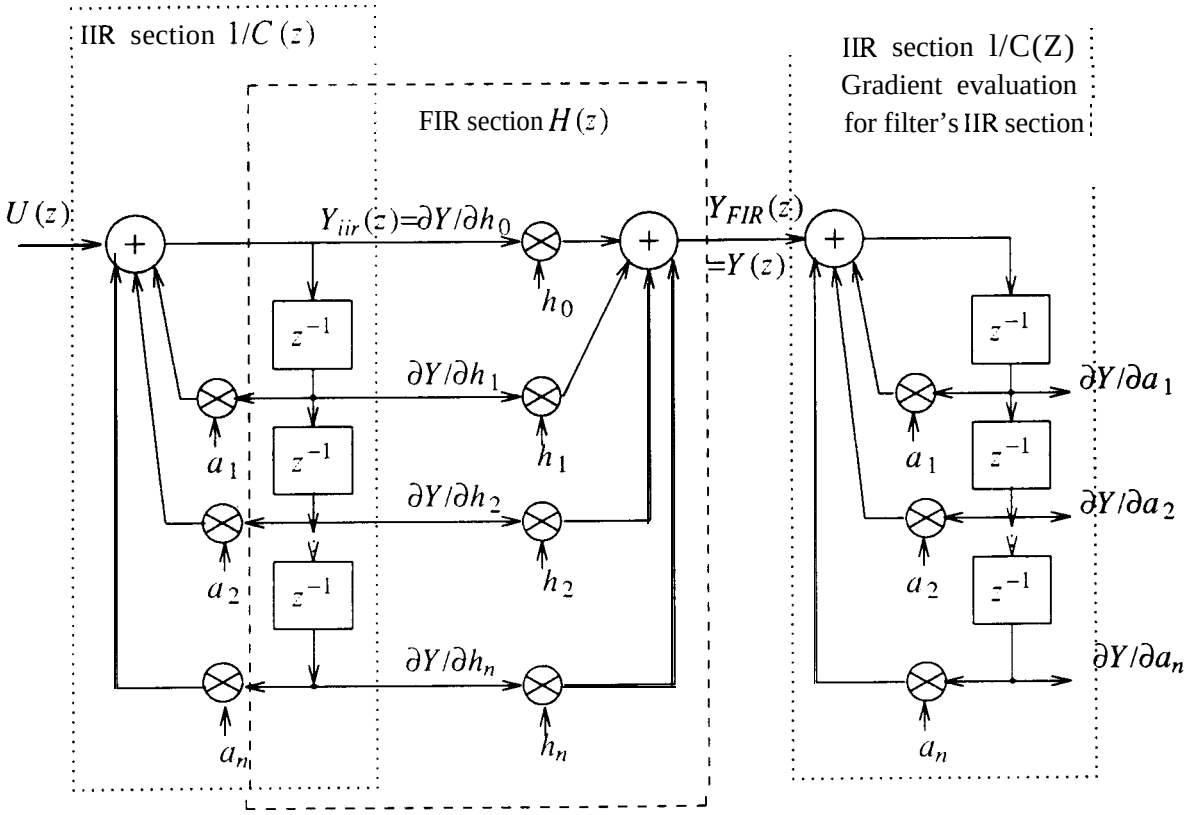


Fig.2.5 An adaptive output-error direct-form filter and evaluation of gradients for its coefficients.

check. Although the direct-form structure is very popular in the literature, it does not have an easy stability check and it also has poor coefficient sensitivities which may result in large residual MSE. On the other hand. an adaptive cascade IIR biquad filter has an easy stability check and good coefficient sensitivities. It can handle multiple poles just by allowing two sections to have identical coefficients, while the parallel form cannot (unless the structure is modified during operation to include cascaded or forth-order sections). For the spectral analysis application, the resonant frequencies of the biquads of an adaptive cascade filter will be identical to the frequencies of the input sinusoids after convergence and in the absence of noise [5]. This is not the case [5] for the parallel structures [41,42] which exhibit bias. A cascade biquad filter is shown in the upper part of Fig.2.6. If the order, n, of the system is odd, one "biquad" is a first-order filter. The filter output can be written as

$$Y(z) = B_m(z)B_{m-1}(z) \cdots B_1(z)U(z) \tag{2.23}$$

where B_i is the transfer function of the ith biquad. Let $\mathbf{p}_i$ indicate the coefficient vector of the ith biquad, then we have

$$\frac{\partial Y(z)}{\partial \mathbf{p}_i} = B_m(z) \cdots B_{i+1}(z)\frac{\partial B_i(z)}{\partial \mathbf{p}_i}B_{i-1}(z) \cdots \boldsymbol{B}_1(z)U(z) \tag{2.24}$$

Considering that the output of the $(i-1)th$ biquad is

$$Y_{i-1}(z) = B_{i-1}(z) \cdots \boldsymbol{B}_1(z)U(z) \tag{2.25}$$

as indicated in Fig.2.6. we have

$$\frac{\partial Y(z)}{\partial \mathbf{p}_i} = B_m(z) \cdots B_{i+1}(z)\frac{\partial B_i(z)}{\partial \mathbf{p}_i}Y_{i-1}(z)$$

$$= \mathbf{G}_i(z)B_m(z) \cdots B_{i+1}Y_{i-1}(z) \tag{2.26}$$

where $i = 1, 2, \ldots, m, Y_0 = U$, and

$$\mathbf{G}_i(z) = \frac{\partial B_i(z)}{\partial \mathbf{p}_i}$$

The implementation of the gradient filters is shown in the lower part of Fig.2.6. The gradient computation cost is quadratic in terms of the system order n and is quite high.

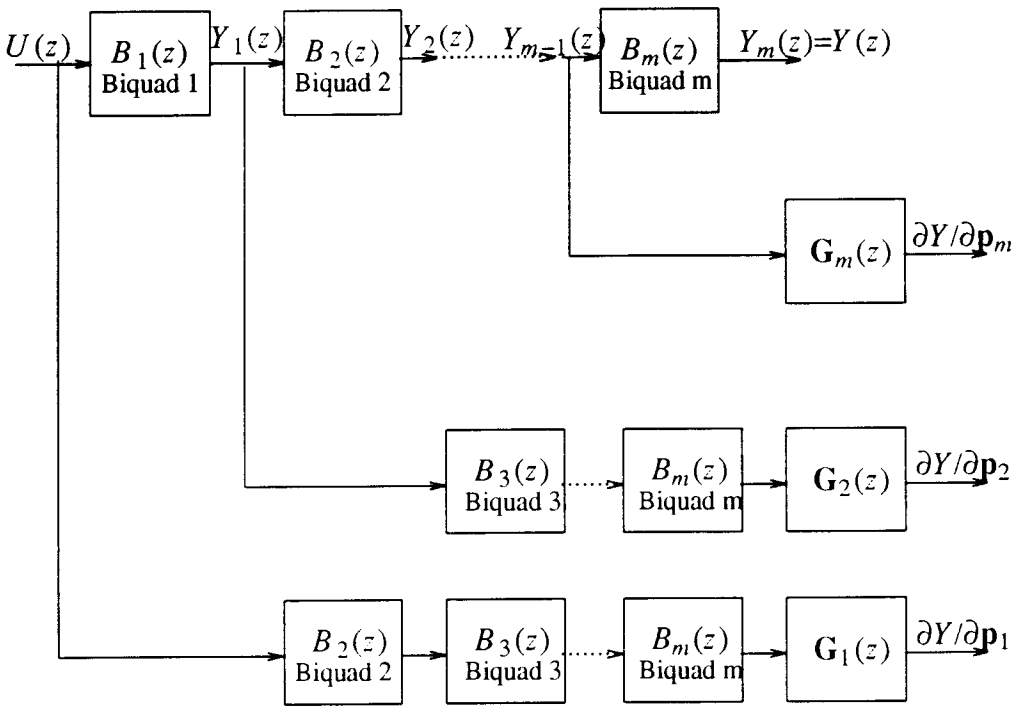


Fig.2.6 A cascade biquad structure shown in the upper part, together with its gradient evaluation in the lower part.

2.5.3 Adaptive Linear Recursive State-Space Filters

Implementation of adaptive linear IIR filters with the recursive state-space structure has been studied [7,8]. It was found that some recursive state-space structures may have much faster adaptation speeds and much better round-off noise performance than the direct-form structure in some cases. An adaptive linear recursive state-space filter can be described by the following equations [7,8]

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}u(k) \tag{2.27a}$$

$$y(k) = \mathbf{C}^T\mathbf{x}(k) + du(k) \tag{2.27b}$$

where x is the state variable, **A** is the feedback matrix, d is the feedthrough coefficient, and **B** and C are the input coefficient vector and the output coefficient vector, respectively.

Taking the z-transform of Equations (2.27a) and (2.27b), differentiating both sides of the equations by a_{ij}, which is the element at the *ith* row and *jth* column of the matrix **A,** and solving for $\partial Y(z)/\partial a_{ij}$, we have

$$\frac{\partial Y(z)}{\partial a_{ij}} = \mathbf{C}^T(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{e}_i X_j(z) \tag{2.28}$$

where $X_j(z)$ is the *jth* element of the state vector $\mathbf{X}(z)$ and $\mathbf{e}_i$ is a vector with the *ith* element being unity and the others being zero. Taking the transpose of both sides of the above equation results in

$$\frac{\partial Y(z)}{\partial a_{ij}} = \mathbf{e}_i^T(z\mathbf{I} - \mathbf{A}^T)^{-1}\mathbf{C}X_j(z) \tag{2.29}$$

Similarly, the formula for evaluation of the gradient for the input coefficient vector **B** can be obtained

$$\frac{\partial Y(z)}{\partial b_i} = \mathbf{e}_i^T (z\mathbf{I} - \mathbf{A}^T)^{-1} \mathbf{C} U(z) \tag{2.30}$$

Let $\mathbf{F}_j$ be the gradient vector for all the elements on the *jth* column of A and let Q indicate the gradient vector for all the elements of $\mathbf{B}$, namely

$$(\mathbf{F}_j(k))_i = \frac{\partial y(k)}{\partial a_{ij}}, \quad q_i(k) = \frac{\partial y(k)}{\partial b_i} \tag{2.3 1}$$

where $(\mathbf{F}_j)_i$ and q_i are the *ith* elements of $\mathbf{F}_j$ and $\mathbf{Q}$, respectively. Writing Equations (2.29) and (2.30) in the time domain gives

$$\mathbf{F}_j(k+1) = \mathbf{A}^T \mathbf{F}_j(k) + \mathbf{C} x_j(k) \tag{2.32}$$

and

$$\mathrm{Q(k+l)} = \mathbf{A}^T \mathbf{Q}(k) + \mathbf{C} u(k) \tag{2.33}$$

Hence, the gradients for the coefficients A and B can be computed recursively, using the systems in Equations (2.32) and (2.33), which are similar in structure to the filter itself. The gradients for the output coefficient vector C and the feedthrough coefficient d can be easily obtained

$$\frac{\partial Y(z)}{\partial \mathbf{C}} = \mathbf{X}(z) \tag{2.34}$$

$$\frac{\partial Y(z)}{\partial d} = U(z) \tag{2.35}$$

Using the LMS algorithm, the filter coefficients are updated according to

$$a_{ij}^{k+1} = a_{ij}^k + 2\mu_a e(k)(\mathbf{F}_j(k))_i \tag{2.36}$$

$$b_i^{k+1} = b_i^k + 2\mu_b e(k) q_i(k) \tag{2.37}$$

$$c_i^{k+1} = c_i^k + 2\mu_c e(k) x_i(k) \tag{2.38}$$

$$d^{k+1} = d^k + 2\mu_d e(k) u(k) \tag{2.39}$$

where μ_a, μ_b, μ_c, and μ_d are step sizes.

Computing gradients for **A** and **B** using Equations (2.32) and (2.33) is very efficient: one gradient filter is able to compute gradients for all the elements of **B** or all the elements of a column of **A,** compared to a simple-minded method which needs one gradient filter for each element of **B** or **A.** It is also shown [7,8] that it is possible to adapt a single column or row of a recursive state-space filter and reduce the total number of gradient filters to one. These single row or single column adaptive filters are shown to have superior convergence properties compared to direct-form filters in over-sampled applications when the desired poles can be estimated.

2.6 Adaptive Nonlinear Filters

System linearity is often assumed in many adaptive signal processing applications. Nonlinearities in real applications limit system performance. Thus, modem signal processing theory and practice are becoming more and more concerned with the design of efficient nonlinear filters. One approach to designing adaptive nonlinear filters is based on the truncated Volterra series. From a theoretical point of view, the Volterra filter is attractive since it can deal with a general class of nonlinear systems while its output is still linear with respect to various higher power system coefficients or impulse responses.

For a nonlinear system satisfying certain conditions, the output $y(k)$ can be expanded into a Volterra series

$$y(k) = \sum_{i_1=0}^{\infty} h_1(i_1)u(k-i_1) + \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} h_2(i_1,i_2)u(k-i_1)u(k-i_2) + \cdots +$$

$$\sum_{i_1=0}^{\infty} \cdots \sum_{i_m=0}^{\infty} h_m(i_1, i_2, \cdots, i_m) u(k-i_1) u(k-i_2) \cdots u(k-i_m) + \cdots \tag{2.40}$$

where $h_1, h_2, \ldots, h_m, \ldots,$ are the coefficients of the system. The sum with h_1 can be considered as a convolution of the input signal, $u(k)$, with the impulse response, $h_1(k)$, of a linear system. This sum is a linear term. The sum with h_i will be referred to as *ith* power term, since if the input $u(k)$ is multiplied by a scalar a, this term will yield a factor α^i. In particular, the sum with h_2 and the sum with h_3 will be referred to as the quadratic term and the cubic term, respectively. The linear term models the linearity of the system; the rest of the series models the nonlinearity.

Generally speaking, the series has an infinite number of terms and each term is an infinite sum. The computation and memory requirements make it impossible to base adaptive filters on this series if no simplification is made. In practice, the series is able to model many system reasonably well if it just contains the several major terms and each term is a finite sum. Thus, the truncated series can be employed to construct adaptive filters. Furthermore, we can consider that $h_2(i_1, i_2)$, $h_3(i_1, i_2, i_3)$, and $h_m(i_1, i_2, \cdots, i_m)$ *are* symmetric, namely, the indexes of $h_2(i_1, i_2)$, $h_3(i_1, i_2, i_3)$, or $h_m(i_1, i_2, \cdots, i_m)$ *are* exchangeable. Then, an adaptive nonlinear filter can be based on the following truncated series

$$y(k) = \sum_{i_1=0}^{n_1} h_1(i_1) u(k-i_1) + \sum_{i_1=0}^{n_2} \sum_{i_2=i_1}^{n_2} h_2(i_1, i_2) u(k-i_1) u(k-i_2) + \cdots +$$

$$\sum_{i_1=0}^{n_m} \sum_{i_2=i_1}^{n_m} \cdots \sum_{i_m=i_{m-1}}^{n_m} h_m(i_1, i_2, \cdots, i_m) u(k-i_1) u(k-i_2) \cdots u(k-i_m) \tag{2.41}$$

where m is the total number of terms in the filter, n_1 is the order of the linear term, and $n_2, n_3, \ldots$ and n_m are the orders of the nonlinear terms. Note the changes in the upper

and lower limits of the summations in the nonlinear terms, which are the results of assuming that $h_2, h_3, \ldots$ and h_m *are* symmetric and assuming each sum is finite. Obviously, this filter has a finite impulse response, hence the name adaptive nonlinear FIR filter.

Updating the Volterra filter coefficients is often based on the LMS concept and is performed as follows [16-20]

$$h_j^{k+1}(i_1, i_2, \cdots, i_j) = h_j^k(i_1, i_2, \cdots, i_j) + 2\mu_j e(k)u(k-i_1)u(k-i_2)\cdots u(k-i_j) \quad (2.42)$$

where $j = 1, 2, \ldots m$ and μ_j is the step size for the *jth* power term. A quadratic Volterra filter is the one with the linear term and the quadratic term. Study of an adaptive quadratic filter has shown [16] that when the input signal is Gaussian, the LMS-based quadratic Volterra filter converges to the optimum solution asymptotically in the mean and the convergence speed depends on the squared ratio of maximum to minimum eigenvalues of the input correlation matrix.

A variant of the LMS algorithm, the sign algorithm [22], updates the filter coefficients using only the sign of the gradient. It uses only the direction of the gradient and loses the proportionality of the correction terms to the error. Therefore, the sign algorithm offers a slower and less accurate convergence in comparison to the stochastic algorithm, while leading to reduced-complexity implementation.

The gradient-based adaptive algorithms use a step size to control the convergence. The step size is fixed in most existing nonlinear adaptive algorithms. A technique for adjusting the step size was proposed in [21] for the sign algorithm presented in [22] in order to obtain faster convergence and reduce final MSE without excessively affecting

the advantage of the sign algorithm in the implementation complexity. It was shown that the structure resulting from the modified sign algorithm still remains much simpler than that related to the stochastic iteration algorithm.

An adaptive quadratic Volterra filter with minimum mean square error criterion was presented in a series of papers [16,29,31]. The order n_1 of the linear term and the order n_2 of the quadratic term were chosen as the same and were indicated by N. The quadratic Volterra series was written as

$$y(k) = \mathbf{H}_1^T \mathbf{u}(k) + tr\{\mathbf{H}_2[\mathbf{u}(k)\mathbf{u}^T(k) - \mathbf{R}_u]\} \tag{2.43}$$

where

$$\mathbf{H}_1 = [\, h_1(0)\ h_1(1) \cdots h_1(N-1)\,]^T$$

$$\mathrm{u(k)} = [\, u(k)\, u(k-1) \cdots u(k-N+1)\,]^T$$

$$\mathbf{H}_2 = \begin{bmatrix} h_2(0,0) & \cdots & h_2(0,N-1) \\ \vdots & \ddots & \vdots \\ h_2(N-1,0) & \cdots & h_2(N-1,N-1) \end{bmatrix}$$

$$\mathbf{R}_u = \begin{bmatrix} r(0) & \cdots & r(N-1) \\ \vdots & \ddots & \vdots \\ r(N-1) & \cdots & r(0) \end{bmatrix}$$

where r indicates the correlation of the input signal u:

$$r(j) = E(u(k)u(k-j))$$

The term $-tr(\mathbf{H}_2\mathbf{R}_u)$ is the dc component and is included to keep the estimate unbiased. When the input signal is Gaussian, the least mean square solution of the quadratic Volterra filter is as following:

$$\mathbf{H}_1 = \mathbf{R}_u^{-1}\mathbf{v}_{du} \tag{2.44}$$

$$\mathbf{H}_2 = \frac{1}{2}\mathbf{R}_u^{-1}\mathbf{T}_{du}\mathbf{R}_u^{-1} \tag{2.45}$$

where $\mathbf{v}_{du}$ is the correlation vector whose elements are defined as

$$v_{du}(i) = E(d(k)u(k-i))$$

and $\mathbf{T}_{du}$ is the cross-bicorrelation matrix whose elements are defined as

$$t_{du}(i,j) = E(d(k)u(k-i)u(k-j))$$

Direct solution of the matrixes $\mathbf{H}_1$ and $\mathbf{H}_2$ based on Equations (2.44) and (2.45) is called the batch-processing method [31].

Most adaptive Volterra filters work in the time domain. A frequency domain adaptive Volterra filter was presented in [18]. For a finite memory of length N, the filter converges to the equivalent time domain nonlinear adaptive filter presented previously. For a quadratic Volterra filter, the proposed method requires 0 (N^2) multiply-adds for N outputs as opposed to an $O(N^3)$ for the time domain.

The algorithms discussed above are based on the LMS concept. An algorithm based on a fast Kalman filter algorithm was presented in [19]. The algorithm developed is for a quadratic Volterra filter. It uses a stochastic approximation in the least square recursion equation for the quadratic weight matrix. The convergence of the algorithm for the quadratic weights was established. In a simulation experiment, the algorithm proposed converged faster than an LMS-based adaptation algorithm.

An adaptive nonlinear FIR filter was proposed based on a three-section model [27] rather than a Volterra series. It has three cascaded sections: linear system with memory, followed by a nonlinear memoryless system and then followed by another linear system with memory. This structure results in an efficient adaptive nonlinear filter which is a

special case of an adaptive Volterra filter. The LMS algorithm was applied to update the coefficients.

2.7 Applications of Adaptive Nonlinear Filters

Adaptive Volterra filters find their application in many fields. Typical applications include echo cancellation [23,24,26], equalization [28,30] and noise reduction [17,20] for communication channels.

In recent years adaptive echo cancellation has been extensively investigated in connection with some new services introduced in the telephone network because it allows for full-duplex transmission on two-wire circuits with full bandwidth in both directions. Application of this concept can be found in digital subscriber loop modems and voiceband data modems. A theoretical study of echo cancellation was conducted in [24]. The coefficients of the continuous-time Volterra series were represented by generalized Fourier series. It was proved that the proposed echo canceler converges and reduces the echo to zero in the absence of noise. A nonlinear echo canceler based on a Volterra filter was presented in [26]. The hardware realizations were based on combinatorial networks obtained by using distributed arithmetic. An adaptive nonlinear echo canceler was developed for data signals in [23]. A binary series expansion of a nonlinear function of the data was derived, which has the form of the Volterra series and has finite terms. Based on this expansion, an LMS adaptation, together with its hardware implementation, was developed.

Nonlinear distortion is now a significant factor hindering high-speed data transmission over telephone channels. Equalization must be adaptive since nonlinear distortion varies from one connection to another and also varies with time for each particular connection. Receivers utilizing nonlinear equalizers have been studied in [28,30] for passband QAM which is the preferred modulation scheme for achieving high data rates. The channel model was a three-section model discussed in the previous section. Equalization was achieved in [28] by utilizing an adaptive nonlinear filter in series with the nonlinear data channel to invert the channel. Simulations on recorded data from real channels have demonstrated that nonlinear decision-feedback equalization can significantly reduce the error rate for a variety of channel characteristics. Cancellation of nonlinear intersymbol interference in [30] is performed by employing an adaptive nonlinear filter in parallel with the channel to provide an estimate of the interference. The cancellation approach removes nonlinear interference terms without excessive noise enhancement and allows effective implementation with ROM's to generate the needed nonlinear signal variables. An orthogonalized version of the Volterra series was employed, resulting in an increased ability to correct for channel nonlinearities.

An adaptive nonlinear FIR filter was studied for noise cancellation in [17], where noise channels are assumed to consist of a nonlinear memoryless section followed by a linear dispersive section. An adaptive nonlinear filter was designed using this model. Including nonlinearity in the noise canceler was shown to increase the ability of the canceler to remove noise from the received signal. In a simulation, the nonlinear canceler achieved 22.3 dB of noise suppression, whereas the linear canceler was capable of suppressing the interference by only 6.4 dB.

Modern digital radio systems utilize highly bandwidth-efficient QAM modulation techniques to make use of the crowded microwave radio spectrum. The system becomes more sensitive to all types of linear and nonlinear distortion as the number of constellation points grows, To achieve maximum efficiency, the high-power amplifier of a transmitter operates in saturation, resulting in a nonlinear distortion in the transmitted signal. This nonlinear distortion becomes a critical issue. An adaptive linearization technique was presented in [36] which compensates for amplifier nonlinearity. It predistorts the signal before the amplifier. The algorithm operates in real time and is data directed. Development of the algorithm assumed that the power amplifier is memoryless and required that the signal not be filtered before the amplifier. Hence, the adaptive linearizer is also memoryless. The performance of the linearizer was further analyzed in [35].

Although most of the applications of adaptive filters are in communications systems, some work has been done to apply these filters in other fields. An adaptive quadratic Volterra filter was employed to model and forecast the sway motion response of a moored vessel to random sea waves [29,31]. Two procedures, batch-processing and adaptive, to implement the the filter were presented in these two papers and were discussed in the previous section. The batch-processing method uses the correlation and cross-bicorrelation functions to evaluate the quadratic filter coefficients. On the other hand, the adaptive method updates the filter weights as a new observation is available. Experimental results based on a scaled model wave basin test were used to demonstrate the utility of the nonlinear filters.

2.8 Summary

In this chapter, the principles of the adaptive linear and nonlinear filters were discussed and the research work in this field was briefly reviewed. The LMS algorithm, on which all the adaptation algorithms presented in this thesis are based, was introduced. The adaptive linear transversal filter was discussed. As well, adaptive linear IIR filters, with emphasis on the direct-form, cascade form, and state-space form filters, were described. Finally, the Volterra theory and adaptive nonlinear FIR filters were presented, together with their applications.

References

[1] M.L. Honig and D.G. Messerschmitt, *Adaptive Filters - Structures, Algorithms, and Applications,* Boston: Kluwer Academic Publishers, 1984.

[2] B. Widrow and S.D. Stearns, *Adaptive Signal Processing,* Englewood Cliffs, New Jersey: Prentice-Hall, 1985.

[3] J.R. Treichler, CR. Johnson and M.G. Larimore, *Theory and Design of Adaptive Filters, New* York, New York: John Wiley & Sons, 1987.

[4] K.W. Martin and M.T. Sun, "Adaptive Filters Suitable for Real-Time Spectral Analysis," *IEEE Trans. on Circuits and Systems, vol. CAS-33,* pp. 218-229, Feb. 1986.

[5] T. Kwan and K.W. Martin, "Adaptive Detection and Enhancement of Multiple Sinusoids Using a Cascade IIR Filter," *IEEE Trans. on Circuits and Systems,* vol. 36, pp.937-947, July 1989.

[6] J.J. Shynk, "Adaptive IIR Filtering," *IEEE ASSP Magazine, pp.4 - 21,* April 1989.

[7] D.A. Johns, "Analog and Digital State-Space Adaptive IIR Filters," *Ph.D. Thesis,* University of Toronto, 1989.

[8] D.A. Johns, W.M. Snelgrove, and A.S. Sedra, "Adaptive Recursive State-Space Filters Using a Gradient Based Algorithm," *IEEE Trans. on Circuits and Systerns,* vol. 37, pp.673-684, June 1990.

[9] C.R. Johnson, Jr., "Adaptive IIR Filtering: Current Results and Open Issues," *IEEE Trans. on Information Theory,* vol.IT-30, pp.237-250, March 1984.

[10] Y.H. Tam, P.C. Ching, and Y.T. Chan, "Adaptive Recursive Filters in Cascade Form," *IEE Proc.,* vol. 134, Pt. F, Comm., Radar & Signal Processing, pp.245 252, June 1987.

[11] I.L. Ayala, "On a New Adaptive Lattice Algorithm for Recursive Filters," *IEEE Trans. Acoustics, Speech, and Signal Processing,* vol. ASSP-30, pp. 316-319, April 1982.

[12] M. Nayeri and W.K. Jenkins, "Alternate Realizations to Adaptive IIR Filters and Properties of Their Performance Surfaces," *IEEE Trans. on Circuits and Systems,* vol. CAS-36, pp. 485-496, April 1989.

[13] N.I. Cho, C.H. Choi, and S.U. Lee, "Adaptive Line Enhancement by Using an IIR Lattice Notch Filter," *IEEE Trans. on Acoustics, Speech, and Signal Processing* , vol. *37,* pp. 585-589, April 1989.

[14] J.J. Shynk, "Adaptive IIR Filtering Using Parallel-Form Realizations," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. *37,* pp. 519-533, April 1989.

[15] F.F. Yassa, "Optimality in the Choice of the Convergence Factor for Gradient-Based Adaptive Algorithms." *IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-35,* pp. 48-59, Jan. 1987.

[16] T. Koh and E.J. Powers, "An Adaptive Nonlinear Digital Filter with Lattice Orthogonalization," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp.37-40, *1983.*

[17] M.J. Coker and D.N. Simkins, "A Nonlinear Adaptive Noise Canceler." *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp.470-473, *1980.*

[18] D. Mansour and A.H. Gray, "Frequency Domain Non-linear Adaptive Filter," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp.550-553, *198* 1.

[19] C.E. Davila, A.J. Welch, and H.G. Rylander, "A Second-Order Adaptive Volterra Filter with Rapid Convergence," *IEEE Trans. Acoustics, Speech, and Signal Processing,* vol.ASSP-35, pp. 1259-1263, Sept. 1987.

[20] J.C. Stapleton and S.C. Bass, "Adaptive Noise Cancellation for A Class of Nonlinear, Dynamic Reference Channels," *Proc. of IEEE International Symposium on Circuits and Systems,* pp.268-271, 1984.

[21] G.L. Sicuranza and G. Ramponi, "A Variable-Step Adaptation Algorithm for Memory-Oriented Volterra Filters," ***IEEE*** *Trans. Acoustics, Speech, and Signal Processing,* vol.ASSP-35, pp.1492-1494, Oct. 1987.

[22] G.L. Sicuranza and G. Ramponi, "Adaptive Nonlinear Digital Filters Using Distributed Arithmetic," *IEEE Trans. Acoustics, Speech, and Signal Processing,* vol.ASSP-34, pp.5 18-526, June 1986.

[23] 0. Agazzi, D.G. Messerschmitt, and D.A. Hodges, " Nonlinear Echo Cancellation of Data Signals," *IEEE Trans. Commun.,* vol. COM-30, pp. 2421-2433, NOV. 1982.

[24] E.J. Thomas, "Some Considerations on the Application of the Volterra Representation of Nonlinear Networks to Adaptive Echo Cancelers," *The Bell System Technical J.,* vol.50, pp.2797-2805, Oct. 1971.

[25] M.G. Bellanger, *Adaptive Digital Filters and Signal Analysis, New* York: Marcel Dekker, Inc., 1987.

[26] G. L. Sicuranza, A. Bucconi, and P. Mitt-i, "Adaptive Echo Cancellation with Nonlinear Digital Filters," *Proc. of IEEE International Cotference on Acoustics, Speech, and Signal Processing,* pp.3.10.1-4, 1984.

[27] C.F.N. Cowan and P.F. Adams, "Nonlinear System Modeling: Concept and Application," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp.45.6.1-4, March 1984.

[28] D.D. Falconer, "Adaptive Equalization of Channel Nonlinearities in QAM Data Transmission Systems," *The Bell System Technical J.,* vol.57, pp.2589-2611, Sept. 1978.

[29] T. Koh and E.J. Powers, "Second-Order Volterra Filtering and Its Application to Nonlinear System Identification," *IEEE Trans. Acoustics, Speech, and Signal Processing,* vol.ASSP-33, pp. 1445-1455, Dec. 1985.

[30] E. Biglieri, A. Gersho. R.D. Gitlin, and T.L. Lim, "Adaptive Cancellation of Nonlinear Intersymbol Interference for Voiceband Data Transmission." *IEEE J. Selected Areas in Communications,* vol.SAC-2, pp.765-777, Sept. 1984.

[31] T. Koh, E.J. Powers, R.W. Miksad, and F.J. Fischer, "Application of Nonlinear Digital Filters to Modeling Low-Frequency, Nonlinear Drift Oscillations of Moored Vessels in Random Seas," *Proc. of the 16th Annual Offshore Technology Conference,* pp.309-314, May 1984.

[32] H. Fan and W.K. Jenkins, "An Investigation of an Adaptive IIR Echo Canceler: Advantages and Problems," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol.36, *pp.* 1819- 1834, Dec. 1988.

[33] R.A. David, "A Modified Cascade Structure for IIR Adaptive Algorithms." *Proc. of 15th Asilomar Conference on Circuits Systems and Computers,* pp. 175-179, Nov. 1981.

[34] R.A. David, "A Cascade Structure for Equation Error Minimization," *Proc. of 16th Asilomar Conference on Circuits, Systems, and Computers,* pp. 182-186, Nov. 1982.

[35] S. Pupolin and L.J. Greenstein, "Performance Analysis of Digital Radio Links with Nonlinear Transmit Amplifiers," *IEEE J. on Selected Areas in Communications,* vol.SAC-5, pp.534-546, April 1987.

[36] A.A.M. Saleh and J. Salz, "Adaptive Linearization of Power Amplifiers in Digital Radio Systems," *Bell System Technical J.,* vol. *62,* pp. 1019- 1033, April 1983.

[37] D.F. Marshall, W.K. Jenkins, and J.J. Murphy, "The Use of Orthogonal Transforms for Improving Performance of Adaptive Filters," *IEEE Trans. on Circuits and Systems, vol. 36,* pp.474-484, April 1989.

[38] S. Haykin, *"Adaptive Filter Theory,"* 2nd Ed., New Jersey: Prentice-Hall, 1991

[39] A. Feuer and E. Weinstein, "Convergence Analysis of LMS Filters with Uncorrelated Gaussian Data," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. ASSP-33, pp. 222-230, Feb. 1985.

[40] L.I. Horowitz and K.D. Senne, "Performance Advantage of Complex LMS for Controlling Narrow-Band Adaptive Arrays," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. ASSP-29, pp. 722-736, June 198 1.

[41] D. Hush and N. Ahmed, "Detection and Identification of Sinusoids in Broadband via a Parallel Recursive ALE," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1985.*

[42] R.A. David, "Detection of Multiple Sinusoids Using a Parallel ALE," *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1984.*

Chapter Three

Adaptive Backpropagation Cascade IIR Filter

3.1 Introduction

An adaptive linear IIR filter has advantages in computation when a system is better modeled by a pole-zero transfer function than by a zero-only function, especially when poles are close to the unit circle in the z-domain. Several structures have been proposed for adaptive linear IIR filters, including direct form [1-4], lattice form [5-7], cascade-form [8-11], parallel-form [12,131, and recently, state-space structures [14]. Among them, the direct form is most popular in the literature. However, an adaptive filter may go unstable during adaptation and it is difficult to ensure stability of a direct-form filter with an order above two. It also has very poor sensitivity performance, which means that a slight change in a coefficient will result in a large change in filter output. This is undesirable for an adaptive filter since its coefficients are constantly affected by measurement noise and quantization noise. Both cascade form and parallel form have an easy stability check and low sensitivities. The parallel form has difficulties implementing a multiple pole.

A cascade IIR structure was developed for both the output-error formulation [10] and the equation-error formulation [11]. It implements the filter denominator in cascade form and the numerator in transversal form. An adaptive cascade filter, composed of IIR notch biquads, was developed for the output-error formulation in [8], which is suitable for detecting and enhancing multiple sinusoids in applications in

communications and radar. Another cascade IIR filter was presented in [9] using the equation-error formulation, where the second-order sections are expressed in terms of their roots and these roots, rather than the section coefficients, are adapted.

One problem of adaptive cascade filters is the complexity of computing filter gradients, which is normally quadratic in the filter order. To solve this problem, an efficient cascade IIR filter has been proposed based on a novel concept of backpropagating the desired signal [15]. The filter consists of a transversal section and cascaded all-pole second-order sections. The computation for adaptation is about the same as that required by the filter itself when the LMS algorithm is used. It has been shown that the equation-error formulation is only a special case of the method of backpropagating the desired signal. The adaptive filter presented here has a similar structure to those in [10,11], but requires much less computation.

3.2 Backpropagation Formulation

The popular output-error formulation minimizes the error computed at the filter output side. This section proposes a different scheme in which a desired signal is backpropagated and intermediate errors are generated, then the filter adjusts its coefficients to minimize the intermediate errors.

The complexity of gradient computation of a conventional output-error cascade filter is due to the fact that the filter objective is minimization of the error at the filter output and gradient signals of a section have to pass the subsequent sections to form gradient signals at the filter output side. If some kind of intermediate errors can be gen-

erated and intermediate errors, instead of the output error, are minimized, the computation will be more efficient. A structure with cascaded sections is shown in the upper part of Fig.3.1. The transfer function $T_i(z)$ can be arbitrary as long as its inverse is stable. The desired signal can be backpropagated into the system with cascaded inverse filter sections. Then, we can employ the intermediate desired signals to generate the intermediate error signals and adapt the coefficients. We now apply this idea to a cascade IIR filter.

The filter shown in the upper part of Fig.2.6 is not suitable for backpropagating the desired signal due to possible instability caused by inverting the numerators of the pole-zero biquads. If the desired signal is not backpropagated through the transversal section, the filter structure shown in the upper part of Fig.3.2 satisfies the stability

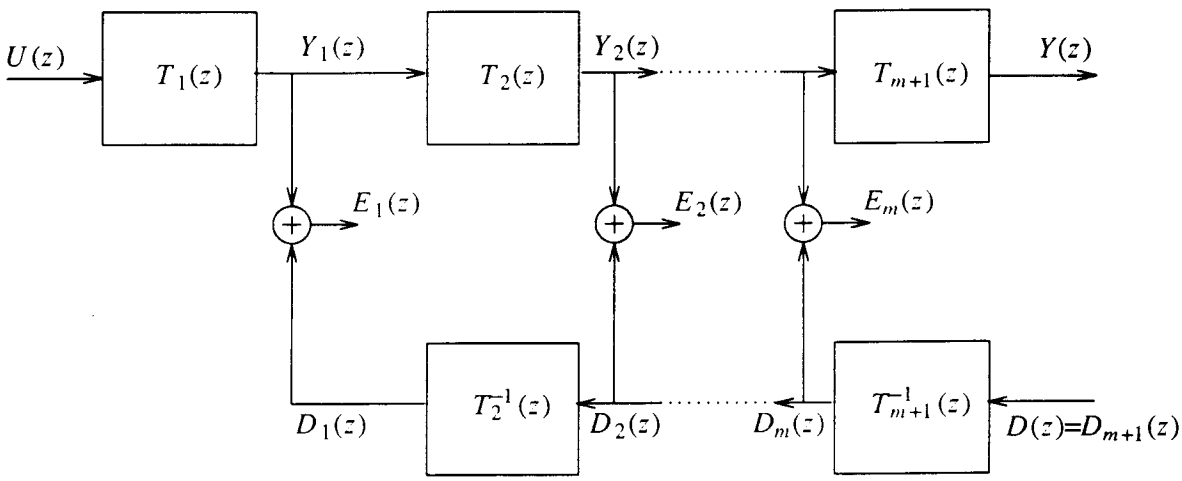


Fig.3.1 Backpropagation of the desired signal for a general cascade filter.

requirement of the backpropagation method. An *nth* order filter is described by a transversal section

$$Y_{fir}(z) = H(z)U(z) \tag{3.1}$$

and m all-zero second-order sections

$$Y_i(z) = \frac{1}{C_i(z)} Y_{i-1}(z) \qquad i = 1, 2, \cdots, m \tag{3.2}$$

where

$$H(z) = \sum_{i=0}^{n} h_i z^{-i},$$

$$C_i(z) = 1 - a_{i1} z^{-1} - a_{i2} z^{-2},$$

$$Y_0(z) = Y_{fir}(z),$$

and

$$Y_m(z) = Y(z)$$

The parameter m is equal to $n/2$ if the order n is even, otherwise it is equal to $(n+1)/2$ and one of the "second-order" section is in fact a first-order section. The intermediate desired signals and the intermediate errors are generated as shown in Fig.3.2. The transversal section H and the all-zero second-order sections C_i can be adapted to minimize the intermediate errors.

It is clear that the coefficient vector h of the transversal section H can be updated like that of an LMS transversal filter:

$$\mathbf{h}^{k+1} = \mathbf{h}^k + 2\mu_h e_1(k)\mathbf{u}(k) \tag{3.3}$$

where $\mathbf{u}(k) = (\, u(k)\ u(k-1)\ \cdots\ u(k-n)\,)^T$ and $\mathbf{h} = (\, h_0\ h_1\ \cdots\ h_n\,)^T$.

Since the signals $D_{i+1}(z), D_{i+2}(z), \ldots, D_{m+1}(z)$ (where $D_{m+1}(z) = D(z)$) are independent of the coefficients of the filter section C_i, the derivatives of the signal D_i

with respect to the coefficients of the section C_i are nonrecursive:

$$\frac{\partial D_i(z)}{\partial \mathbf{a}_i} = -(\ z^{-1}\ \ z^{-2}\ )^T D_{i+1}(z) \tag{3.4}$$

where the vector $\mathbf{a}_i = (\ a_{i1}\ \ a_{i2}\ )^T$. These coefficients can also be updated like those of an LMS transversal filter:

$$\mathbf{a}_i^{k+1} = \mathbf{a}_i^k + 2\mu_{a_i} e_i(k)\mathbf{d}_{i+1}(k) \tag{3.5}$$

where $\mathbf{d}_{i+1}(k) = (\ d_{i+1}(k-1)\ \ d_{i+1}(k-2)\ )^T$.

Each all-zero second-order section C_i is guaranteed to be stable and has a global minimum, although there is a possible bias in coefficient estimates when a measurement noise is present. An all-pole second-order section $1/C_i$ copies coefficients from its corresponding all-zero second-order section C_i, as indicated by the dashed lines in Fig.3.2. The intermediate errors are related to the output error by

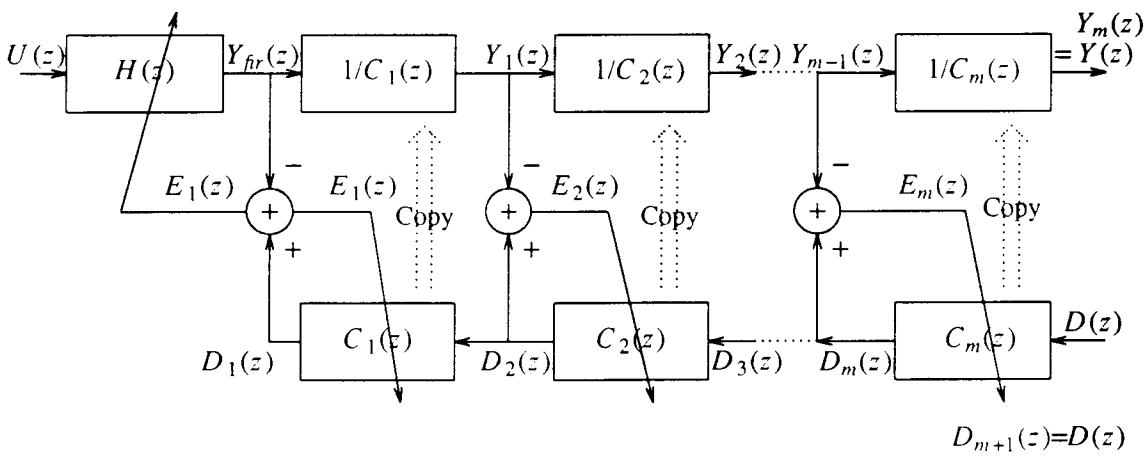


Fig.3.2 Adaptive Backpropagation Cascade Filter (BCF).

$$E_i(z) = C_i(z)C_{i+1}(z) \cdots C_m E(z) \quad i = 1, 2, \cdots, m$$

The intermediate errors are the filtered versions of the output error. Multiple objective functions are simultaneously minimized and they are closely related. It is clear that if the intermediate errors are zero, the output error is zero and perfect matching between the filter output and the desired signal is achieved. However, in other cases, minimization of the intermediate errors is not equivalent to minimization of the output error. To see the effects of noise, we can write the intermediate error as

$$E_i(z) = E_i'(z) + C_i(z)C_{i+1}(z) \cdots C_m(z)N(z) \qquad i = 1, 2. \cdots, m$$

where N is the noise on the desired signal, and E_i' is the error signal without noise. The filter attempts to minimize the mean squares of the intermediate error signals plus filtered noises. Hence, the noise will definitely introduce bias in coefficient estimates just as in the equation-error formulation.

In the rest of the chapter, the sections C_i will be referred to as the all-zero second-order sections and the section H will be referred to as the transversal section. The sections C_i and H will be collectively called the FIR sections.

3.3 Convergence Analysis

The all-zero second-order sections C_i and the transversal section H are adapted in a similar way in which an LMS transversal filter is adapted. However, rigorous analysis of the convergence of the backpropagation cascade IIR filter is much more difficult because there exists interaction among the different FIR sections.

Similar to analysis of mean coefficients of an adaptive LMS transversal filter discussed in Section 2.4, we can assume adaptation is so infrequent that the current input signal vector of each FIR section ($H, C_1, \ldots$ or C_m) is uncorrelated with its previous values. Additionally, Wiener solution of each FIR section is assumed to be independent of the desired signal of that section. Under those strong assumptions, it is expected that the mean value of the coefficient error goes asymptotically to zero if

$$0 < \mu_i < \frac{1}{\lambda_{i\max}} \tag{3.7}$$

and the mean value of the coefficient error goes asymptotically to zero with a finite variance if

$$0 < \mu_i < \frac{1}{3tr(\mathbf{R}_i)} \tag{3.8}$$

where $i = 0, 1, 2, \cdots, m$, and $i = 0$ is for the transversal section H, λ_{imax} is the maximum eigenvalue of the correlation matrix $\mathbf{R}_i$ of the input signal of the *ith* section. The bounds $1/\lambda_{imax}$ and $1/(3tr\mathbf{R}_i)$ will be referred to as the zero-mean upper bound and finite-variance upper bound, respectively. Interaction between different sections may require lower upper bounds.

The convergence time is expected to be

$$\tau_i = \frac{\lambda_{imax}}{\alpha_i \lambda_{imin}} \tag{3.9}$$

for a step size $\mu_i = \alpha_i/\lambda_{imax}$, where α_i is a constant between 0 and 1, and λ_{imin} is the minimum eigenvalue of $\mathbf{R}_i$.

The convergence time for the whole cascade IIR filter should be greater than or equal to the worst convergence time of all the FIR sections:

$$\tau \geq \max (\tau_0, \tau_1, \tau_2, \cdots, \tau_m) \tag{3.10}$$

3.4 Stability Monitoring

An adaptive filter will be unstable if a pole of an adaptive IIR filter stays outside the unit circle long enough. This instability can be prevented by checking the pole locations. One major advantage of the cascade structure is its easy stability check.

It can be shown that the stability region of an all-pole second-order section is a triangle which is defined by [2]

$$1 + a_{i1} - a_{i2} > 0, \quad 1 - a_{i1} - a_{i2} > 0, \text{ and } 1 + a_{i2} > 0. \tag{3.11}$$

The triangle is drawn in Fig.3.3.

This stability condition can be easily monitored during adaptation. An unstable update might be corrected by reducing step sizes. Once an unstable all-pole second-order section is detected in an iteration, the filter coefficients are computed again using smaller step sizes for the feedforward and feedback coefficients with the same gradients and error signal(s). If there is still at least one unstable all-pole second-order section, the filter coefficients will not be updated for that iteration. This is one of many possible ways of implementing stability monitoring and it is the one used in the simulations of this chapter.

An adaptive filter is essentially a time-variant system. The concept of pole is an approximation. The stability monitoring technique discussed above is based on this approximation and is not strictly required for convergence. When an adaptive filter enters an unstable region without stability monitoring, the MSE will increase and the

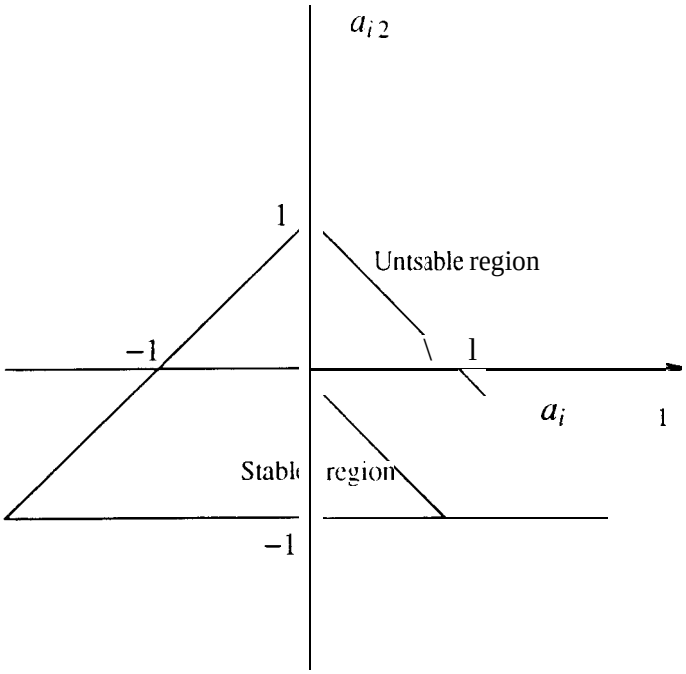


Fig.3.3 The stability triangle of an all-pole second-order section.

adaptation algorithm tends to push it towards a stable region. Whether the filter remains stable depends on such factor as how far the filter has gone into the unstable region. Stability monitoring may reduce convergence rates if an adaptive filter is always forced back into a stable region after its poles enter an unstable region.

3.5 Comparison with Equation-Error Formulation

The equation-error formulation was developed for a direct-form adaptive filter [2]. In the equation-error approach, the feedback signal is replaced by the desired signal so

that the feedback coefficients are updated in an all-zero, nonrecursive form. The filter output is

$$Y(z) = A(z)D(z) + H(z)U(z) \tag{3.12}$$

where

$$A(z) = \sum_{i=1}^{n} a_i z^{-i}, \quad H(z) = \sum_{i=0}^{n} h_i z^{-i}$$

Fig.3.4 gives a popular pictorial description of the Equation-error Direct-form Filter (EDF), where Y_{output} is the filter output. The error signal is

$$E(z) = (1 - A(z))D(z) - H(z)U(z) \tag{3.13}$$

which suggests Fig.3.5. Comparing Fig.3.5 with Fig.3.1, we find that Fig.3.5 shows that

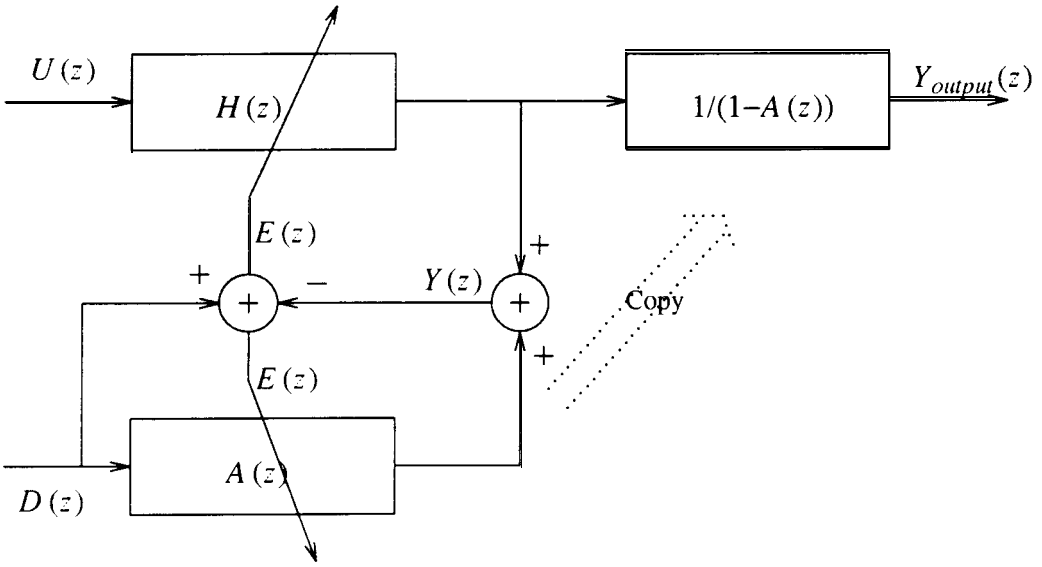


Fig.3.4 Equation-error formulation.

the equation-error formulation is just a special case of the backpropagation formulation illustrated in Fig.3.1 when there are only two cascaded sections.

3.6 Simulation Results

The algorithms proposed in this chapter have been simulated on a model matching problem, in which an adaptive filter attempts to match the transfer function of a reference system. A third order system has been used as a reference system in the simulations:

$$y(k)=0.5765y(k-1\,\mathbf{j}-0.7810\sim(k-2)+0.3821\sim(k-3)+$$
$$u(k)+1.6751u(k-1)+1.6751u(k-2)+u(k-3) \qquad (3.14)$$

where the system poles are 0.5122 and $0.0321\pm0.8643i$ and zeroes are -1 and

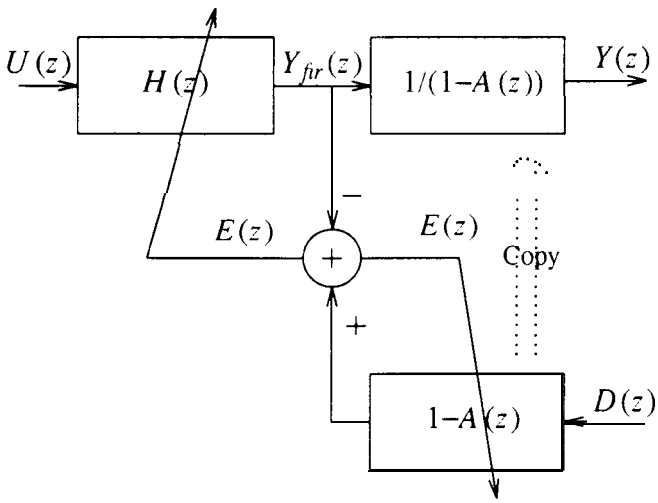


Fig.3.5 Alternative view of the equation-error formulation.

$-0.3375 \pm 0.9413i$.

For the cascade filter. the section C_2 is an all-pole second-order section whose optimal coefficient vector is $\mathbf{a}_2 = (\ 0.06429\ \ -0.748\ )^T$. The section C 1 is an all-pole first-order section whose optimal coefficient is $a_{11} = 0.5122$. The optimal coefficient vector of the transversal section H is h $= (\ 1\ \ 1.6751\ \ 1.6751\ \ 1)^T$.

In all the tests, the mean square errors (MSE) were computed using a data block of 100 samples. The input was a white Gaussian signal with unit variance ($0dB$). The initial values of the adaptive filter coefficients were set to zero. Three sets of simulations have been performed using the three adaptive filters: Output-error Direct-form Filter (ODF) of Fig.2.5, BCF of Fig.3.2, and EDF of Equation (3.11) or Fig.3.4.

In the first set of simulations, there was no additive noise on the reference signal (desired signal) and the step sizes were chosen so that the adaptive filters reached the computational noise floor (about *-300dB)* in the least number of iterations. The step sizes for the sections C_1 and C_2 were chosen the same for convenience. The convergence curves of the first set of simulations are the lower ones in Figs.3.6-8. Both the BCF and the EDF employ the backpropagated desired signals. So, it is interesting to compare the BCF with the EDF. Figs.3.7 and 3.8 show that the BCF had smoother curve and bigger step sizes. The ODF and the EDF had to use smaller step sizes because of the higher sensitivities of the direct-form structure. That the ODF and the EDF had spikier curves is also directly due to the higher sensitivities. These spikes are undesirable and although they can be reduced by using smaller step sizes, this will result in even slower convergence.

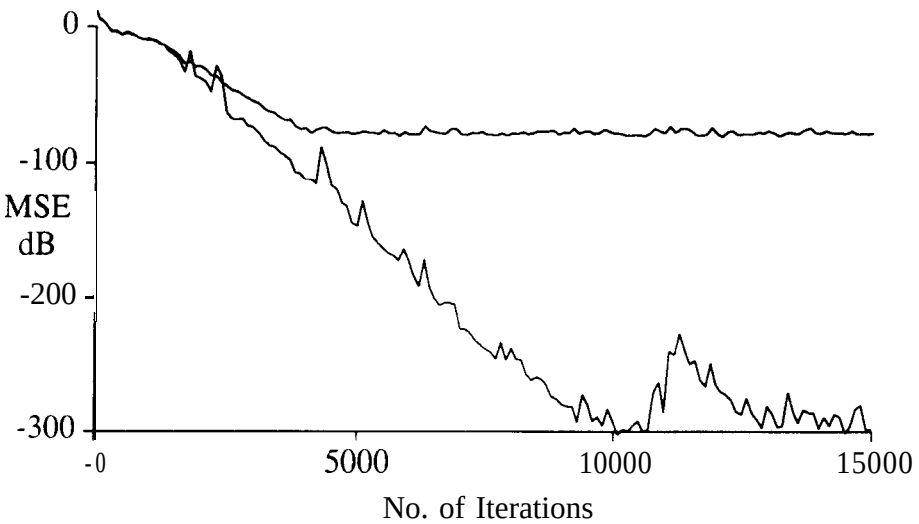


Fig.3.6 Convergence curves for the Output-error Direct-form Filter (ODF). Upper curve: additive noise of -80 dB, step size for IIR section = 0.0015 and step size for transversal section = 0.015. Lower curve: no additive noise, step size for IIR section = 0.002 and step size for transversal section = 0.03.

In practice, the reference signal is often contaminated by an additive noise, called measurement noise. An independent white noise of $-80dB$ was added to the reference signal to investigate the performance of the filters in the presence of measurement noise. The second set of simulations were performed under this condition. Suppose the adaptive filters are used to suppress echo in a data transmission channel. In such an application, the MSE is required to be less than about $-60dB$. Here, we require the MSE of an adaptive filter be below $-70dB$, allowing a safe margin. The step sizes were chosen so that the filters satisfied this MSE requirement in the least number of iterations. The convergence curves of the second set of simulations are upper ones in

Figs.3.6-3.8. The BCF converged after $2.3k$ iterations. The EDF converged at $5.0k$ iterations, while the ODF at $3.8k$ iterations. Fig.3.9 shows MSE contour with an adaptation path for the BCF. The adaptation path of the BCF is not normal to the contours because the BCF minimizes the intermediate errors and the contours were drawn using the output error. No visible bias in the filter coefficients was observed in the contour of the BCF because the noise level was modest.

In the above simulations, no stability check was employed. Instability of an adaptive filter can occur, which might be caused by, for example, a surge of measurement noise, large step size. and/or large gradients due to steep performance surface. A third set of simulations were performed based on the second set of simulations. All the conditions in the third set of simulations were the same as those of the second set, except that there was a measurement noise surge from sample 600 to 1000. The ODF, the BCF, and the EDF went unstable without stability monitoring when the measurement noise surge floor became high. Then stability monitoring was activated for the BCF, and the simulation was performed again. It remained stable and converged well. As expected, it worked well even if the noise level was very high. Fig.3.10 shows the convergence curve of the BCF with a noise of $26dB$ (standard deviation of 20), which shows a typical behavior of the BCF with stability monitoring. The filter worked normally before and after the noise surge. It had a high MSE level (but remained stable) during the surge because the gradient estimate was greatly corrupted.

In the following, the theoretical and practical maximum step sizes allowed for convergence are computed and compared. No measurement noise is added to the desired signal.

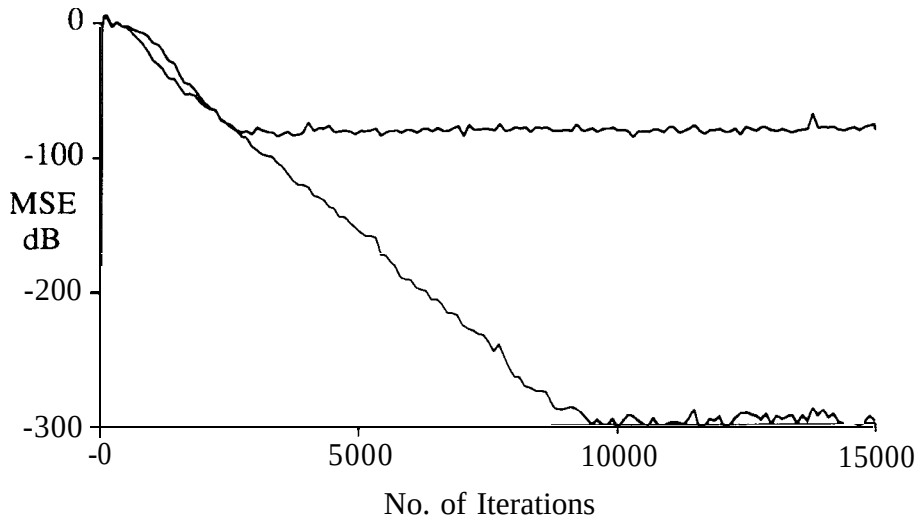


Fig.3.7 Convergence curves for the Backpropagation Cascade Filter (BCF). Upper curve: additive noise of -80 dB, step size for all-pole second-order section = 0.004 and step size for transversal section = 0.049. Lower curve: no additive noise, step size for all-pole second-order section = 0.006 and step size for transversal section = 0.09.

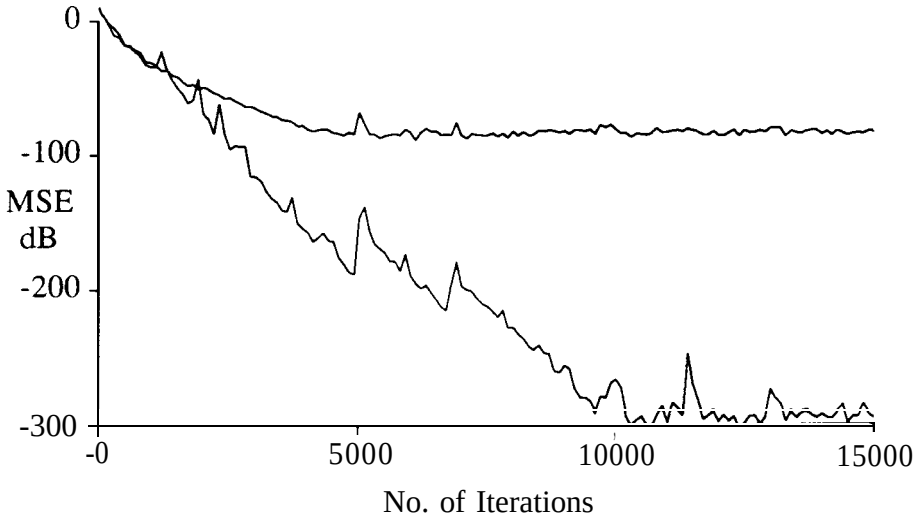


Fig.3.8 Convergence curves for the Equation-error Direct-form Filter (EDF). Upper curve: additive noise of -80 dB, step size for feedback section = 0.003 and step size for transversal section = 0.015. Lower curve: no additive noise, step size for feedback section = 0.005 and step size for transversal section = 0.07.

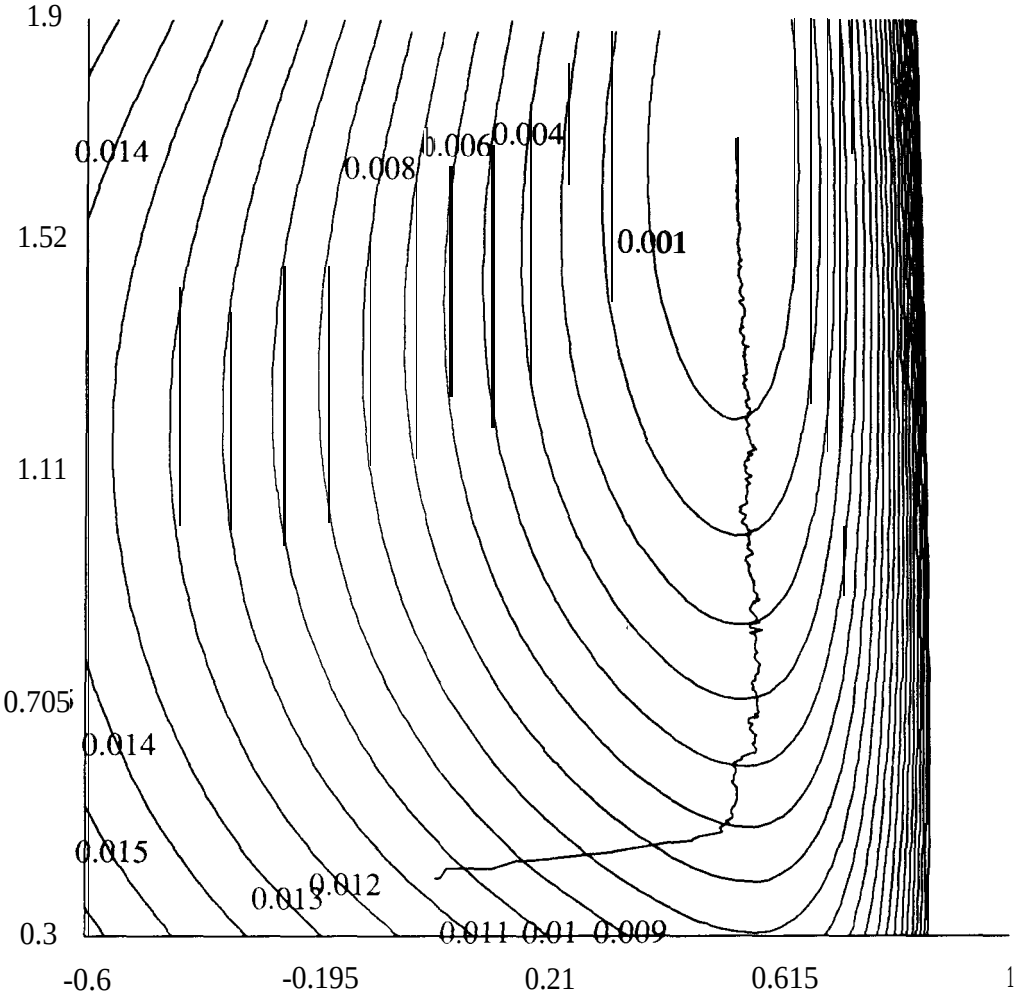


Fig.3.9 Contour plot for BCF. axis is a_{21}, and y axis is h (1). $\mu_a = \mu_h$ = 0.0007.

Fig.3.10 Convergence curve for BCF with measurement noise of 26 dB. Step size for all-pole second-order sections = 0.004 and step size for transversal section = 0.049.

Since the input signal is white Gaussian signal with a unit variance, the zero-mean upper bound for the step size of the transversal section H is unity, namely

$$\mu_h = 1 \tag{3.15}$$

and the finite-variance upper bound is

$$\mu_h = 0.08 \tag{3.16}$$

The input signal of the all-zero second-order section C_2 is the desired signal - the output of the reference physical system. The eigenvalues of the input-signal correlation matrix of this section are

$$eigenvalues\ of\ \mathrm{R}_{d_3} = (\ 3.7290\ \ 22.5310\ )^T \tag{3.17}$$

The zero-mean upper bound for the step size of the second-order section C_2 is

$$\mu_2 = 0.044 \tag{3.18}$$

and the finite-variance upper bound is

$$\mu_2 = 0.013 \tag{3.19}$$

To compute the correlation matrix of the input signal d_2 of the all-zero first-order section C_1, the optimal coefficient values were assigned to C_2. The eigenvalue of the correlation matrix is 22.85. The zero-mean upper bound for the step size of C_1 is

$$\mu_1 = 0.0438 \tag{3.20}$$

and the finite-variance upper bound is

$$\mu_1 = 0.015 \tag{3.21}$$

Simulations were performed to see what the practical values of the maximum step sizes are. When experimenting a step size for an FIR section (for example, section H), we set optimal values to other two FIR sections (for examples, sections C_1 and C_2). It was found that the practical maximum step sizes allowed for the sections H, C_1, and C_2 were

$$\mu_h = 0.19$$

$$\mu_1 = 0.0175 \tag{3.23}$$

and

$$\mu_2 = 0.0175 \tag{3.24}$$

They are between their corresponding zero-mean bounds and the finite-variance upper bounds.

Simulations were petformed to show the effect of the interaction of different sections on the choice of step sizes. The coefficients of the all three FIR sections were adapted initially from zeros. The filter converged when all the step sizes were reduced

to their corresponding practical maximum step sizes divided by 3.5. This shows that the interaction among different sections makes smaller step sizes necessary.

3.7 Summary

This chapter has studied adaptive cascade IIR filters which have an easy stability check and low parameter sensitivities. A novel concept has been proposed, which suggests backpropagating the desired signal through the inverse all-pole second-order sections and producing intermediate errors to be minimized. This concept was applied to a cascade IIR structure. resulting in an efficient adaptive cascade IIR filter. It has been shown that the equation-error formulation is just a special case of backpropagation of the desired signal.

References

[1] H. Fan and W.K. Jenkins, "An Investigation of an Adaptive IIR Echo Canceler: Advantages and Problems,"*IEEE Tram, on Acoustics, Speech, and Signal Processing,* vol.36, pp. 1819- 1834, Dec. 1988.

[2] J.J. Shynk, "Adaptive IIR Filtering." *IEEE ASSP Magazine, pp.4 - 21*, April 1989.

[3] C.R. Johnson. Jr., "Adaptive IIR Filtering: Current Results and Open Issues," *IEEE Tram. on Information Theory,* vol.IT-30, pp.237-250, March 1984.

[4] F.F. Yassa, "Optimality in the Choice of the Convergence Factor for Gradient-Based Adaptive Algorithms," *IEEE Trans. Acoustics, Speech, and Signal Processing,* vol. ASSP-35, pp. 48-59, Jan. 1987.

[5] D. Parikh, N. Ahmed, and S.D. Stearns, "An Adaptive Lattice Algorithm for Recursive Filters," *IEEE Tram. Acoustics, Speech, and Signal Processing,* vol. ASSP-28, pp. 110-112, Feb. 1980.

[6] N.I. Cho, C.H. Choi, and S.U. Lee, "Adaptive Line Enhancement by Using an IIR lattice Notch Filter," *IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 37,* pp. 585-589, April 1989.

[7] I.L. Ayala, "On a New Adaptive Lattice Algorithm for Recursive Filters," *IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-30,* pp. 316-319, April 1982.

[8] T. Kwan and K.W. Martin, "Adaptive Detection and Enhancement of Multiple Sinusoids Using a Cascade IIR Filter," *IEEE Trans. on Circuits ad Systems,* vol. 36, pp.937-947, July 1989.

[9] Y.H. Tam, P.C. Ching, and Y.T. Chan, "Adaptive Recursive Filters in Cascade Form," *IEE Proc.,* vol. 134, Pt. F, Comm., Radar & Signal Processing, pp.245 252. June 1987.

[10] R.A. David, "A Modified Cascade Structure for IIR Adaptive Algorithms" *Proc. of 15th Asiiomar Conference on Circuits, Systems, and Computers,* pp. 175-179, Nov. 1981.

[11] R.A. David, "A Cascade Structure for Equation Error Minimization," *Proc. of 16th Asilomar Conference on Circuits, Systems, and Computers,* pp. 182-186, Nov. 1982.

[12] M. Nayeri and W.K. Jenkins, "Alternate Realizations to Adaptive IIR Filters and Properties of Their Performance Surfaces" *IEEE Trans.on Circuits and Systems,* vol. *CAS-36.* pp. 485-496, April 1989.

[13] J.J. Shynk, "Adaptive IIR Filtering Using Parallel-Form Realizations," *IEEE Trans. on Acoustics, Speech, ad Signal Processiq,* vol. *37,* pp. 519-533, April 1989.

[14] D.A. Johns, W.M. Snelgrove, and A.S. Sedra, "Adaptive Recursive State-Space Filters Using a Gradient Based Algorithm," *IEEE Trans.on Circuits and Systems,* vol. *37,* pp.673-684, June 1990.

[15] F.X.Y. Gao and W.M. Snelgrove, "An Efficient Adaptive Cascade IIR Filter," *Proc. of IEEE International Symposium on Circuits and Systems,* pp.444-447, June 1991.

Chapter Four

Adaptive Linearization Schemes for Weakly Nonlinear Systems

4.1 Introduction

System linearity is desired in many applications where nonlinearities exist. Some applications where linearization is necessary include

- Integrated continuous-time filters, where resistors are sometimes replaced by transistors [1] which suffer from substantial nonlinearity at large signal swings.
- Optical communication, where distortions caused by the noniinearities in the analog drive circuitry and LED or laser can be significant [2].
- Sound reproduction systems, where a loudspeaker has a few percent of nonlinear distortions [3-6,11,12].
- Digital microwave radio systems. where a critical issue in bandwidth-efficient QAM is the nonlinearity of the high-power amplifier in a satellite. Adaptive predistortion methods have been proposed to compensate for the nonlinear distortion [7,8].

There are some drawbacks to the existing linearization approaches. Most of the linearization methods for integrated continuous-time filters require device matching which can only be satisfied to a certain degree due to manufacturing fluctuations. The feedback technique has difficulties linearizing systems containing a lot of delay, such as air-path delay in a loudspeaker system. Most of the existing methods rely on fixed

circuits or devices, thus their performance will be degraded by aging, temperature, and an ever-changing environment.

Adaptive approaches may provide a good solution for some of the applications. Three new adaptive linearization schemes [9] and application of one of the schemes to a loudspeaker [10] are presented in this chapter. The three schemes are linearization by cancellation at the output, linearization with a post-processor (post-distortion), and linearization with a pre-processor (pre-distortion). Adaptive FIR filters are employed to furnish necessary estimates. The post-distortion scheme and the pre-distortion scheme are suitable for weakly nonlinear systems. The weaker the nonlinearities are, the more reduction in nonlinearity these two schemes can achieve. The scheme of linearization by cancellation at the output can be applied to problems with stronger nonlinearities and is able to give perfect nonlinear cancellation if the adaptive nonlinear filter produces a perfect estimate of the nonlinear part of the physical system. Each scheme may have applications where it is the preferred method.

As an application, linearization of a loudspeaker is investigated. A loudspeaker has nonlinearities which sometimes severely degrade the fidelity of the sound reproduced. The major nonlinearities in a loudspeaker include nonlinear suspension and non-uniform flux density [3-6,11,12]. The effect of suspension nonlinearity is proportional to the amplitude of the cone movement and thus can be reduced by some conventional techniques, such as a well designed vented baffle or a suitable horn. but at the cost of increasing size or limiting power. The distortion caused by the non-uniform flux density can be reduced by a careful design using conventional design techniques. All these considerations add extra constraints in design. The adaptive pre-distortion

approach proposed in this chapter may be used alternative to or in addition to the conventional design approaches and may result in a substantial reduction in nonlinear distortions or a gain in design flexibility or acceptable power levels.

4.2. Linearization by Cancellation at the Output

As discussed in Section 2.6, the Volterra series represents a nonlinear system by two subsystems: one purely linear and another purely nonlinear. This is described notationally by [1]

$$\begin{aligned} y_p(k) &= y_{Lp}(k) + y_{Np}(k) \\ &= [L_p(u)](k) + [N_p(u)](k) \end{aligned} \tag{4.1}$$

where y_{Lp} is the output of the linear subsystem with linear operator L_p, y_{Np} is the output of the purely nonlinear subsystem with nonlinear operator N_p, and $[L_p(u)](k)$ indicates an operation L_p on the sequence u evaluated at time k.

It is obvious that we can linearize a nonlinear system by subtracting an estimate of the output of the purely nonlinear subsystem from the output of the physical system. This estimate $N(u)$ can be obtained from an adaptive nonlinear filter. The adaptive linearization scheme is shown in Fig.4.1. This scheme is simple and effective.

However, for some applications, such as a loudspeaker system, it is hard to perform signal subtraction at the output side of a system. In some cases, it is desirable or necessary to pre-distort a signal at the input side of a system, while in other cases, post-distorting of signals may be required.

[1] In this thesis, whenever it is necessary to distinguish the variables of a physical system from those of an adaptive filter, the subscript p is used for the variables of the physical system.

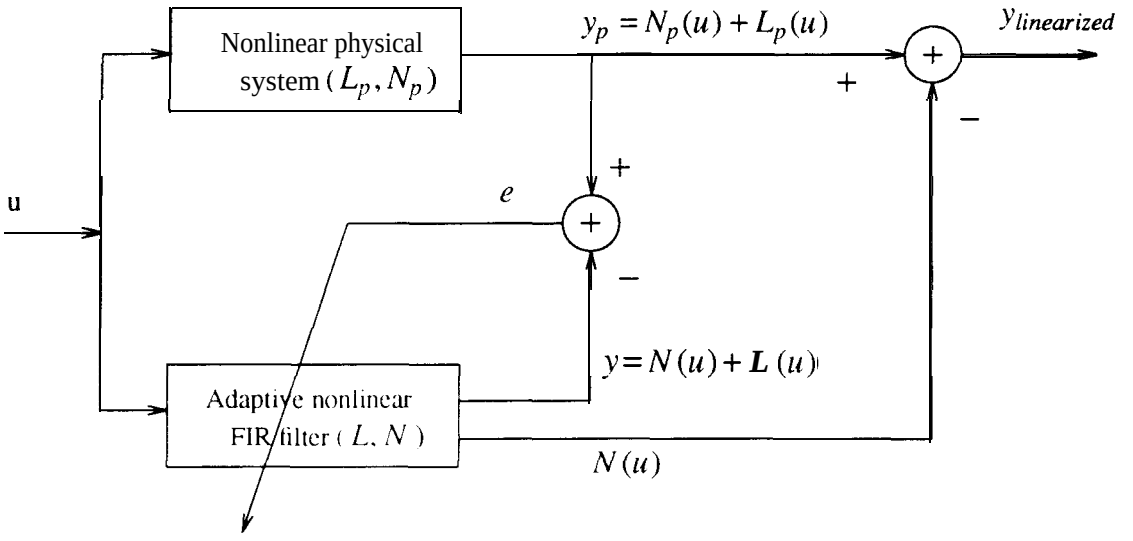
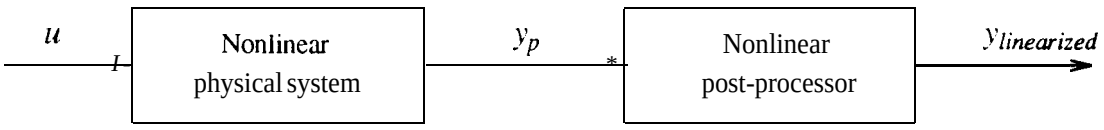


Fig.4.1 Adaptive linearization by canceling the effect of the nonlinearity at the output.

4.3. Linearization Using a Post-Processor

For some applications, it is preferred to post-distort signals. A post-processor can be applied to linearize such a system, as shown in Fig.4.2. One method is proposed here for a weakly nonlinear system.

In the following discussion, inverse modeling of the linear behavior of a nonlinear system will be used. Let L^{-1} indicate the linear operator obtained by an adaptive linear filter which performs inverse modeling of a physical system described by Equation (4.1). Then, we can have L^{-1}, satisfying



Fi.g.4.2 A nonlinear post-processor is placed at the output side of the nonlinear physical system to post-distort the signal.

$$L^{-1}L_p = z^{-\delta} \tag{4.2}$$

where $z^{-\delta}$ indicates a delay of δ samples and δ usually must be nonzero to allow a causal L^{-1}. If the nonlinearity of a physical system is weak, a post-processor with output

$$y(k) = y_p(k-\delta) - [N(L^{-1}(y_p))](k) \tag{4.3}$$

can reduce (though not eliminate) the nonlinear distortion, thus, linearizing the system. The notation N indicates an estimate of the nonlinear operator N_p of the physical system. We can verify this idea by some simple algebraic manipulations. The delayed output of the physical system can be written as

$$y_p(k-\delta) = [L_p(u)](k-\delta) + [N_p(u)](k-\delta)$$

Then the output of the nonlinear post-processor is

$$\begin{aligned} y(k) &= [L_p(u)](k-\delta) + [N_p(u)](k-\delta) - [N(L^{-1}(L_p(u) + N_p(u)))](k) \\ &= [L_p(u)](k-\delta) + [N_p(u)](k-\delta) - [N(z^{-\delta}(u) + L^{-1}(N_p(u)))](k) \end{aligned} \tag{4.4}$$

where Equation (4.2) is used. the operator $z^{-\delta}(u) = u(k-\delta)$. Assuming that the nonlinearity is weak, namely

$$|[L_p(u)](k)| \gg |[N_p(u)](k)|$$

we have

$$|u(k-\delta)| \gg |[L^{-1}([N_p(u))(k)|] \tag{4.5}$$

where Equation (4.2) is again employed. Then, assuming the nonlinear operator N is continuous, namely,

$$N(u+\varepsilon) \to N(u) \text{ for } \varepsilon \to 0$$

we have

$$\begin{aligned} y(k) &\doteq [L_p(u)](k-\delta) + [N_p(u)](k-\delta) - [N(u)](k-\delta) \\ &\doteq [L_p(u)](k-\delta) \end{aligned} \tag{4.6}$$

if the adaptive filter obtains a good estimate of N_p. The remaining nonlinearity in the output is of higher order and the output of the processor is the linearized output $y_{linearized}$.

The ratio of the linear signal to the residual nonlinear distortion can be estimated. Because of the assumption of weak nonlinearity in Equation (4.5), the post-processor's nonlinear part $[N(z^{-\delta}(u)+L^{-1}(N_p(u)))](k)$ can be approximated by

$$[N(z^{-\delta}(u\ j + L^{-1}(N_p(u)))](k) = [N(u)](k-\delta) + [N'(u)](k-\delta)[L^{-1}(N_p(u))](k) \tag{4.7}$$

where the nonlinear operator N' is the derivative of the nonlinear operator N and is defined by the following relation [13]

$$\lim_{x \to x_0} \frac{||N(x) - N(x_0) - N'(x_0)(x - x_0)||}{||x - x_0||} = 0$$

and x and x_0 are a variable and a point, respectively, in a Banach space. Considering Equation (4.7) and Equation (4.4), we can see the ratio of the linear signal to the residual nonlinear distortion is

$$\gamma \doteq \frac{\| [L_p(u)](k-\delta) \|_2}{\| [N'(u)](k-\delta)[L^{-1}(N_p(u))](k) \|_2} \tag{4.8}$$

To gain some insight, let us suppose the original nonlinear signal is weaker by α, where α is greater than one. Then, after linearization the ratio becomes

$$\gamma \doteq \alpha^2 \frac{\| [L_p(u)](k-\delta) \|_2}{\| [N'(u)](k-\delta)[L^{-1}(N_p(u))](k) \|_2} \tag{4.9}$$

This shows that if the original distortion is smaller by α, the ratio is increased by α^2 after linearization.

A simple example can help further illustrate the ideas. Suppose that the nonlinear physical system is described by a Taylor series, the memoryless case of a Volterra **series,**

$$y_p = 2u(k) + 0.06u^2(k) \tag{4.10}$$

The post-processor can be designed using Equation (4.3),

$$y_{linearized}(k) = y_p(k) - [N(L^{-1}(y_p))](k)$$

where the delay δ is zero due to memorylessness. Then,

$$\begin{aligned} y_{linearized}(k) &= 2u(k) + 0.06u^2(k) - 0.06(2^{-1}(2u(k) + 0.06u^2(k)))^2 \\ &= 2u(k) - 0.0036u^3(k) - 0.000054u^4(k) \\ &\doteq 2u(k) \end{aligned} \tag{4.11}$$

Assuming the norm of the input signal is unity, the original ratio of linear signal to distortion is

$$\gamma = \frac{2}{0.06}$$

After linearization, this ratio becomes

$$\gamma = \frac{2}{0.0036}$$

The signal-to-distortion ratio in dB is almost doubled by the linearization technique. As another check, we can also use Equation (4.8) to estimate this ratio after linearization. Considering $[N_p(u)](k) = 0.06u^2(k)$, $[N'(u)](\sim) \doteq [N_p'(u)](k) = 0.12u(k)$, $[L^{-1}(u)](k) \doteq [L_p^{-1}(u)](k) = 2^{-1}u(k)$, $[L_p(u)](k) = 2u(k)$, we have

$$\gamma = \frac{||2u(k)||_2}{||0.12u(k)2^{-1}(0.06u^2(k))||_2}$$

$$= \frac{2}{0.0036}$$

which is consistent with that obtained above.

To implement the linearization scheme, the operators L^{-1} and N are needed. Adaptive linear and nonlinear FIR filters can be used to provide these estimates. The adaptive implementation using adaptive FIR filters is shown in Fig.4.3. The adaptive nonlinear FIR filter models the "forward" behavior of the physical system and gives the operators $\boldsymbol{L}$ and N, which are estimates of L_p and N_p. The adaptive linear FIR filter models the "inverse" behavior of the linear part of the physical system and gives the operator L^{-1}, an estimate of L_p^{-1} with a difference of a delay operator. The input of the adaptive linear filter can be either the output of the physical system or the output of the linear subsystem of the adaptive nonlinear filter (see dashed lines in Fig.4.3).

The linear FIR filter of the processor is copied from the adaptive linear FIR filter, and the purely nonlinear FIR filter of the processor is a copy of the nonlinear operator N of the adaptive nonlinear filter. It is best to wait to perform the copying until the adaptive filters get reasonably good estimates. If the input of the adaptive linear filter is the

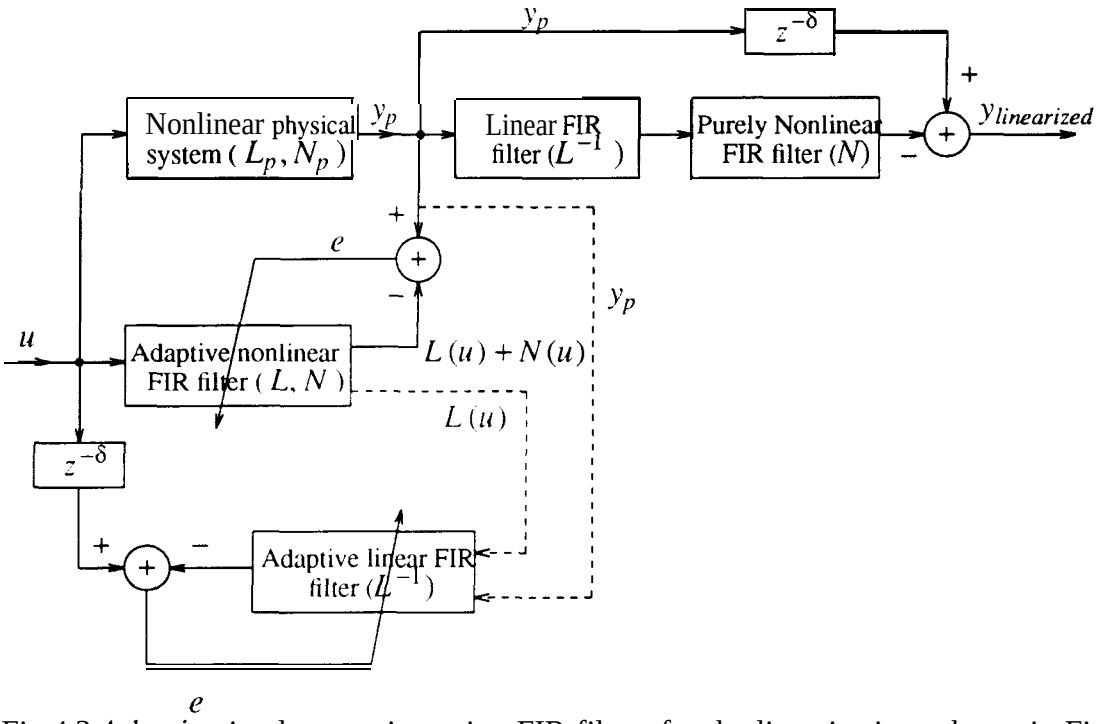


Fig.4.3 Adaptive implementation using FIR filters for the linearization scheme in Fig.4.2. Either one of the two dashed lines could be used. The linear FIR filter L^{-1} is copied from the adaptive linear FIR filter L^{-1}, and the nonlinear filter N is a copy of the nonlinear part of the adaptive nonlinear FIR filter.

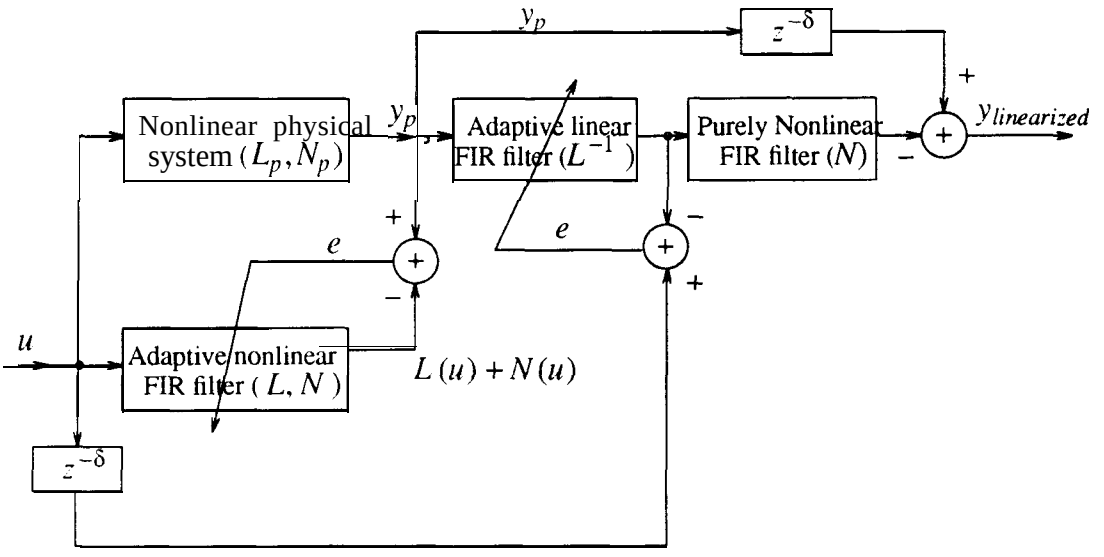


Fig.4.4 An efficient implementation of linearization scheme in Fig.4.3 when using the physical system output as the input of the adaptive linear inverse modeling filter L^{-1}.

output of the physical system, then Fig.4.3 can be easily modified to Fig.4.4 so that the adaptive linear filter can serve as the linear filter of the processor and computation can be reduced.

4.4. Linearization Using a Pre-Processor

For other linearization applications, a nonlinear processor is needed to *pre-distort* signals. as shown in Fig.4.5. A nonlinear processor with the following nonlinear mapping

$$y_i(k) = u(k-\delta) - L^{-1}(N(u)) \tag{4.12}$$

can perform the task. This can be verified easily. The output of the physical system is

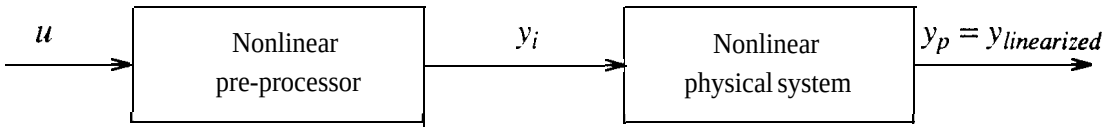


Fig.4.5 A nonlinear pre-processor is placed at the input side of the nonlinear physical system to pre-distort the signal.

$$\begin{aligned} y_p(k) &= [L_p(y_i)](k) + [N_p(y_i)](k) \\ &= [L_p(z^{-\delta}(u) - L^{-1}(N(u)))](k) + [N_p(z^{-\delta}(u) - L^{-1}(N(u)))](k) \\ &\doteq [L_p(u)](k-\delta) \end{aligned} \tag{4.13}$$

where Equations (4.2) and (4.5) are used. Hence, the output of the physical system is the linearized output. namely, $y_p = y_{linearized}$.

It can also be shown that the ratio of the linear signal to the residual distortion for the pre-distortion technique is about the same as that of the post-distortion technique:

$$\gamma = \frac{\| [L_p(u)](k-\delta)) \|_2}{\| [N_p'(u)](k-\delta)[L^{-1}(N(u))](k) \, \mathbf{I} \|_2} \tag{4.14}$$

This scheme can also be implemented using adaptive filters, as shown in Fig.4.6. The input of the adaptive linear filter is either the output of the physical system or the output of the linear subsystem of the adaptive nonlinear filter. The linear FIR filter is copied from the adaptive linear FIR filter and the purely nonlinear FIR filter is copied from the nonlinear part of the adaptive nonlinear FIR filter. As in the case of linearization using a post-processor, it is better to copy after the adaptive filters have run for some time and have good estimates.

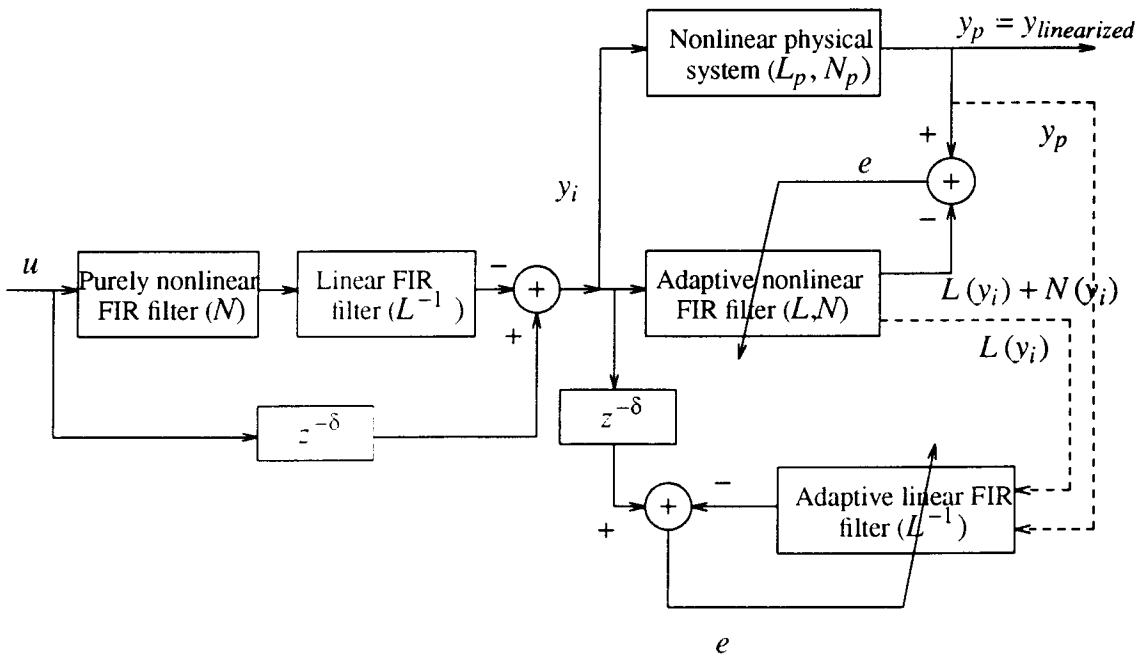


Fig.4.6 Adaptive implementation using FIR filters for the linearization scheme in Fig.4.5. Either one of the two dashed lines should be used. The linear FIR filter L^{-1} is copied from the adaptive linear FIR filter L^{-1}, and the nonlinear filter N is a copy of the nonlinear part of the adaptive nonlinear FIR filter.

4.5. Application: Linearization of a Loudspeaker

As discussed earlier, a loudspeaker sometimes has significant nonlinear distortions. For loudspeaker linearization, the scheme of linearization by cancellation at the output and the scheme with a post-processor are not suitable because these schemes require processing of sound signals after sound waves have left the loudspeaker. However, the scheme with a pre-processor handles signals in electrical form. Hence, it can be realized using the DSP technique. The pre-distortion scheme can be used and the general struc-

ture for this application is depicted in Fig.4.7. This section discusses a loudspeaker model and nonlinear distortions in a loudspeaker. The basic direct radiator loudspeaker is chosen for study due to its simplicity and popularity.

4.5.1 A Loudspeaker Model

A loudspeaker is composed of an electrical part and a mechanical part as shown in Fig.4.8. The electrical part is simply the voice coil. The mechanical part consists of the cone, the suspension, and the air load. The two parts interact through the magnetic field. The mechanical part can also be described by an equivalent electrical circuit, which will be called the mechanical circuit.

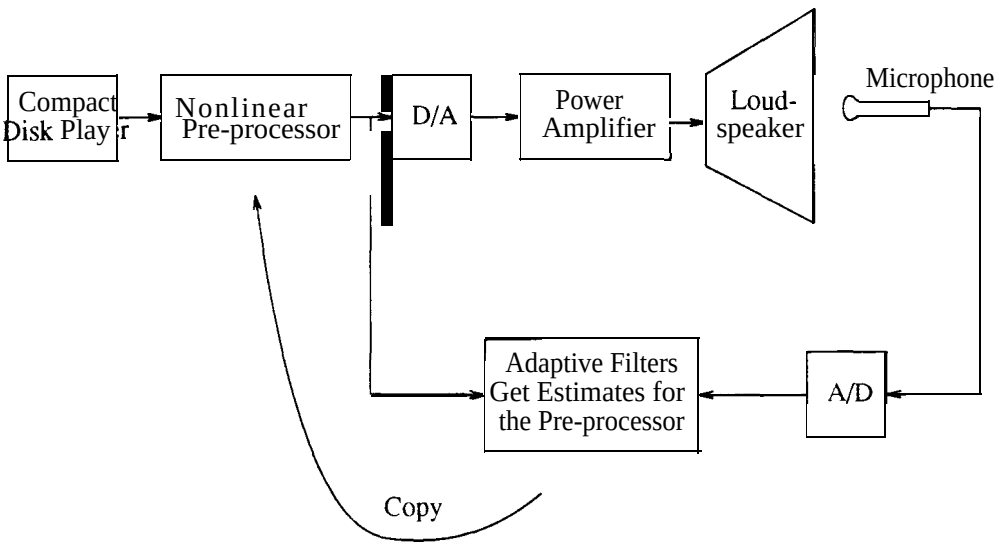


Fig. 4.7 Adaptive linearization of a loudspeaker using an adaptive nonlinear pre-processor.

The electrical circuit and the mechanical circuit of a loudspeaker are shown in Fig.4.9 [4,6]. In terms of analogies, the dimensions in the electrical circuit corresponding to length, mass, force and time in the mechanical system are charge, self-inductance, generator voltage, and time. Thus. we can write the differential equation for the mechanical circuit:

$$m\frac{d^2x}{dt^2} + r_M\frac{dx}{dt} + \frac{x}{C_M} = Bli \tag{4.15}$$

Referring to the electrical circuit shown in Fig.4.9, the following equation can be written:

$$e = ir + L\frac{di}{dt} + Bl\frac{dx}{dt} \tag{4.16}$$

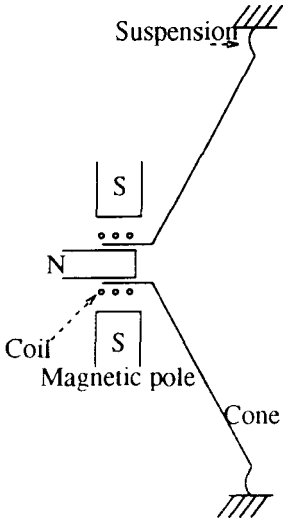
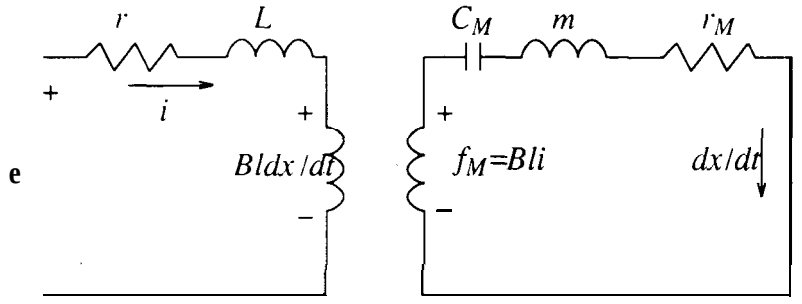


Fig.4.8 A conceptual structure of a basic loudspeaker.



Electrical circuit Mechanical circuit

Fig.4.9 Equivalent electrical and mechanical circuits of a loudspeaker. In the electrical circuit, e indicates the internal voltage of the generator, r represents the total electrical resistance of the generator and the voice coil, L is the inductance of the voice coil, i is the amplitude of the current in the voice coil, E is the voltage produced in the electrical circuit by the mechanical circuit and $E = Bldx/dt$, where B is the magnetic flux density in the air gap, 1 is the length of the voice coil conductor, and I is the cone displacement. In the mechanical circuit, m represents the total mass of the coil, the cone and the air load, r_M indicates the total mechanical resistance due to dissipation in the air load and the suspension system, C_M is the compliance of the suspension, and f_M is the force generated in the voice coil and is equal to Bli.

4.5.2 **Distortions in a Loudspeaker**

Generally, the force in the voice coil is a nonlinear function of displacement so that the compliance of the suspension system is a function of the displacement. The suspension nonlinearity affects distortion mainly at low frequencies. At frequencies of about 300 Hz or above, the total harmonic distortion of a loudspeaker is usually fairly low (of the order of 1%) and not appreciably affected by the suspension nonlinearity. As the frequency decreases, however. the distortion rises rapidly in loudspeakers having a suspension nonlinearity. For instance, a 10 inch dynamic loudspeaker with a nonlinear suspension has been measured to produce 10% total harmonic distortion with an input of 2 watts at 60 Hz [5]. The force deflection characteristic of the loudspeaker cone suspension system can be usually approximated by a polynomial

$$f = \alpha x + \beta x^2 + \gamma x^3 \tag{4.17}$$

Then, the compliance of the suspension system can be obtained

$$C_M = \frac{x}{f} = \frac{1}{\alpha + \beta x + \gamma x^2} \tag{4.18}$$

Substituting the above equation into Equation (4.15), we have

$$m\frac{d^2x}{dt^2} + r_M\frac{dx}{dt} + \alpha x + \beta x^2 + \gamma x^3 = Bli \tag{4.19}$$

Another source of harmonic distortion is non-uniform flux density up to the maximum amplitude of operation. The distortion caused by non-uniform flux density is small, usually less than 1%. as long as the amplitude of movement is small. However, the distortion is severe if the output signals are large. The flux density B is a function of the displacement x and may be approximated by a polynomial [12]

$$B(x) = B_0 + B_1 x + B_2 x^2 \tag{4.20}$$

This model can be confirmed using the measurement curve in [11]. The nonlinearity affects both the electrical circuit and the mechanical circuit, as suggested by Equations (4.15) and (4.16).

Then we substitute Equation (4.20) into (4.16) and (4.19) and discretize the two equations using the Euler approximation,

$$\left.\frac{dx}{dt}\right|_{t=k\tau} = (x(k+1) - x(k))/\tau$$

where τ is the sampling period and $x(k)$ is used to indicate $x(k\tau)$ for convenience. Letting $x_1 = i$, $x_2 = x$, and $x_3 = dx_2/dt$, we have the following difference equation in state-space form

$$\mathbf{x}(k+1) = \begin{bmatrix} a_{11} & 0 & a_{13} \\ 0 & 1 & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix} u(k)$$

$$+ \begin{bmatrix} p_{11} x_2(k) x_3(k) + p_{12} x_2^2(k) x_3(k) \\ 0 \\ p_{31} x_2^2(k) + p_{32} x_2^3(k) + p_{33} x_1(k) x_2(k) + p_{34} x_1(k) x_2^2(k) \end{bmatrix}$$

$$y(k) = (\,0\ 1\ 0\,)^T \mathbf{x}(k) \tag{4.21}$$

where the terms indicated by zero or unity are always zero or unity, $u = e$, $a_{11} = 1 - \tau r/L$, $a_{13} = -\tau l B_0/L$, $a_{23} = \tau$, $a_{31} = \tau B_0 l/m$, $a_{32} = -\tau\alpha/m$, $a_{33} = 1 - \tau r_M/m$, $b_1 = \tau/L$, $p_{11} = -\tau B_1 l/L$, $p_{12} = -\tau B_2 l/L$, $p_{31} = -\tau\beta/m$, $p_{32} = -\tau\gamma/m$, $p_{33} = \tau B_1 l/m$, and $p_{34} = \tau B_2 l/m$.

4.6. Numerical Examples

Numerical experiments on several different systems have been performed to test the adaptive linearization schemes and the results are presented in this section.

4.6.1 Physical Systems Described by Volterra Series

In the following tests, a physical system was modeled by a Volterra series with a linear term, a quadratic term and a cubic term. The adaptive nonlinear filter had the same orders as the physical system, and initially, all the coefficients of the adaptive filters were set to zero. For the scheme of linearization by cancellation at the output, the output of the nonlinear part of the physical system was the original distortion. The residual distortion after linearization was the difference between the output of the linear part of the physical system and the linearized output. To measure the residual distortion for the other two schemes, we have used a reference system which was a copy of the linear part of the physical system. Its input was $u(k-\delta)$, a delayed version of the original input signal since the linearized signal was delayed by this amount in these two schemes. The residual distortion was measured as the difference between the output of the reference system and the linearized signal.

Test I

The physical system had orders $n_{p1}=10$, $n_{p2}=3$, and $n_{p3}=2$. The linear part was

$$y_{Lp}(k)=0.2u(k)+0.5u(k-1)+0.3u(k-2)+1.2u(k-3)+0.7u(k-4)+0.05u(k-5)$$
$$+0.01u(k-6)+0.01u(k-7)+0.01u(k-8)-0.008\sim(k-9)-0.005\sim(k-10)$$

The quadratic term was

$$y_{quadratic}(k) = 0.01u^2(k) - 0.001u(k)u\ (\text{k-l})\ - 0.001u(k)u(k-2) + 0.008u(k)u(k-3)$$
$$+ 0.011u^2(k-1) + 0.003u(k-1)u(k-2) + 0.001u(k-1)u(k-3)$$
$$+ 0.009u^2(k\text{ -2}) + 0.002u(k-2)u(k\text{ -3}) + 0.008u^2(k\text{ -3})$$

and the cubic term was

$$y_{cubic} = 0.005u^3(k) + 0.003u^2(k)u(k-1) - 0.005u^2(k)u(k-2) + 0.009u(k)u^2(k-1)$$
$$- 0.006\sim(k)u(k-1)u(k\text{ -2}) - 0.007u(k)u^2(k\text{ -2}) + 0.008u^3(k-1)$$
$$- 0.001u^2(k-1)u(k-2) + 0.002u(k-1)u^2(k-2) + 0.001u^3(k-2)$$

The coefficients were chosen so that the nonlinear component of the output was a few percent of the linear component. The mean square (MS) value of original nonlinear distortion of the physical system was $-24.1dB$. The MS value of the linear signal, namely, the non-distorted signal, was $2.9dB$. The order of the adaptive linear filter was chosen as $n = 50$. The step sizes were 0.005 for h_1 of the adaptive nonlinear filter, and 0.001 for h_2, h_3 of the adaptive nonlinear filter and h of the adaptive linear filter.

The reduction in distortion versus the number of iterations is shown in Fig.4.10 for the scheme of linearization by cancellation at the output. This curve shows that at $6k$ iterations the distortion was reduced to $-107dB$, and after $20k$ iterations the distortion was reduced to $-290dB$ (the computational noise floor), which is essentially perfect. In this case, subtraction of the nonlinear output was performed starting at time zero. The output distortion was actually worse than the original distortion for the first $1k$ iterations due to transients. In some applications. it is necessary to avoid this, and so this subtraction should be performed after the adaptive filter gets better estimates. The performance of the adaptive nonlinear filter can also be deduced from this figure since the curve of the MS error for the forward identification of the physical system had a difference of just a

few dB with the curve shown in Fig.4.10.

The results for Test 1 are tabulated in Table 4.1. In this test the scheme with a non-linear post-processor and the scheme with a pre-processor reduced the distortions to $-53.1dB$ and $-46.6dB$, respectively, from $-24.1dB$. These two techniques exhibited small distortion residuals and they doubled the linear signal to distortion ratio in dB.

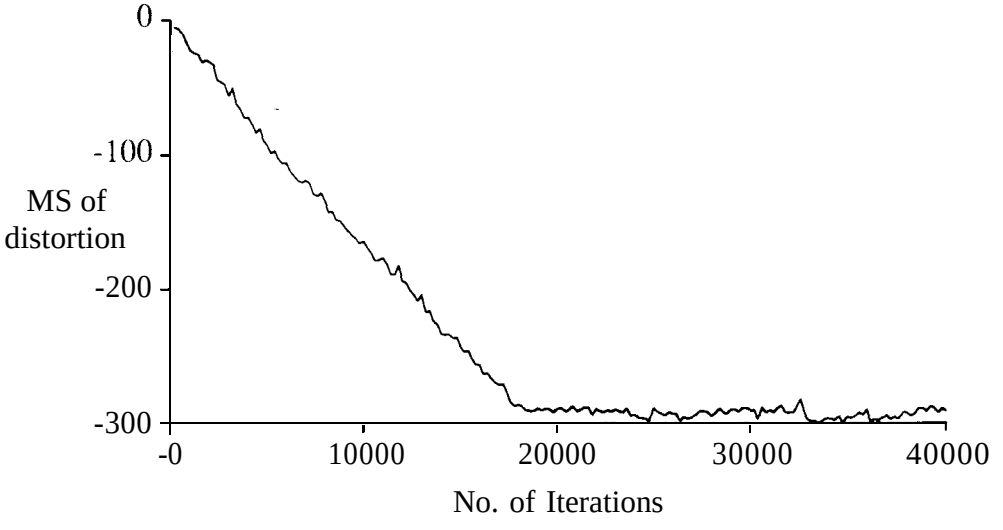


Fig.4.10 Reduction in distortion for linearization by cancellation at the output in Test 1.

Test 2

All the conditions in both Test 1 and Test 2 were the same, except for the nonlinear parts. The nonlinear coefficients of the physical system in Test 2 were chosen so that there was a lower distortion in the original physical system than that of Test 1. The results are summarized in Table 4.1. At $40k$ iterations, the distortion was reduced to the computational noise floor for the scheme of linearization by cancellation at the output, to $-67dB$ for the scheme with a post-processor, and to $-59dB$ for the scheme with a pre-processor from the original distortion of $-30dB$. The distortion reductions in Test 2 for the schemes with a pre-processor and a post-processor were larger than those in Test 1 since the original distortion in Test 2 was smaller, satisfying the assumption of weak nonlinearity better.

In all tests, no significant differences in the results were observed whether the output of the physical system or the output of the linear part of the adaptive nonlinear filter was used as the input signal to the adaptive linear inverse filter.

Table 4.1 Results for Tests 1 and 2

Test	Initial distortion	cancel at output	Post-distort	Pre-distort
1	-24.1 dB	CNF	-53.1 dB	-46.6 dB
2	-30 dB	CNF	-67 dB	-59 dB

CNF: computational noise floor.

4.6.2 **Loudspeaker Linearization**

This section presents simulation results for the pre-distortion scheme on the loudspeaker model in Equation (4.21). The loudspeaker in the simulation had the following parameters

$$\mathbf{x}(k+1) = \begin{bmatrix} -0.1 & 0 & -0.2 \\ 0 & 1 & 1 \\ 0.6 & -0.5 & -0.15 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 0.4 \\ 0 \\ 0 \end{bmatrix} u(k)$$

$$+ \begin{bmatrix} -0.04x_2x_3 - 0.05x_2^2x_3 \\ 0 \\ -0.08x_2^3(k) + 0.01x_1(k)x_2(k) + 0.02x_1(k)x_2^2(k) \end{bmatrix}$$

$$y(k) = (\,0 \;\; 1 \;\; 0\,)^T \mathbf{x}(k)$$

where the sample period τ was set to be unity and the parameter β was chosen as zero, as in [4], since β is very small **in** practice

A reference linear filter having the linear parts of the loudspeaker model was used. The mean square values of the linear part and nonlinear part of the output signal were $-10.0dB$ and $-39.4dB$, respectively. The orders of the forward-modeling adaptive nonlinear FIR filter were $n_1 = 17, n_2 = 10, n_3 = 10$, the step sizes were $\mu_1 = 0.01$, $\mu_2 = 0.0001$, and $\mu_3 = 0.000\,1$ f‘or the nonlinear filter. The order of the inverse modeling linear filter was 6, and the step size was $\mu = 0.07$. The delay δ was chosen to be 3 and the input signal of the reference linear filter was delayed by this amount. The initial coefficients of the adaptive filters were set to zero.

At $80k$ iterations, MSE for inverse modeling by the linear filter was $-34dB$, which could not be reduced further by a linear filter due to the existence of nonlinearity in the system. MSE for forward-identification by the nonlinear filter was $-67.6dB$ after $80k$

iterations. After the linearization took effect at $80k$ iterations, the nonlinear distortion was reduced from the original value of $-39.4dB$ to $-66dB$, that is, 5% of the original distortion. In other words, the ratio of the linear signal to the nonlinear distortion was increased to $56dB$ from $29.4dB$, nearly doubled.

4.7. Summary

Three new adaptive linearization schemes have been developed. The schemes are attractive in that the resultant systems are not complicated, making them easy to implement in both hardware and software. The post-distortion scheme and the pre-distortion scheme are designed for weakly nonlinear systems and the weaker the nonlinearities are, the greater reduction in nonlinearity they can achieve. The scheme of linearization by cancellation at the output can be applied to problems with stronger nonlinearities and can achieve perfect nonlinear cancellation if the adaptive nonlinear filter produces a perfect estimate of the nonlinear part of the physical system. These methods may find applications in acoustical systems. communications systems, etc. The pre-distortion scheme was proposed to linearize a loudspeaker. Simulations on a mathematical loudspeaker model have shown its promise.

References

[1] Y. Tsividis, "Continuous-Time MOSFET-C Filters in VLSI," *IEEE Journal on Solid-State Circuits,* vol.SC-21, pp. 15-29, Feb. 1986.

[2] H. Kressel (eds.), *Semiconductor Devices for Optical Communication, (Topics in Applied Physics vol. 39), New* York: Springer-Vet-lag, 1980.

[3] K.B. Benson (eds.), *Audio Engineering Handbook,* Toronto: McGraw-Hill Book Company, 1988.

[4] H.F. Olson, *Acoustical Engineering,* Toronto: D. Van Nostrand Company, Inc., 1964.

[5] F. Langford-smith, *Radiotron Designer's Handbook,* Wireless Press, 1953.

[6] M. Rossi, *Acoustics and Electroacoustics,* Norwood, Massachusetts: Artech House, Inc., 1988.

[7] A.A.M. Saleh and J. Salz, "Adaptive Linearization of Power Amplifiers in Digital Radio Systems," *Bell System Technical J.,* vol. 62, pp.1019-1033, April 1983.

[8] G. Karam and H. Sari, "Analysis of Predistortion, Equalization, and IS1 Cancellation Techniques in Digital Radio Systems with Nonlinear Transmit Amplifiers," *IEEE Tram. on Communications, vol. 37,* pp. 1245-1253, Dec. 1989.

[9] F.X.Y. Gao and W.M. Snelgrove, "Adaptive Linearization Schemes for Weakly Nonlinear Systems Using Adaptive Linear and Nonlinear FIR Filters," *Proc. of 33rd Midwest Symposium on Circuits and Systems,* Calgary, 1990.

[10] F.X.Y. Gao and W.M. Snelgrove, "Adaptive Linearization of A Loudspeaker," *Proc. of International Conference on Acoustics, Speech, and Signal Processing,* pp.3589-3592, May 199 1.

[11] M.H. Knudsen, J.G. Jensen, V. Julskjaer, and P.Rubak, "Determination of Loudspeaker Drive Parameters Using a System Identification Technique," *J. Audio Engineering Society,* Vol.37, pp.700-708, Sept. 1989.

[12] W. Klippel, "Dynamic Measurement and Interpretation of the Nonlinear Parameters of Electrodynamic Loudspeakers," *J. Audio Engineering Society, Vol.38,* pp.944-955, Dec. 1990.

[13] J. Dieudonne, *Foundations of Modern Analysis, New* York: Academic Press, 1969.

Chapter Five

Adaptive Nonlinear Recursive State-Space Filters

5.1 Introduction

Adaptive nonlinear filters previously reported are often, directly or indirectly, based on Volterra theory and have finite impulse responses, as discussed in Chapter Two. They can be considered as extensions of adaptive linear FIR transversal filters to nonlinear problems. Adaptive nonlinear FIR filters share advantages and disadvantages with adaptive linear FIR filters. The problem of computation cost in the case of adaptive nonlinear FIR filters is much more serious than that in the case of adaptive linear FIR filters since their cost increases superlinearly, rather than linearly, with system memory length.

Adaptive linear IIR filters have aroused some interest, e.g. [1-5], due to their potential advantage in computation over linear FIR filters. Very few results have been reported on adaptive nonlinear IIR filters in the context of signal processing. An adaptive nonlinear IIR filter was presented in [6] using the Volterra series with a bilinear structure. Adaptive nonlinear recursive state-space (ANRSS) filters were first introduced in [7] and are presented in this chapter. They are more general in form than the adaptive Volterra filter with a bilinear structure in [6] and are expected to alleviate the problem of high computational cost of adaptive nonlinear FIR filters in long-memory applications. Since the physics of the nonlinear system concerned is often known in

practice and our understanding of the system can be improved by some identification methods which give such important information as an estimate of the order and significance of a term, the structure of the nonlinear system can be assumed to be known. Then, the adaptive filter only has to adapt parameters which are not exactly known or which drift with time.

In this chapter, after introducing ANRSS filters, efficient gradient computation algorithms are developed to improve their efficiency. The stability, the convergence performance, and the potential applications of the ANRSS filters are investigated. Finally, simulation results are presented.

5.2. FIR Volterra Filters Are Computationally Expensive

The computational disadvantage of adaptive nonlinear FIR filters reported in the literature can be easily shown by numerical experiments on a simple example. Consider a nonlinear first-order physical system, with quadratic nonlinearity, which is described by

$$x_p(k+1) = a_p x_p(k) + b_p u(k) + p_{p1} x_p^2(k) + p_{p2} u(k) x_p(k)$$
$$y_p(k) = c_p x_p(k) \tag{5.1}$$

where x_p and y_p are the state variable and the output variable. respectively. An adaptive nonlinear FIR filter was used to match the input-output relationship of this system. The input signal was a white Gaussian signal with unit variance.

The physical system used in the simulations was

$$x_p(k+1) = -0.9x_p(k) + 0.8u(k) + 0.01x_p^2(k) + 0.03u(k)x_p(k)$$

$$y_p(k) = x_p(k) \tag{5.2}$$

The ratio of the nonlinear component to the linear component of the system output is $-21dB$. The parameters were chosen so that the impulse response is not very long and the nonlinearity is not very strong hence the problem is not very tough for an adaptive Volterra filter.

Four tests of adaptive FIR filters were performed on this example and the results are presented in Table 5.1. In all tests, step sizes were $\mu_1 = 10^{-3}$ for the linear term, $\mu_2 = 10^{-4}$ for the quadratic term (if the adaptive filter had one), and $\mu_3 = 10^{-5}$ for the cubic term (if any). In the first test. the nonlinear coefficients, p_{p1} and p_{p2}, of the reference system were set to zero. An adaptive linear filter of order 70 showed an MSE of $-60dB$ after $5k$ iterations. Order $n_1 = 70$ was then used for the linear term of the Volterra filter for the rest tests.

In all the following tests, the physical system was the one described in Equation (5.2). In the second test, the adaptive filter was still linear, with the same order as in the first test. The adaptive linear filter could achieve a residual MSE of only $-16dB$, the noise floor due to the nonlinearity of the physical system. In the third test, a quadratic term with $n_2 = 50$ was added to the adaptive filter, resulting in a reduction of $21.7dB$ in MSE over Test 2. This means that about ninety percent of the nonlinearity of the physical system has been modeled by the quadratic term. In the fourth test, a cubic term with $n_3 = 10$ was also added to the adaptive filter, resulting in a further reduction in MSE by $0.5dB$. This reduction was very small compared to that obtained by adding the quadratic term since the quadratic term is the dominant nonlinear term for this physical system.

In Test 4, the adaptive filter performed $7.84k$ multiplications per sample, which is computationally demanding. If a lower residual is desired and the physical system has a longer impulse response and a stronger nonlinearity, the adaptive nonlinear filter should consist of longer and higher power terms. It is challenging to implement such an adaptive filter on a single IC chip. For example, a Motorola 56001 chip can perform 10.25 million multiplications per second. It is able to perform 232 multiplications per sample at the 44.1 kHz audio rate, or 128 1 multiplications per sample at the 8 kHz speech rate. Therefore, such a single chip cannot handle the computation load required by the adaptive nonlinear FIR filter in this example (tests 3 and 4) even at the speech rate.

It is well known that adaptive linear IIR filters have a potential computational advantage over adaptive linear FIR filters, which has sparked active research on adaptive linear IIR filters. Similarly, we can expect that an adaptive nonlinear IIR filter is

Table 5.1 Numerical Results for the First-Order Example Using FIR Filters

Test	MSE(dB)	Multiplication/iter	Iterations
1	-60	0.14k	5k
2	-16	0.14k	20k
3	-37.7	5.65k	80k
4	-38.2	7.84k	80k

Test 1: Both the physical system and adaptive filter were linear.
Test 2: The physical system was nonlinear and the adaptive filter was linear.
Test 3: The physical system was nonlinex and the adaptive filter was nonlinear with linear and quadratic terms.
Test 4: The physical system was nonlinex and the adaptive filter was nonlinear with linear. quadratic, and cubic terms.

potentially more economical than an adaptive nonlinear FIR filter. A recursive state-space structure is a quite general nonlinear IIR structure. An ANRSS filter should be easily implemented on a single chip for some applications. To make comparison between an adaptive nonlinear FIR filter and an ANRSS filter, some numerical tests, corresponding to those in Table 5.1, have been performed and the results will be presented in Section 5.7. These results indicate that for the first-order example, an ANRSS filter is able to match the reference physical system perfectly, with 0.4% of the computation required by the adaptive nonlinear FIR filter per iteration and with 6% of its convergence time.

Another motivation of introducing the nonlinear recursive state-space structure is its suitability for analog implementation. Although the ANRSS filters are presented in the digital domain in this chapter. they are also applicable in the continuous-time domain. A programmable *linear* recursive state-space filter has been implemented in analog technology [5]. Analog implementation of the ANRSS filters proposed in this chapter can be very similar to that of an adaptive linear recursive state-space filter.

5.3. Filter Formulation and Gradient Computation

The physics of many practical systems is known and can often be described by a state-space equation. The system parameters are unknown or slowly time-varying and this makes an adaptive filter necessary in certain applications. Suppose a physical system is described by a nonlinear recursive state-space equation of order n_p:

$$\mathbf{x}_p(k+1) = \mathbf{A}_p\mathbf{x}_p(k) + \mathbf{B}_p u(k) + \mathbf{g}_p(\mathbf{p}_p, u(k), \mathbf{x}_p(k)) \tag{5.3.a}$$

$$y_p(k) = \mathbf{C}_p^T\mathbf{x}_p(k) + d_p u(k) \tag{5.3.b}$$

where $\mathbf{A}_p$ is the system feedback matrix, $\mathbf{B}_p$ is the system input coefficient vector, $\mathbf{x}_p$ is the state vector, $\mathbf{g}_p$ is a nonlinear function, and $\mathbf{p}_p$ is a vector of coefficients for the nonlinearity. The order of the system is assumed to be known. The exact values of $\mathbf{A}_p, \mathbf{B}_p$, $\mathbf{C}_p$, and $\mathbf{p}_p$ are not known. The right hand side of Equation (5.3a) can be considered as a truncated Taylor Series expansion of a general function $\mathbf{f}_p(\mathbf{x}_p, u)$ at $\mathbf{x}_p = 0$ and $u = 0$. The nonlinear function $\mathbf{g}_p$ is thus a truncated multi-dimensional Taylor series without linear terms and its coefficients are the elements of $\mathbf{p}_p$. Following is a second-order example with the Taylor series truncated to quadratic terms:

$$\mathbf{g}_p = \begin{bmatrix} p_{p1}x_{p1}(k)x_{p2}(k) + p_{p2}x_{p1}(k)u(k) + p_{p3}x_{p2}(k)u(k) + p_{p4}u^2(k) + p_{p5}x_{p1}^2(k) + p_{p6}x_{p2}^2(k) \\ p_{p7}x_{p1}(k)x_{p2}(k) + p_{p8}x_{p1}(k)u(k) + p_{p9}x_{p2}(k)u(k) + p_{p10}u^2(k) + p_{p11}x_{p1}^2(k) + p_{p12}x_{p2}^2(k) \end{bmatrix}$$

The nonlinear function $\mathbf{g}_p$ is differentiable with respect to both $\mathbf{x}_p$ and $\mathbf{p}_p$. The point $\mathbf{x}_p = 0$ and $u = 0$ is the equilibrium point since all the terms, $\mathbf{A}_p\mathbf{x}_p, \mathbf{B}_p u$, and $\mathbf{g}_p$, on the right-hand side of Equation (5.3a) are equal to zero at this point.

An ANRSS filter employs the structure of the physical system in Equation (5.3) and adapts its coefficients $\mathbf{A}, \mathbf{B}, \mathbf{C}, d$, and $\mathbf{p}$ to minimize the mean square (MS) of the

difference between its output and a desired signal. The well-known LMS algorithm will be used to update the coefficients. The filter coefficient vector is updated according to

$$\mathbf{w}^{k+1} = \mathbf{w}^{k} + 2\mu e(k)\frac{\partial y(k)}{\partial \mathbf{w}} \tag{5.4}$$

where the vector w includes all the coefficients to be adapted. In this algorithm, the gradient of each parameter to be updated should be available.

The gradients of the adaptive filter output with respect to an element of C and the feedthrough coefficient d can be easily written as

$$\frac{\partial y(k)}{\partial c_i} = x_i(k), \quad \frac{\partial y(k)}{\partial d} = u(k)$$

where x_i is the *ith* element of the state vector **x**.

If the gradients of the state vector x with respect to the elements of A, **B**, and **p** are defined as

$$\mathbf{F}_{ij}(k) = \frac{\partial \mathbf{x}(k)}{\partial a_{ij}}, \quad \mathbf{Q}_i(k) = \frac{\partial \mathbf{x}(k)}{\partial b_i}, \quad \text{and} \quad \mathbf{H}_i(k) = \frac{\partial \mathbf{x}(k)}{\partial p_i},$$

where a_{ij} is the element on the *ith* row and the *jth* column of the A matrix, b_i and p_i are the *ith* elements of B and **p**, respectively, it can be shown that the gradients of the adaptive filter output with respect to these filter coefficients can be written as

$$\frac{\partial y(k)}{\partial a_{ij}} = \mathbf{C}^T\mathbf{F}_{ij}(k), \quad \frac{\partial y(k)}{\partial b_i} = \mathbf{C}^T\mathbf{Q}_i(k), \quad \text{and} \quad \frac{\partial y(k)}{\partial p_i} = \mathbf{C}^T\mathbf{H}_i(k) \tag{5.5}$$

where $\mathbf{F}_{ij}(k), \mathbf{Q}_i(k)$ and $\mathbf{H}_i$ are computed recursively from the following equations:

$$\mathbf{F}_{ij}(k+1) = \mathbf{A}\mathbf{F}_{ij}(k) + \mathbf{e}_i x_j(k) + \left(\frac{\partial \mathbf{g}(\mathbf{p}, u(k), \mathbf{x}(k))}{\partial \mathbf{x}(k)}\right)\mathbf{F}_{ij}(k) \tag{5.6}$$

$$\mathbf{Q}_i(k+1) = \text{AQ;(k)} + \mathbf{e}_i u(k) + \left(\frac{\partial \mathbf{g}(\mathbf{p}, u(k), \mathbf{x}(k))}{\partial \mathbf{x}(k)}\right)\mathbf{Q}_i(k) \tag{5.7}$$

$$\mathbf{H}_i(k+1) = \mathbf{A}\mathbf{H}_i(k) + \left(\frac{\partial \mathbf{g}(\mathbf{p}, u(k), \mathbf{x}(k))}{\partial \mathbf{x}(k)}\right)\mathbf{H}_i(k) + \frac{\partial \mathbf{g}(\mathbf{p}, u(k), \mathbf{x}(k))}{\partial p_i} \tag{5.8}$$

where $\mathbf{e}_i$ is a vector with unity in the *ith* element and zero in others. Comparing Equations (5.5), (5.6), (5.7), and (5.8) with Equation (5.3), it is seen that the gradients are computed with systems very similar in structure to the adaptive filter itself.

5.4. Reductions in Gradient Computation

From the above discussion. we know that one gradient filter with complexity similar to that of the adaptive filter itself is needed to adapt each element of A, **B**, or p. This demands a substantial amount of computation. Two methods of reducing the computation will be discussed in this section.

5.4.1 Keeping the Input Coefficient **Vector Fixed**

The computation can be reduced if adapting **B** can be avoided. There is a way to do so if it is known which terms of $\mathbf{B}_p$ are zero and which are not, and the differences between $\mathbf{B}_p$ and B just result in scaled states and coefficients. This idea is best explained with an example. Suppose the physical system concerned is a second-order system described by

$$x_{p1}(k+1) = a_{p11}x_{p1}(k) + a_{p12}x_{p2}(k) + b_{p1}u(k) + p_{p1}x_{p1}^3(k)x_{p2}^2(k) \tag{5.9}$$

$$x_{p2}(k+1) = a_{p21}x_{p1}(k) + a_{p22}x_{p2}(k) + b_{p2}u(k) + p_{p2}u(k)x_{p1}^2(k) \tag{5.10}$$

$$y_p = c_{p1}x_{p1}(k) + c_{p2}x_{p2}(k) + d_p u(k) \tag{5.11}$$

where all the coefficients are unknown. Let us first assume that both elements of the input coefficient vector $\mathbf{B}_p$ are nonzero. For given $\mathbf{B}_p$ and **B**, there exist two nonzero scalars α_1 and α_2, which relate $\mathbf{B}_p$ and **B:**

$$b_1 = \alpha_1 b_{p1}, \quad b_2 = \alpha_2 b_{p2} \tag{5.12}$$

Next, multiplying both sides of Equation (5.9) by α_1 and both sides of Equation (5.10) by α_2 and performing some simple algebraic manipulations, we arrive at

$$x_1(k+1) = a_{11}x_1(k) + a_{12}x_2(k) + b_1u(k) + p_1x_1^3(k)x_2^2(k) \tag{5.13}$$

$$x_2(k+1) = a_{21}x_1(k) + a_{22}x_2(k) + b_2u(k) + p_2u(k)x_1(k) \tag{5.14}$$

$$y = c_1x_1(k) + c_2x_2(k) + du(k) \tag{5.15}$$

where

$$x_i = \alpha_i x_{pi}, \quad c_i = \frac{c_{pi}}{\alpha_i}, \quad \text{for } i = 1 \text{ and } 2, \quad a_{11} = a_{p11}, \quad a_{12} = a_{p12}\frac{\alpha_1}{\alpha_2}, \tag{5.16}$$

$$a_{22} = a_{p22}, \quad a_{21} = a_{p21}\frac{\alpha_2}{\alpha_1}, \quad d = d_p, \quad p_1 = \frac{p_{p1}}{\alpha_1^2\alpha_2^2}, \quad p_2 = p_{p2}\frac{\alpha_2}{\alpha_1^2} \tag{5.17}$$

The new system described by Equations (5.13), (5.14), and (5.15) is obtained by scaling the original system. This scaling maintains the structure of the original system. From Equations (5.16) and (5.17), we know that the ***a's*** must be nonzero. Hence, if some elements of $\mathbf{B}_p$ **are** zero (or nonzero), the corresponding elements of **B** must also be zero (or nonzero), as suggested by Equation (5.12).

Therefore, to use an adaptive nonlinear filter to match a physical system described by Equations (5.9), (5.10), and (5.11), we can set the input coefficient vector of the adaptive filter to be a constant vector with the same zero-nonzero pattern as that of the physical system, and the adaptive filter can just adapt the feedback matrix **A,** the output coefficient vector **C**, the feedthrough coefficient d and the nonlinear coefficients **p** to match the physical system, resulting in a scaled system model. The n gradient filters for the input coefficient vector **B are** then not needed. It can be shown that this is generally true for the case where the nonlinear function vector $\mathbf{g}_p(\mathbf{p}_p, u(k), \mathbf{x}_p(k))$ is an n-

dimensional Taylor series without linear terms. A direct-form equation is an example where the zero-nonzero pattern of the input coefficient vector is known.

In practice, **B** would be set to estimates from the physics of the nonlinear system. This can not only provide a good starting point, but also avoid some numerical difficulties arising from scaling. Further, adapting **B** and C simultaneously would result in difficulties in convergence due to redundant degrees of freedom.

5.4.2 **The Approximate Stochastic-Gradient Method**

The technique of gradient approximation has been widely and successfully applied in many practical optimization problems. This technique can be applied to the ANRSS filters to reduce the computation. If the system is weakly nonlinear (the magnitude of the signal from the nonlinear part $\mathbf{g}_p$ of the physical system is much smaller than that from the linear part $\mathbf{A}_p\mathbf{x}_p(k)+\mathbf{B}_p u(k)$), **we** can compute the approximate gradients for filter coefficients by neglecting the nonlinear part, thus considering the gradient filters as linear.

When neglecting the nonlinearity, the gradients for **A** and **B** of an ANRSS filter can be computed like those of an adaptive linear recursive state-space filter [4,5]

$$\mathbf{F}_j(k+1) = \mathbf{A}^T\mathbf{F}_j(k) + \mathbf{C}x_j(k) \tag{5.18}$$

and

$$\mathbf{Q}(k+1) = \mathbf{A}^T\mathbf{Q}(k) + \mathbf{C}u(k) \tag{5.19}$$

where $q_i(k)=\partial y(k)/\partial b_i$ and the ith element of $\mathbf{F}_j(k)=\partial y(k)/\partial a_{ij}$. One gradient filter is able to generate gradients for all the elements of one column of matrix **A,** and one gradient filter for all the elements of **B**, resulting in a significant reduction in the computa-

tion. Evaluation of the gradient for B is also discussed here since it may sometimes be necessary to adapt B. The approximate gradient for $\mathbf{p}$ can be computed from

$$\mathbf{H}_i(k+1) = \mathbf{A}\mathbf{H}_i(k) + \frac{\partial \mathbf{g}(\mathbf{p}, u(k), \mathbf{x}(k))}{\partial p_i} \tag{5.20}$$

and

$$\frac{\partial y(k)}{\partial p_i} = \mathbf{C}^T \mathbf{H}_i(k) \tag{5.21}$$

where

$$\mathbf{H}_i(k) = \frac{\partial \mathbf{x}(k)}{\partial p_i}.$$

Although the nonlinearity is ignored when computing the approximate gradients, it is still used for computing the adaptive filter output. If exact gradients for all coefficients adapted are used. the method will be referred to as the stochastic-gradient method. On the other hand, if approximate gradients for all coefficients other than C are used, it will be referred to as the approximate stochastic-gradient method. As for C, the exact gradient is easily available and therefore always used. The approximate stochastic-gradient method is only suitable for the case of weak nonlinearity, say, the ratio of nonlinearity versus linearity being 10%. If the nonlinearity is strong, the performance of the approximate-gradient method will be degraded and the stochastic-gradient method should be used instead.

5.5. Stability Considerations and Convergence Analysis

As is well known. stability and convergence analysis are challenging problems for adaptive linear IIR filters. The issues are even more challenging for adaptive nonlinear

IIR filters. In this section. these two issues are addressed for an ANRSS filter.

If both the starting point and the optimal point of an adaptive nonlinear IIR filter are in the stable region. it is reasonable to anticipate the whole adaptation path lies inside the stable region if sufficiently small step sizes are used. If the adaptive filter were to enter unstable region, its output would become large and so would the MSE. A gradient-based adaptation algorithm would automatically force the system to return to the stable region. Hence, the step sizes can be chosen to be small enough to maintain the stability. The chances of instability can be reduced if the starting point is chosen to be close to the optimal point. For some applications, the optimal point can be estimated. An adaptive nonlinear IIR filter can start at this estimated point and performs fine tuning. Furthermore, if the system concerned has weak nonlinearity and the magnitudes of input signals are known to be bounded, the adaptive filter stability is mainly determined by the linear part of the adaptive filter. Monitoring the adaptive filter stability may be performed only on the linear part of the adaptive filter. To ensure that the nonlinear part does not cause instability, smaller step sizes can be used for its nonlinear coefficients.

According to Lyapunov's first stability theorem [16], the stability near the point of equilibrium of an autonomous nonlinear system is determined by the poles of the time-invariant linearized system:

$$\mathbf{x}_p(k+1) = \mathbf{A}_p x_p(k)$$

This theorem uses a linear system to approximate a nonlinear system and it is obvious that it applies only to small deviations about the point of equilibrium. Similarly, we would like to use a linear system with inputs to approximate a nonlinear system with

inputs at the point of equilibrium and expect that the stability of the linear system gives some information about the stability of the nonlinear system especially when the inputs are small. Since the right-hand side of Equation (5.3a) can be considered as a truncated n-dimensional Taylor Series, the equation of the adaptive filter using this model can be rewritten as

$$\mathbf{x}(k+1) = \mathbf{A}'(\mathbf{x}(k), u(k))\mathbf{x}(k) + \mathbf{B}u(k) + \mathbf{g}'(u(k)) \tag{5.22a}$$

$$y(k) = \mathbf{C}^T\mathbf{x}(k) + du(k) \tag{5.22b}$$

where the matrix A' is the time-variant feedback matrix depending on both the input signal u and the state variable $\mathbf{x}$, and the function g' is a purely nonlinear function of the input signal u and not a function of the state variable x. The term $\mathbf{B}u(k) + \mathbf{g}'(u(k))$ can be considered as the input of the linearized filter and thus has no effect on the filter stability. The feedback matrix A' determines the poles of the linearized filter. This equation takes account of a characteristic of an ANRSS filter: its dependence upon the value of the state variables and the value of the input signal for stability. Hence, it is reasonable to expect that the stability of the system in Equation (5.22) is closely related to that of the system in Equation (5.3). For each iteration, we can compute the matrix A' and monitor the poles of the linearized filter described in Equation (5.22). A similar stability monitoring technique was presented in [17]. The monitoring process is generally expensive but is trivial for some special cases, such as first- and second-order systems. To illustrate the idea, let us use a second-order system as an example:

$$\begin{bmatrix} x_{p1}(k) \\ x_{p2}(k) \end{bmatrix} = \begin{bmatrix} a_{p11} & a_{p12} \\ a_{p21} & a_{p22} \end{bmatrix} \begin{bmatrix} x_{p1}(k) \\ x_{p2}(k) \end{bmatrix} + \begin{bmatrix} g_{p1} \\ g_{p2} \end{bmatrix} + \begin{bmatrix} b_{p1} \\ b_{p2} \end{bmatrix} u(k)$$

where

$$\begin{bmatrix} g_{p11} \\ g_{p2} \end{bmatrix} = \begin{bmatrix} p_{p1}x_{p1}(k)x_{p2}(k)+p_{p2}x_{p1}(k)u(k)+p_{p3}x_{p2}(k)u(k)+p_{p4}u^2(k)+p_{p5}x_{p1}^2(k)+p_{p6}x_{p2}^2(k) \\ p_{p7}x_{p1}(k)x_{p2}(k)+p_{p8}x_{p1}(k)u(k)+p_{p9}x_{p2}(k)u(k)+p_{p10}u^2(k)+p_{p11}x_{p1}^2(k)+p_{p12}x_{p2}^2(k) \end{bmatrix}$$

Rewriting the above equation into the form of Equation (5.22), we have

$$\mathbf{A}' = \begin{bmatrix} a_{p11}+p_{p1}x_{p2}(k)+p_{p2}u(k)+p_{p5}x_{p1}(k) & a_{p12}+p_{p3}u(k)+p_{p6}x_{p2}(k) \\ a_{p21}+p_{p7}x_{p2}(k)+p_{p8}u(k)+p_{p11}x_{p1}(k) & a_{p22}+p_{p9}u(k)+p_{p12}x_{p2}(k) \end{bmatrix}$$

and

$$g' = \begin{bmatrix} p_{p4}u^2(k) \\ p_{p10}u^2(k) \end{bmatrix}$$

The pole determined by the time-variant matrix A' is used to monitor the stability. The stability analysis discussed above is an approximate technique. It has its limitations, for example, it would fail when the input signal u has a large magnitude.

During adaptation, knowledge of the filter stability at the current point in the coefficient space is important. One way is to compute a stability region at the point of equilibrium and decide if the current point is in the region. Many methods have been proposed to establish stability regions of nonlinear dynamical systems [14,15]. One class of the methods is based on a local Lyapunov function. The stability regions are those where the Lyapunov function satisfies the stability conditions. Generally, there exist no simple close-form solutions. The problem has to be solved numerically and is a quite computationally demanding process.

In the following, we analyze the convergence performance of an ANRSS filter. In fact, the results obtained are also applicable to adaptive linear IIR filters. In terms of the coefficients and the input signal, the physical system output and the adaptive filter output can be written as

$$y_p(k) = f(\mathbf{w}_p, u(k)) \tag{5.23}$$

$$\mathrm{y(k)} = f(\mathbf{w}^k, u(k)) \tag{5.24}$$

where the function f is assumed identical for both the physical system and the adaptive filter, w is the vector with all the updated coefficients, and $\mathbf{w}_p$ contains the physical system coefficients corresponding to those of w. Expanding the physical system output y_p around $\mathbf{w}^k$, we can approximate the output error $\mathrm{e(k)} = y_p(k) - y(k)$ by

$$\begin{aligned} e(k) &= \left(\frac{\partial y_p(k)}{\partial \mathbf{w}_p} \Big|_{\mathbf{w}^k}\right)^T \mathbf{e}_w(k) \\ &= \left(\frac{\partial y(k)}{\partial \mathbf{w}}\right)^T \mathbf{e}_w(k) \end{aligned} \tag{5.25}$$

where $e_{,,,}(k) = \mathbf{w}_p - \mathbf{w}^k$

Considering the LMS updating formula Equation (5.4), we can write

$$\mathbf{e}_{w'}(k+1) = e_{,,,}(k) - 2\mu \frac{\partial y(k)}{\partial \mathbf{w}} e(k) \tag{5.26}$$

Substituting Equation (5.25) into the above equation and taking expectation of both sides of the equation gives

$$E(\mathbf{e}_{w'}(k+1)) = (\mathbf{I} - 2\mu E(\frac{\partial y(k)}{\partial \mathbf{w}}(\frac{\partial y(k)}{\partial \mathbf{w}})^T))E(\mathbf{e}_{w'}(k)) \tag{5.27}$$

where as in the convergence analysis of an adaptive transversal filter in Chapter Two, it is assumed that the coefficient error vector $e_{,,,}(k)$ is uncorrelated with the gradient vector $\partial y(k)/\partial \mathbf{w}$. From Equation (5.25), we can see that the assumption means that the mean of the error signal e is zero, which can be often satisfied in practice. The above equation can be written as

$$E(\mathbf{e}_{w'}(k+1)) = (\mathrm{I} - 2\mu \mathbf{R})E(\mathbf{e}_{w'}(k)) \tag{5.28}$$

where the matrix R is the correlation matrix of the gradient vector $\partial y(k)/\partial \mathbf{w}$.

We can see Equation (5.28) is similar to Equation (2.11). Hence, similar results can be obtained. Let λ_{max} and λ_{min} indicate the maximum eigenvalue and the minimum eigenvalue of the correlation matrix R. Following the similar convergence analysis in Chapter two. it can be shown that the step size μ must satisfy the zero-mean condition

$$0 < \mu < \frac{1}{\lambda_{max}} \tag{5.29}$$

and a sufficient condition is

$$0 < \mu < \frac{1}{3tr(\mathbf{R})} \tag{5.30}$$

The bounds $1/\lambda_{imax}$ and $1/(3tr\mathbf{R}_i)$ will be referred to as the zero-mean upper bound and finite-variance upper bound. respectively. The convergence time is

$$\tau = \frac{1}{\alpha}\frac{\lambda_{max}}{\lambda_{min}} \tag{5.3 1}$$

for a step size α/λ_{max}, where $\boldsymbol{a}$ is a constant between 0 and 1. This shows that the convergence speed of an ANRSS filter depends on the eigenvalue spread of correlation matrix of the gradient signals. Although this matrix is not constant in the adaptation process, this analysis based on the localized information still provides valuable insight into the convergence performance of an ANRSS filter. These theoretical predictions are compared with those in the actual convergence results in Section 5.7.

5.6. Potential Applications

This section discusses potential applications of ANRSS filters, which include linearization of nonlinear systems and echo cancellation in digital communications.

5.6.1 Linearization of a Class of Nonlinear Systems

The nonlinear function $\mathbf{g}_p$ of the physical system in Equation (5.3) brings in nonlinearity and causes distortion. The nonlinear term can be canceled out by subtracting an estimate of it from the right hand side of Equation (5.3.a). The estimate can be obtained by an adaptive nonlinear filter.

The adaptive linearization scheme is illustrated in Fig.5.1. The adaptive filter is built using the model in Equation (5.3). The adaptive filter estimates the nonlinear coefficient vector $\mathbf{p}_p$ and the state vector $\mathbf{x}_p(k)$, which, together with $u(k)$, determine the estimate of $\mathbf{g}_p$.

The adaptive linearization process has two phases: first an identification phase, then a linearization phase. In the identification phase, the output of the nonlinear function g of the adaptive filter is fed to itself so that the adaptive filter is able to match the physical system. Once the adaptive filter matches the physical system, $\mathbf{g}(\mathbf{p},u(k),\mathbf{x}(k))$ is expected to be a good estimate of $\mathbf{g}_p(\mathbf{p}_p, u(k),\mathbf{x}_p(k))$. In the linearization phase, the switch of Fig.5.1 will toggle so that $-\mathbf{g}(\mathbf{p},u(k),\mathbf{x}(k))$ is fed to the physical system and thus linearizes the system. Note that switching out the nonlinearity makes the adaptive filter linear and enables its states to trace the linearized system. The term $z^{-\delta}$ in Fig.5.1 models the possible delay between the system output and measured output.

The linearization scheme can be applied to a loudspeaker system. The basic structure of adaptive linearization of a loudspeaker using an ANRSS filter is shown in Fig.5.2. Assume that the nonlinearity of a loudspeaker is due to its suspension system. As discussed in Chapter Four, the displacement of the cone in a loudspeaker satisfies

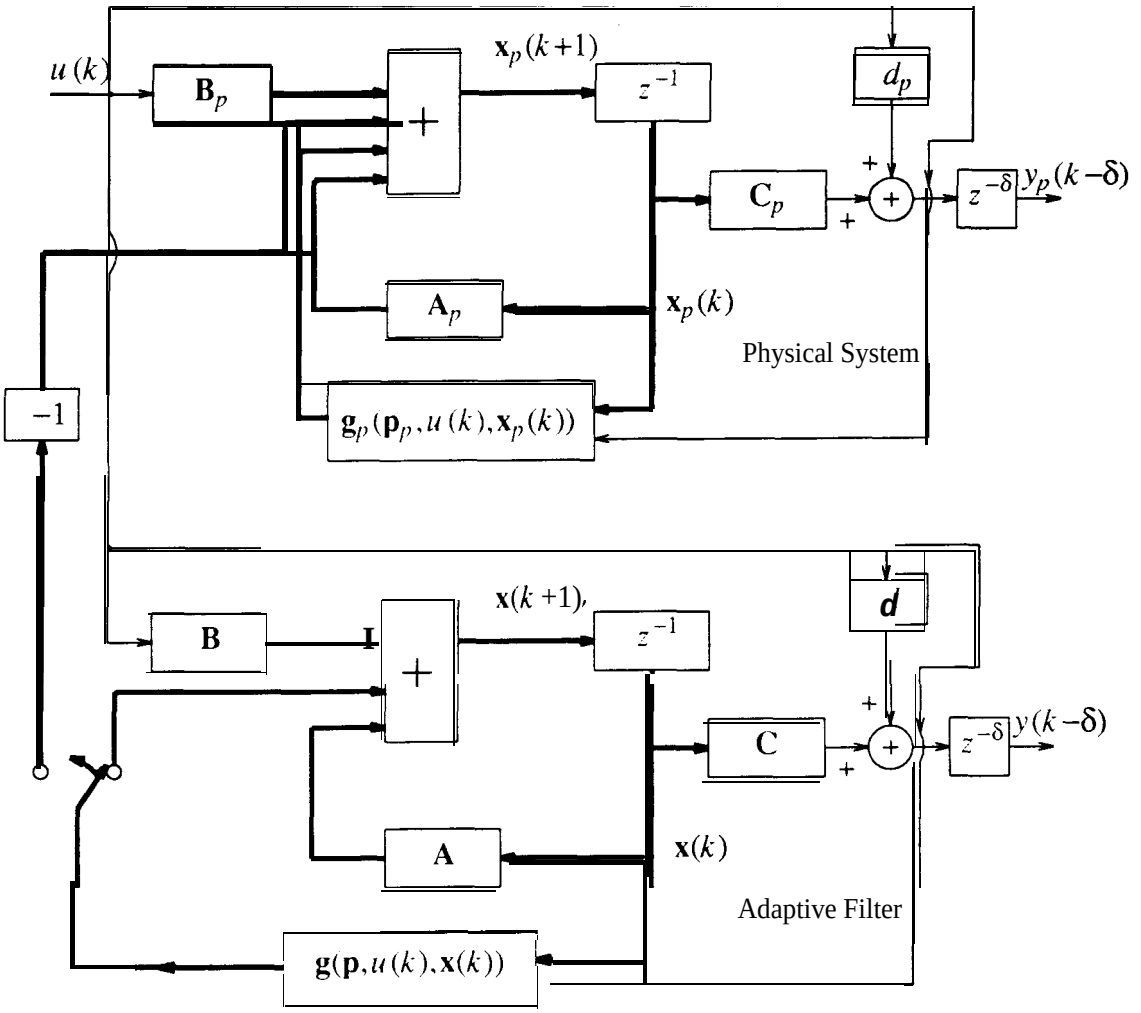


Fig.5.1 The adaptive linearization scheme using the nonlinear state-space filter

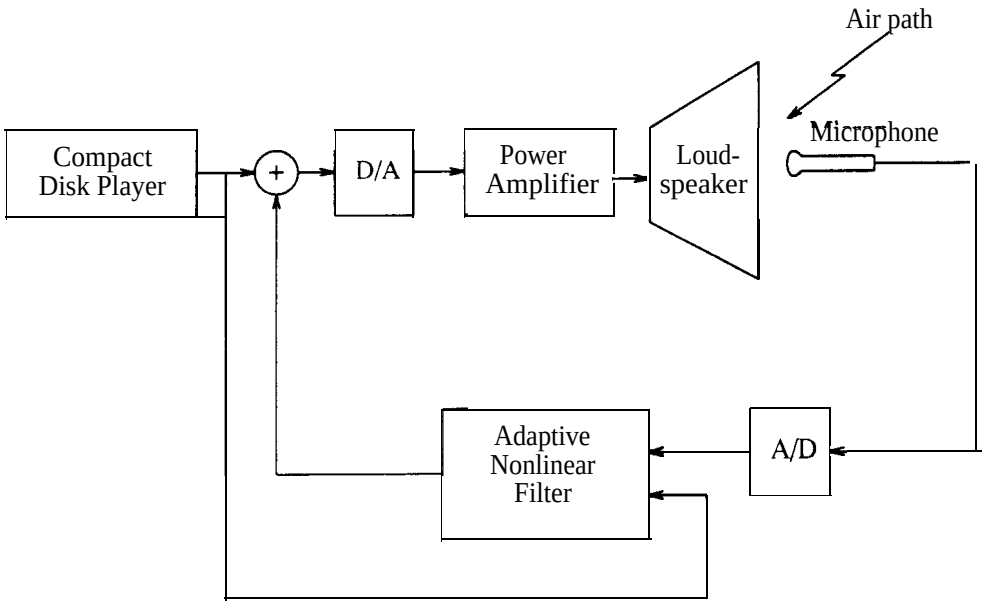


Fig.5.2 Adaptive linearization of a loudspeaker using an ANRSS filter.

$$x(k+2) - a_{p2}x(k+1) - a_{p1}x(k) - p_{p1}x^2(k) - p_{p2}x^3(k) = b_{p2}u(k) \qquad (5.32)$$

where u is the current of the coil. Choosing $x_{p1}(k) = x(k), x_{p2}(k) = x(k+1)$, this difference equation can be written in the state-space form of Equation (5.3) with

$$\mathbf{A}_p = \begin{bmatrix} 0 & 1 \\ a_{p1} & a_{p2} \end{bmatrix} \qquad (5.33)$$

$$\mathbf{g}_p = (\,0 \;\; g_p(\mathbf{p}_p, \mathbf{x}_p(k))\,)^T \quad \mathbf{B}_p = (\,0 \;\; b_{p_2}\,)^T \quad \mathbf{C}_p = (\,1 \;\; 0\,)^T$$

where $g_p(\mathbf{p}_p, \mathbf{x}_p(k)) = p_{p1}x_{p1}^2(k) + p_{p2}x_{p1}^3(k)$ and $\mathbf{p}_p = (\,p_{p1} \;\; p_{p2}\,)^T$. An estimate of the term $-\mathbf{g}_p(\mathbf{p}_p, \mathbf{x}_p(k))/b_{p_2}$ is required to add to the input signal $u(k)$ to cancel out the nonlinearity of the original system.

5.6.2 Echo Cancellation in a Nonlinear Data Transmission Channel

An echo canceler is typically configured in a digital subscriber loop as shown in Fig.5.3. Most cancelers reported are linear, based on the assumption that the echo channel is linear. Nonlinearity exists, however, and can severely limit the success of echo cancellation by a linear canceler. The most notable sources of nonlinearity include the D/A converter [8,11] and the line driver [11] at the transmission end. Nonlinear cancelers, based on nonlinear FIR filters, have been reported in [8-11]. Since the linear part of the channel is often better approximated by a pole-zero model [2], an IIR filter may be preferred in terms of computation efficiency. The echo channel may be modeled as a nonlinear memoryless system followed by a linear dispersive system described by an IIR transfer function.

5.7. Numerical Examples

Three examples are shown to illustrate the utilization and performance of the adaptive filters proposed. The first example is the same as the one for adaptive FIR filters in Section 5.2. This example is simple, however very suitable for illustration. The second example is identification and linearization of a loudspeaker model. The last example is for nonlinear echo cancellation in a data communication channel.

5.7.1 Example 1 - First-Order System

The input coefficient vector $\mathbf{B}_p$ of the first-order example in Equation (5.2) has a known zero-nonzero pattern: the only element is always nonzero. Hence, the adaptive filter input coefficient b was fixed at unity and other coefficients were adapted. The

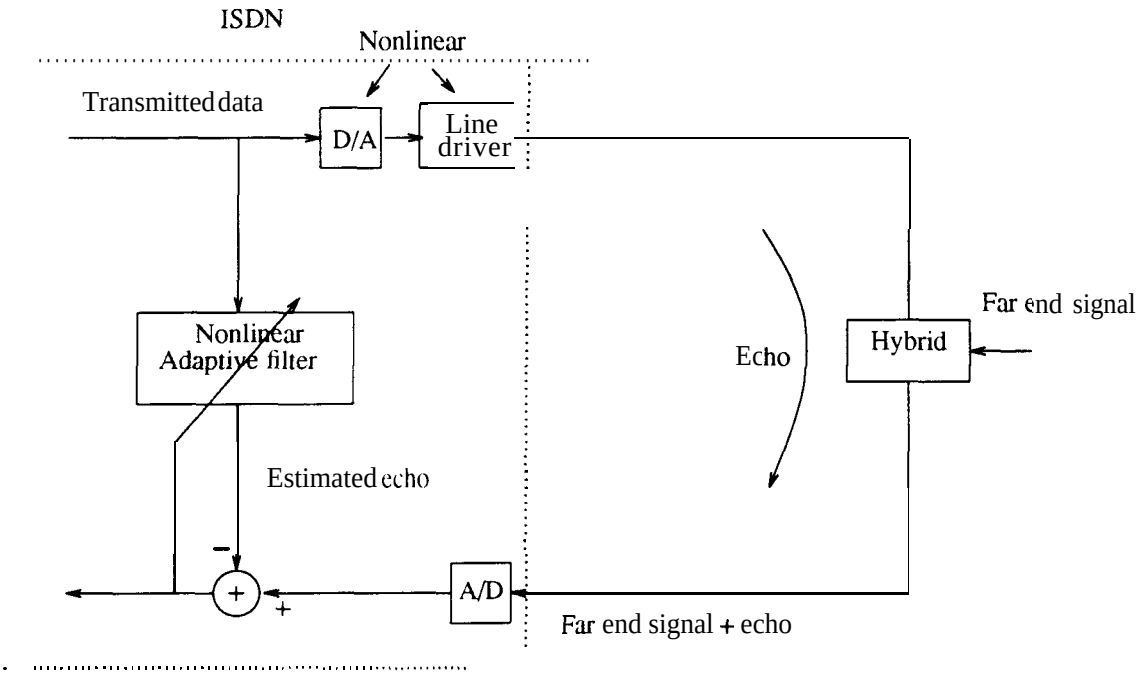


Fig.5.3 Typical configuration of a subscriber loop with an echo canceler.

physical system in Equation (5.2) was used as the reference system. The adaptive filter updated its coefficients a, c, p_1, and p_2, with initial values being zero. The step sizes were $\mu_a = 0.0005$, $\mu_c = 0.01$, and $\mu_p = 0.0005$.

To show the effect of not adapting nonlinear coefficients. a test was run which adapted the linear coefficients only. Curve (a) in Fig.5.4 was from this test. The MS error could go down to only about $-15dB$.

The approximate stochastic-gradient method was simulated next. The convergence curve is depicted in Fig.5.4 as curve (b), which shows that the MSE was reduced from $0dB$ to below $-100dB$ after $1k$ iterations. Two contours have been drawn in

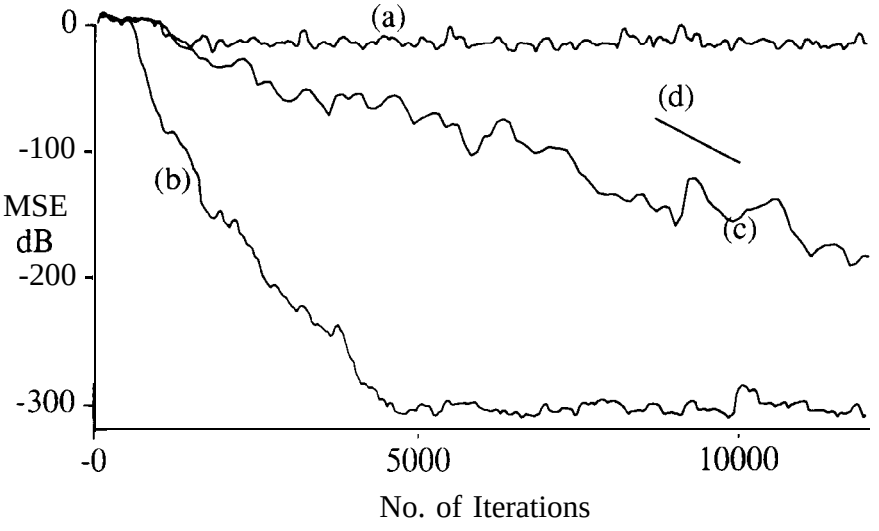


Fig.5.4 Convergence curves for the first-order example:
(a) adapting linear coefficients only;
(b) approximate stochastic-gradient method with different step sizes for the parameters;
(c) approximate stochastic-gradient method with a uniform step size for all the parameters:
(d) theoretical convergence rate for (c).

Figs.55 and 5.6 for the linear and nonlinear coefficients to show the performance surface and the adaptation behavior of the algorithm. Small step sizes were used for the adaptation paths so that the paths are smooth. It is obvious from the contour plots that the paths are generally normal to the contours, which is a characteristic of the steepest descent algorithms.

The stochastic-gradient method (without approximating gradients) was simulated next and very small differences between the results of the approximate stochastic-gradient method and the stochastic-gradient method were observed. The convergence curve and adaptation paths of the approximate stochastic-gradient method are slightly less smooth than those of the stochastic-gradient method since the nonlinearities neglected in computing gradients by the approximate stochastic-gradient method create noise in gradient computation.

We are now in a position to make a comparison between the results of the adaptive FIR filters and adaptive IIR filters on this first-order example. The major results for the adaptive nonlinear IIR and FIR filters are summarized in Table 5.2. For this example, an ANRSS filter is able to match the reference physical system perfectly, with 0.4% of the computation required by the adaptive nonlinear FIR filter per iteration and with 6% of its convergence time. For this example, the adaptive IIR filters definitely outperformed the adaptive FIR filters.

Table 5.2 Major Results for the First-Order Example Using the Adaptive Nonlinear IIR and FIR Filters

Test	MSE(dB)	Multiplication/iter	Iterations
IIR	-300	33	5k
FIR	-38.2	7.84k	80k

The number of multiplications per iteration for IIR filter was that of the stochastic-gradient method.

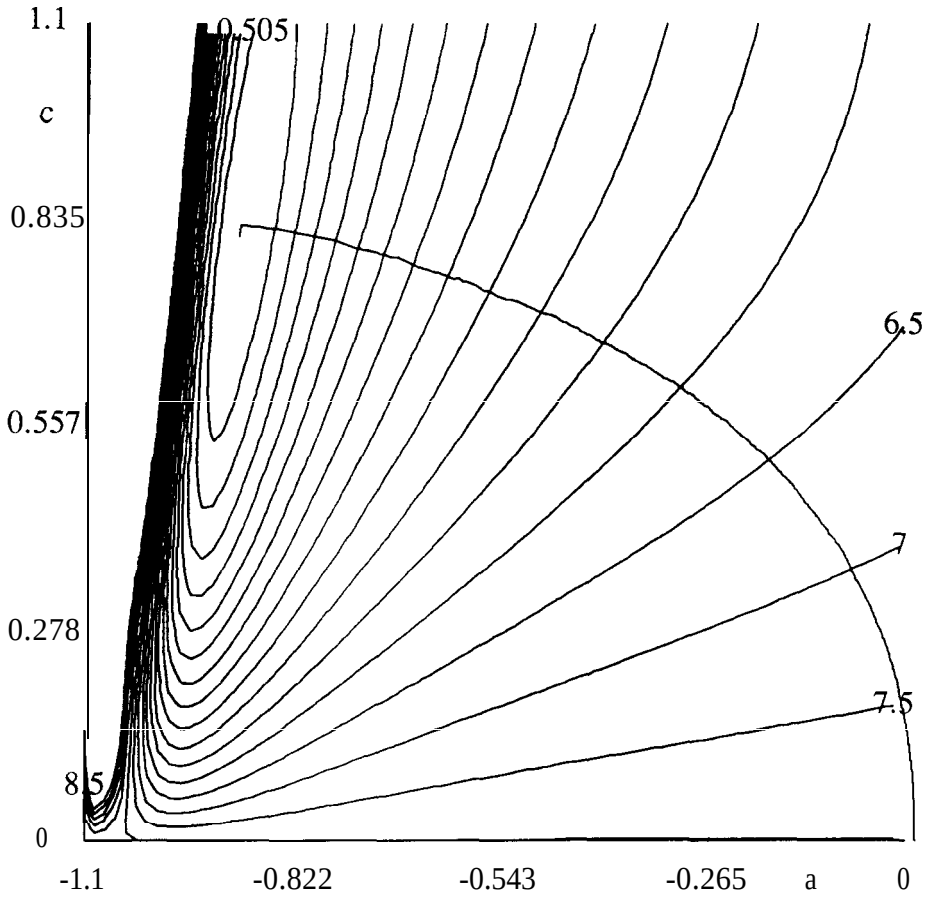


Fig.5.5 Nonlinear state-space filter for the first-order example. μ_a=7e−5, μ_c=7e−5. Approximate stochastic-gradient method.

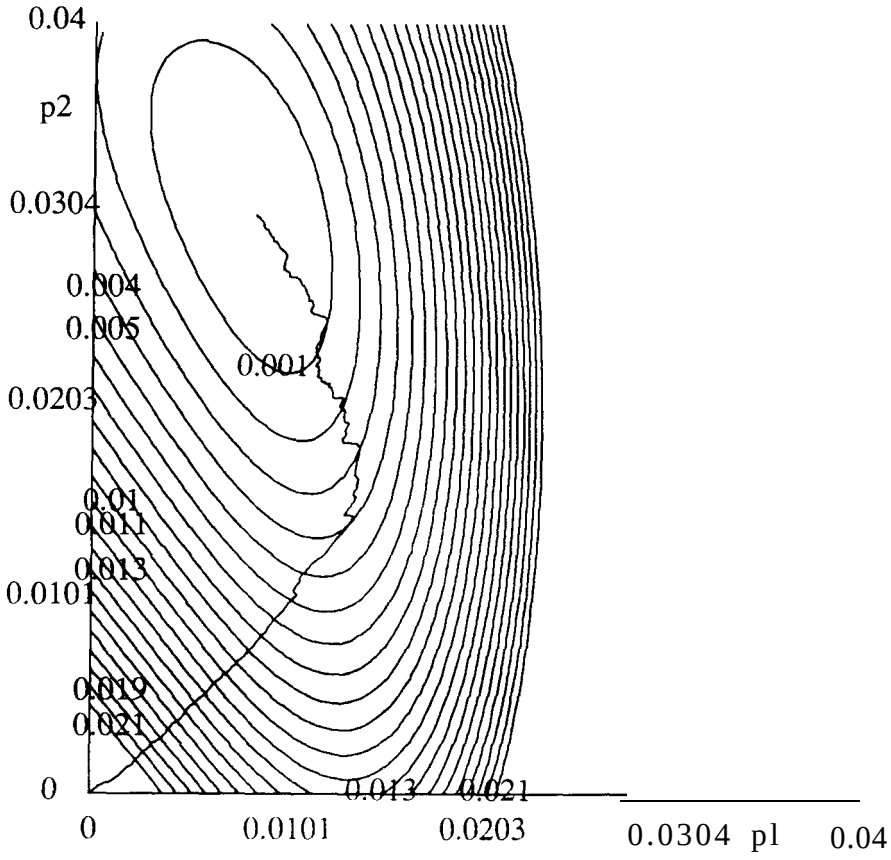


Fig.5.6 Nonlinear state-space filter for the first-order example. μ_p=5e−6. Approximate stochastic-gradient method.

It would be interesting to compare simulation results on convergence with the theoretical results. In the following simulations, the input coefficient b of the filter was set to be equal to the input coefficient b_p of the system and was kept fixed to be consistent with the assumption of the theoretical analysis. Other filter coefficients were adapted and were set to zero initially. The correlation matrix R of the gradient signals for the four parameters adjusted is not constant in the adaptation. As an approximation, it was calculated at the optimal point, namely, the filter coefficients equal to the corresponding system coefficients. The eigenvalues of the matrix are

$$eigenvalue = (\ 1.8852\ \ 6.7106\ \ 35.8057\ \ 178.8985\ )^T$$

The theoretical zero-mean upper bound for the step size is

$$\mu_{zero-mean} = \frac{1}{\lambda_{\max}} = 0.0056$$

and the f-mite-variance upper bound is

$$\mu_{finite-variance} = \frac{1}{3tr(\mathbf{R})} = 0.0015$$

Experiments were made and the practical maximum step size allowed for convergence was found to be

$$\mu_{practical} = 0.0008$$

which is smaller than both the theoretical zero-mean and finite-variance upper bounds. The practical upper bound is supposed to be between these two theoretical bounds. However, this is not the case in the above simulations. The theoretical upper bounds were obtained from the correlation matrix computed at a particular point, the optimal point, while the practical upper bound was determined by the correlation matrixes computed in all the iterations of the adaptation process. The step size $\mu = 0.0008$,

corresponding to $\alpha = 1/7$, was then used and the convergence curve is plotted in Fig.5.4 as curve (c).

According to Equation (5.31) the theoretical time constant for the step size $\mu_{practical}$ is

$$\tau_{theoretical} = \frac{\lambda_{\max}}{\alpha\lambda_{\min}} = 665 \quad iterations$$

where $\alpha = 1/7$. This means that after 665 iterations, the coefficient errors are smaller by e, in other words, reduced by $8.7dB$. The mean square error can be approximated by a quadratic function. without a linear term, of the coefficient error vector. Hence, after 665 iterations, the MSE is expected to be reduced by about $17.4dB$. A straight line is drawn in Fig.5.4 as curve (d) whose slope is the theoretical convergence speed. Although the practical convergence speed varied with time, its overall slope is close to that of the straight line. The estimated practical time constant is

$$\tau_{practical} = 565 \quad iterations$$

which is very close to the theoretically predicted value.

5.7.2 Example 2 - Identification and Linearization of a Loudspeaker

Identification and linearization of a loudspeaker has been simulated. The parameters of the loudspeaker model were $a_{p1} = 0.3$, $a_{p2} = 0.2$, $p_{p1} = 0.006$, $p_{p2} = 0.03$, $b_{p2} = 0.6$, and $c_{p1} = 1$. That p_{p2} was chosen larger than p_{p1} was to be consistent with the fact that the cubic term is dominant in the suspension nonlinearity. The adaptive filter input coefficient vector was set to be a constant vector $(0\ 1)^T$. The adaptive filter updated its coefficients a_1, a_2, p_1, p_2,and c_1, with zero initial values. The step sizes were $\mu_a = 0.02$ for a_1 and a_2, $\mu_p = 0.001$ for p_1 and p_2, $\mu_c = 0.02$ for c_1. The

delay in the air path was chosen as 50 sampling periods. In practice, this delay can be measured by feeding an impulse signal to the loudspeaker or using an adaptive linear transversal filter to estimate it. It is also possible to cascade an adaptive linear transversal filter with an ANRSS filter to perform on-line estimation of the delay. The interaction between the two cascaded filters may influence the convergence of the system.

Both the stochastic-gradient method and the approximate stochastic-gradient method were run. The convergence curves of the two methods are similar, with differences of a few dB in the final stage of the runs. For the sake of brevity, only the curve for the approximate stochastic-gradient method is shown here in Fig.5.7 for identification up to $30k$ iterations. It is seen that the MS error has reached the numerical noise floor at about $-300dB$ after $7k$ iterations.

To measure the performance of the linearized loudspeaker system, we used a reference loudspeaker system, which just had the linear part of the speaker system to be linearized. Its output will be referred to as y_{ref_lin}. At the beginning of adaptation, the loudspeaker system and reference system had zero initial states. The linearization took effect at $10k$ iterations. At the time of switching from the identification phase to the linearization phase, we assigned the state variables of the loudspeaker being linearized to the state variable of the reference system so that these two systems had the same initial states after switching. Thus, the MS value of $y_p - y_{ref\_lin}$ computed before switching measures the original distortion and the value computed after switching measures the residual distortion. The loudspeaker output consisted of a linear signal and a nonlinear distortion, whose mean squares before linearization were $-3dB$ and $-23dB$, respectively. This gives a nonlinear-to-linear ratio of ***-20dB.*** that is, the nonlinear

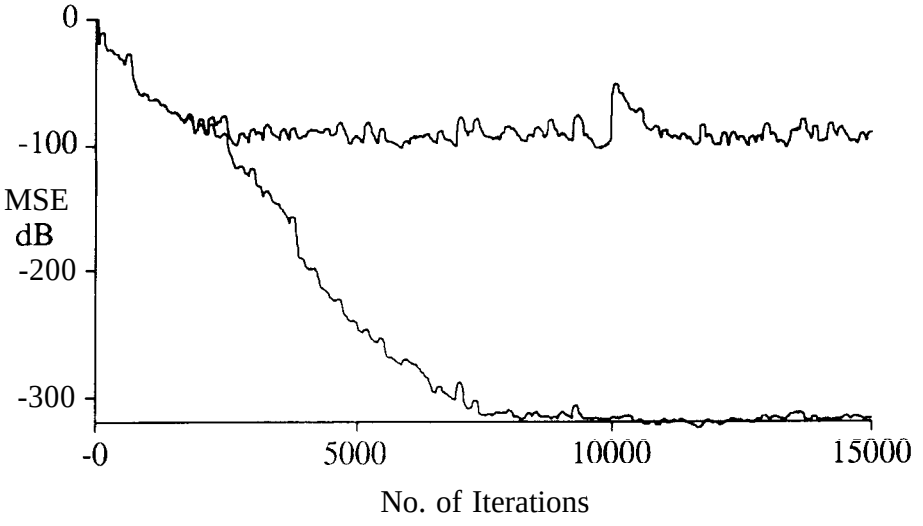


Fig.5.7 Convergence curves for the loudspeaker example. The lower curve is for the case where the filter's model is the same as that of the system's. The upper curve for the case where the filter's model is not exactly the same as that of the system's. In both cases, the approximate stochastic-gradient method was used.

signal is 10% of the linear signal. The nonlinear distortion was reduced from *-3dB* to *-310dB* after linearization. This distortion reduction is so good that it can only be achieved in simulation, and some factors, such as measurement noise and model mismatch, will degrade the performance in a practical situation.

It would be interesting to see whether the adaptation algorithms and linearization scheme are robust: do they work or not if a deviation is present between the filter model and physical system model? Suppose a practical loudspeaker also has a nonzero quartic term in the nonlinear feedback term, that is $g_p(\mathbf{p}_p, \mathbf{x}_p(k)) = p_{p1} x_{p1}^2(k) + p_{p2} x_{p1}^3(k) + p_{p3} x_{p1}^4(k)$ but the adaptive filter just has a nonlinear feedback with

$g(\mathbf{p},\mathbf{x}(k)) = p_1 x_1^2(k) + p_2 x_1^3(k)$. The parameter p_{p3} was chosen to be 2×10^{-5}. Other parameters were the same as before. Simulations were performed using both the stochastic-gradient and the approximate stochastic-gradient methods, and very similar results have been obtained. The convergence curve was plotted in Fig.5.7 for the approximate stochastic-gradient method. The adaptive filter worked well and reduced the MS errors to about $-90dB$, a residual floor determined by the term in the loudspeaker which was not modeled by the adaptive filter. The nonlinear distortion was reduced from about $-23dB$ to $-49dB$.

5.7.3 Example 3 - Echo Cancellation

Assume that the dominant nonlinearity in the echo path is from the D/A converter [8]. Due to processing imperfections, an integrated D/A converter has a systematic nonlinearity. The nonlinearity can often be modeled as a memoryless nonlinear function. One typical nonlinear transfer function for the integral nonlinearity of a MOS D/A converter is [9]

$$y(u) = bu + b_3 u^3 \tag{5.34}$$

The input signal to the adaptive echo canceler and the D/A converter is in digital form. The signal code in the simulations was 2B1Q, namely, pairs of bits encoded as four level pulses are transmitted. A popular model for the linear part of the channel used for simulations [8,9] is

$$y_p(z) = \frac{1}{1 - a_p z^{-1}} \tag{5.35}$$

This model, together with the D/A converter model in Equation (5.35), was used for our first echo cancellation example. The channel parameter a_p was chosen to be -0.4 as in

[8]. The D/A converter parameters were $b_p = 1.01333$, $b_{p3} = -0.01333$ as in [9]. The adaptive filter has the following structure

$$x(k+1) = ax(k) + bu(k) + b_3 u^3(k)$$
$$y(k) = cx(k) \tag{5.36}$$

The input coefficient b was set to unity and not adapted. The coefficients a, c, and b_3 were adapted, with initial values being zero. The step sizes were 0.1 for a and c, and 0.05 for b_3. The adaptive filter suppressed the echo to -3 $10dB$ at $7.2k$ iterations for both the stochastic-gradient method and the approximate stochastic-gradient method.

In the second echo cancellation example, a third-order linear system was used as the model of the linear part of channel with poles at 0.9375 and $0.9375 \pm j\ 0.1776$, zeros at $0.969 \pm$ j0.2323 and a gain factor of 0.22. The linear part of the adaptive filter employed the quasi-orthonormal structure [4,5]. The initial values of the matrix A were determined so that the initial poles of the linear part of the adaptive filter were at 0.9. The input coefficient vector B was set to be $(0\ 0\ 1)^T$ and was not adapted. The output coefficient vector C was adapted, with zero initial values. The step sizes were $\mu_a = 0.0001$, $\mu_c = 0.001$, and $\mu_{b_3} = 0.001$. The plot of mean squared echo residual is shown in Fig.5.8 for the approximate stochastic-gradient method and the stochastic-gradient method gave a similar curve.

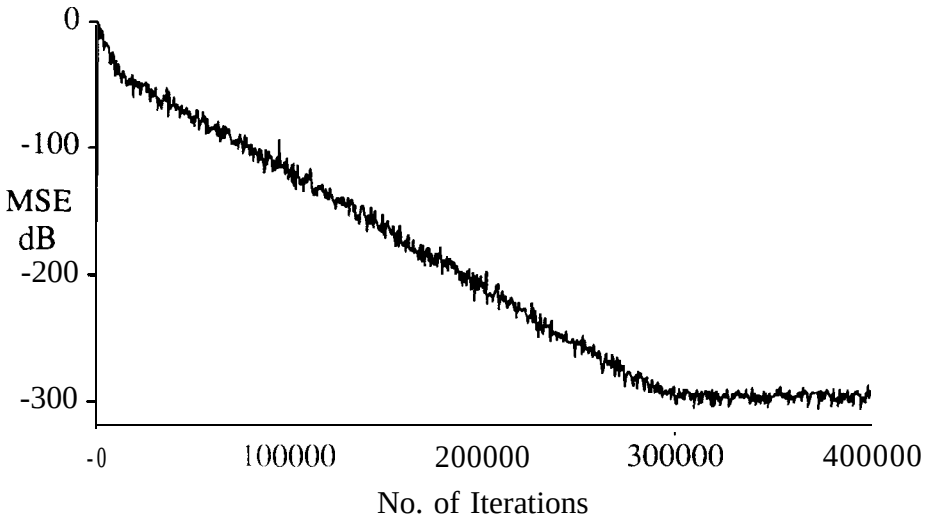


Fig.5.8 Convergence curve for the third-order echo cancellation example. The approximate stochastic-gradient method was used.

5.8. Summary

ANRSS filters have been introduced in this chapter, which are computationally more attractive than adaptive nonlinear FIR filters for some applications. To take advantage of the ANRSS filters, one has to have some knowledge of the system: most importantly the mathematical structure of the system. Knowledge of the estimated values of the system parameters can also be used to improve the filter performance.

Efficient adaptation algorithms have been developed for ANRSS filters. It has been shown that the input coefficient vector need not be adapted if we know the zero-

nonzero pattern of the input coefficient vector of the physical system to be matched. The gradients of the adaptive filter coefficients can be efficiently computed by neglecting the nonlinearity in the system in the case of weak nonlinearity, Although the nonlinearity is neglected when computing gradients, it is still used to evaluate the adaptive filter output. The approximate stochastic-gradient method performed quite well in our simulations. Choices of the step size and stability monitoring of an ANRSS filter have been discussed. Convergence analysis has shown that the adaptive filter convergence relies on the eigenvalue spread of the correlation matrix of the coefficient gradient signals. A scheme for canceling nonlinearity for a class of nonlinear systems was proposed and was applied to linearization of a loudspeaker model with nonlinearity only in the suspension system.

References

[1] J.J. Shynk, "Adaptive IIR filtering," *IEEE ASSP Magazine*, pp. 4-21, vol. 6, April 1989.

[2] H. Fan and W.K. Jenkins, "An Investigation of an Adaptive IIR Echo Canceler: Advantages and Problems." *IEEE Trans. on Acoustics, Speech, and Signal Pi-ocessing*, pp. 1819- 1834, vol.36, Dec. 1988.

[3] T. Kwan and K.W. Martin, "Adaptive Detection and Enhancement of Multiple Sinusoids Using a Cascade IIR Filter," *IEEE Tras. on Circuits and Systems*, pp.937-947, vol. 36, July 1989.

[4] D.A. Johns, W.M. Snelgrove, and A.S. Sedra, "Adaptive Recursive State-Space Filters Using a Gradient Based Algorithm," *IEEE Trans. on Circuits and Systems*, pp.673-684, vol. *37*, June 1990.

[5] D.A. Johns, "Analog and Digital State-Space Adaptive IIR Filters," *Ph.D. Thesis*, University of Toronto, 1989.

[6] F.X.Y. Gao, W.M. Snelgrove, and D.A. Johns, "Nonlinear IIR Adaptive Filtering Using A Bilinear Structure," *Proc. of* ***IEEE*** *International* ***Symposium*** *on* ***Circuits and*** *Systems,* pp. 1740-1743, May 1988.

[7] F.X.Y. Gao and W.M. Snelgrove, "Adaptive Nonlinear State-Space Filters," *Proc. of* ***IEEE*** *International Symposium on Circuits and Systems,* pp.3122-3125, May 1990.

[8] Y. Takahashi, et al "An ISDN Echo-Canceling Transceiver Chip for 2BlQ Coded U-Interface," *Proc.* ***oj'*** *IEEE International Solid-State Circuits Conference,* pp.258-260, *1989.*

[9] K. Murano, S. Unagami, and F. Amano, "Echo Cancellation and Applications," *IEEE Communications Magazine,* vol. 28, pp.49-55, Jan. 1990.

[10] M.J. Smith, C.F.N. Cowan and P.F. Adams, "Nonlinear Echo Cancelers Based on Transposed Distributed Arithmetic." *IEEE Trans. on Circuits and Systems,* vol.35, pp.6- *18,* Jan. 1988.

[11] 0. Agazzi, D.G. Messerschmitt. and D.A. Hodges, "Nonlinear Echo Cancellation of Data Signals," ***IEEE*** *Trans. Commun.,* vol. COM-30, pp. 2421-2433, Nov. 1982.

[12] G. L. Sicuranza, A. Bucconi, and P. Mitri, "Adaptive Echo Cancellation with Nonlinear Digital Filters." *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp.3.10.1-4, 1984,

[13] H. Khorramabadi, et al "An ANSI Standard ISDN Transceiver Chip Set," *Proc. of* ***IEEE*** *International Solid-State Circuits Conference,* pp.256-257, *1989.*

[14] H.D. Chiang and J.S. Thorp, "Stability Regions of Nonlinear Dynamical Systems: a Constructive Methodology," *IEEE Trans. on Automatic Control,* vol. 34, pp. 1229-1241, Dec. 1989.

[15] R. Genesio, M. Tartaglia, and A. Vicino. "On the Estimation of Asymptotic Stability Regions: State of the Art and New Proposals," ***IEEE*** *Trans. on Automatic* ***Control,*** vol. AC-30, pp.747-755, August 1985.

[16] F. Csaki, *Modern Control* ***Theories,*** Pudapest: Akademiai Kiado, 1972.

[17] P. Urwin and B.H. Swanick, "Adaptive Control of Systems with Certain Non-Linear Structures," *Int. J.* ***Control,*** pp.3 1-55, vol.39, 1984.

Chapter Six

Results on Loudspeaker Measurements

6.1 Introduction

In the previous chapters, the algorithms proposed in this thesis have been simulated successfully on mathematical models. This chapter applies the algorithms to measured loudspeaker data. After illustrating the measurement setup and characteristics of the data, we discuss solutions to some practical problems. Then, we present results on identification of the loudspeaker system by an adaptive linear FIR filter, a nonlinear FIR filter, an equation-error filter, a linear state-space filter, a backpropagation cascade filter, and a nonlinear state-space filter. Finally, we apply the predistortion technique to linearize the extracted model of the loudspeaker.

6.2 Loudspeaker Measurements

Measurements were performed in an anechoic chamber at the National Research Council of Canada. The measurement setup is shown in Fig.6.1. Because the measured data were originally meant for a linearization study, signals with low to medium frequencies were of interest and the signal level was chosen to be relatively high. Two low-pass analogue filters with cut-off frequencies at 1 kHz were employed for anti-aliasing and anti-imaging. They were fourth-order Butterworth filters. The D/A and A/D converters have 16 bits. The signal generator produced white noise. The woofer

had a diameter of about 6 inches. The chamber had a size of 11 feet x 11 feet x 18 feet. The microphone was placed 6.56 feet away from the loudspeaker. The sampling rate was 8 kHz and the data were measured with the SPL (sound pressure level) at the microphone adjusted to $85dB$. The number of samples recorded was 16128, which was the maximum number obtainable by the recording system. However, it should be noted that the generator was only able to produce a maximum of about 8192 or 2^{13} independent samples.

An impulse response of the loudspeaker system was also recorded and is plotted in Fig.6.2 for the first 700 samples. As shown in Fig.6.1, the loudspeaker system (simply referred to as the system later) consists of all the components on the signal path from the input of the D/A to the output of the A/D. At the beginning of the response, there

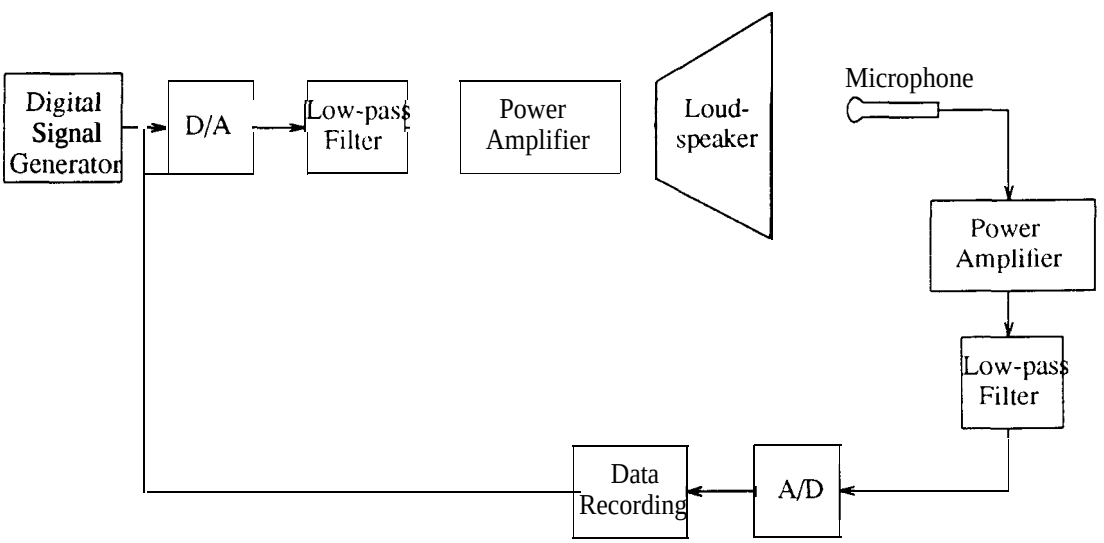


Fig.6.1 Measurement setup.

was a period of low-level noise caused by a delay in the signal path. This period is not shown in the figure so that the measured impulse response can be more conveniently compared with the impulse responses of adaptive filters presented later. The transfer function computed from the measured impulse response is plotted in Fig.6.3. It has a high attenuation at low frequencies (below the loudspeaker resonance) and rolls off above $1k$ Hz due to the analogue filters.

The data have some interesting characteristics. Although the measurement was performed in an anechoic chamber, noise and echoes still exist. The echoes are visible in Fig.6.2 and they appeared near the $490th, 548th$, and *606th* samples, separated by about 58 samples. The data have a DC component from A/D converter offset. In addition to nonlinearities in the loudspeaker, the A/D and D/A converters also contribute noniinearities to the system and they have integral and differential nonlinearities. The

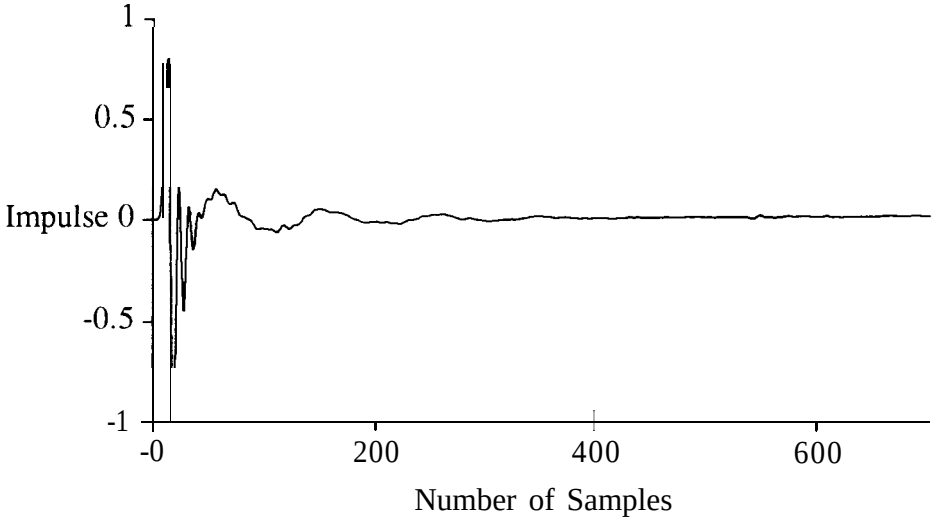


Fig.6.2 The measured impulse response of the loudspeaker system.

transfer function of the system is bandpass, with high attenuations at low and high frequencies. This will impose difficulties for inverse modeling of the system.

6.3 Considerations for Some Practical Problems

The DC component of the measurement data can be considered by adding a DC term in an adaptive filter. For an adaptive filter based on the output-error formulation, the DC term p_{dc} can be adapted to minimize the output error according to

$$p_{dc}^{k+1} = p_{dc}^{k} + 2\mu_{dc} e(k) \tag{6.1}$$

where μ_{dc} is the step size and e is the output error. For an adaptive filter based on the equation-error formulation or the backpropagation formulation, Equation (6.1) could be

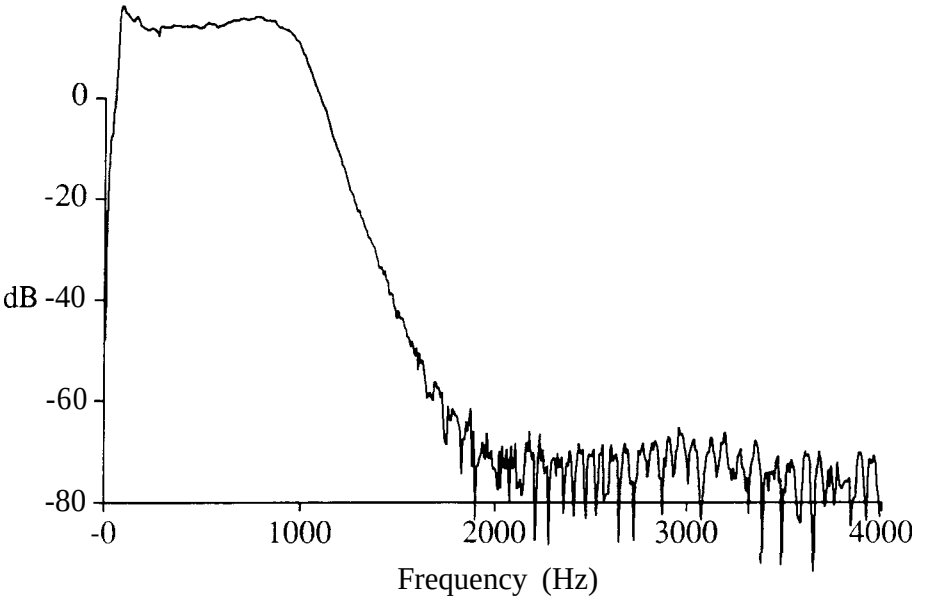


Fig.6.3 The transfer function of the loudspeaker system, computed from the measured impulse response.

used. However, it is more natural to estimate the DC term by computing the average of a block of the measurement data since the output error is not directly minimized in the algorithms. For each sample, the block average can be obtained by

$$p_{dc}^{k+1} = p_{dc}^{k} + (d(k) - d(k-N))/N \tag{6.2}$$

where the initial value of p_{dc} is set to be zero, N is the length of data block, and $d(i)$ is the desired signal and is equal to zero for $i < 0$.

Because the D/A and A/D converters have 16 bits, the recorded data are integers, with a maximum magnitude of 2^{15}. Such large input signals will cause numerical difficulties in nonlinear tilters: very small filter coefficients will result. Hence, the data should be normalized by dividing by 2^{15}.

As pointed out earlier, the transfer function of the loudspeaker system is bandpass, which is common for practical systems. In Chapter Four, a straight-forward inverse method was employed, where an adaptive filter attempts to minimize the mean square of the difference between the system output and a delayed input signal. Inversion of the high-attenuation parts of the transfer function results in high gains in an adaptive inverse filter, which in turn requires a long impulse response or causes slow convergence of the adaptive inverse filter. The values of an inverse function of the loudspeaker transfer function are meaningless at those frequencies above $1kHz$ because of the anti-imaging and anti-aliasing filters. In general, the performance of a loudspeaker at very high and low frequencies is not important because human ears are insensitive below 20 Hz or above 20 kHz [6]. Hence, the solution may be that inversion is avoided on high-attenuation parts of the forward transfer function at those frequencies of no interest. Since both the pre-distortion and post-distortion techniques require

inverse modeling of a practical system, the following discussion on inverse modeling will not be restricted to linearization (pre-distortion) of a loudspeaker system.

In the pre-distortion and post-distortion schemes, the forward linear transfer function of a physical system is sometimes available or has to be obtained anyhow. So the inverse transfer function can be computed directly from the available forward transfer function of the system. Suppose that D_i indicates the Fourier transform of the delayed impulse signal at the ith frequency point and H_i the Fourier transform of the forward transfer function of the system. The inverse transfer function H_i at frequencies of interest can be obtained by computing D_i/H_i. The inverse transfer function $\hat{H}_i$ at frequencies of no interest can be set to any convenient value so that the time-domain impulse response of the transfer function is short.

The second scheme is to use a filter to block the desired signal at those frequencies of no interest where high attenuations exist in the forward transfer function and to pass, without significantly altering, the desired signal at other frequencies as shown in Fig.6.4. Then. the adaptive filter will produce a good inversion of the forward transfer function of the physical system at frequencies of interest. This method is similar to the model-reference adaptive systems (MRAS) in the control literature [1] and is similar to the inversion method mentioned in [2], where no discussion was provided on choice of the filter for the desired signal and the straight-forward inversion method was employed.

The third scheme is to employ frequency weighting, a technique widely used in design of conventional filters [3]. Consider the objective function with frequency-

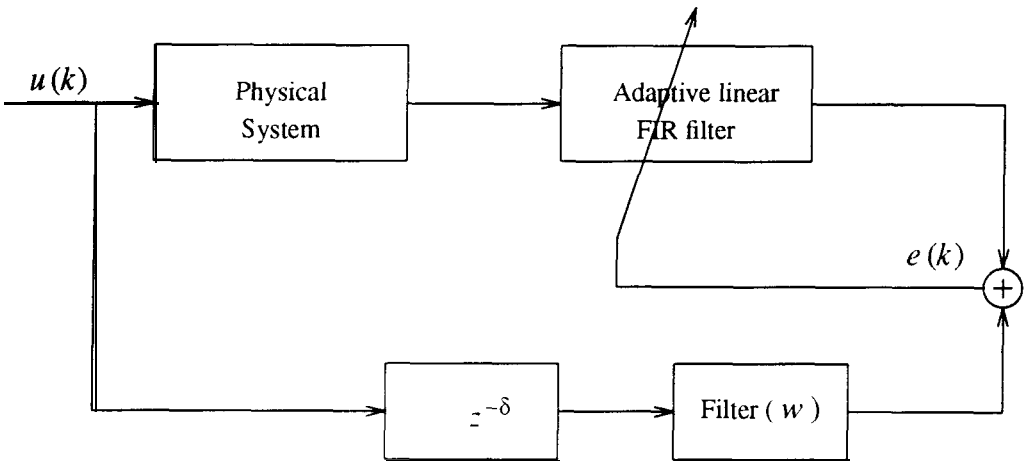


Fig.6.4 Inverse modeling with a filtered desired signal.

weighting

$$B = \sum_{i=1}^{N} (W_i E_i)^2 \tag{6.3}$$

where $E = Y - U_d$, Y is the Fourier transform of the filter output $y(k)$, U_d is the Fourier transform of the delayed input signal and W is a weighting function which is squared so that the time-domain formula has a simpler form. Using Parseval's theorem, the above equation is equal to the following one in the time-domain:

$$B = \sum_{i=1}^{N} (w(k-i) \otimes e(k-i))^2 \tag{6.4}$$

where $w(k)$ and $e(k)$ are the time-domain representations of W and E, and $\otimes$ indicates convolution. Equation (6.4) shows that the objective function is the energy of the filtered error signal $e(k)$. This method is illustrated in Fig.6.5. As in the LMS algorithm (where w (k) is just an impulse), the energy function can be approximated by its instantaneous value:

$$B = (w(k) \otimes e(k))^2 \tag{6.5}$$

It can be shown that the gradient of the objective function with respect to a coefficient p of the adaptive filter is

$$\frac{\partial B}{\partial p} = -2(w(k) \otimes e(k))(w(k) \otimes \frac{\partial y(k)}{\partial p}) \tag{6.6}$$

where y is the filter output. In the case of an adaptive linear FIR filter,

$$\frac{\partial B}{\partial h_i} = -2(w(k) \otimes e(k))(w(k) \otimes u(k-i))$$

$$= -2e_f(k)g(k-i) \tag{6.7}$$

where $e_f(k) = w(k) \otimes e(k)$ and $g(k) = w(k) \otimes u(k)$.

Although the first scheme employs the information already available and frequency-domain processing is often more efficient than time-domain processing, it requires a lot of complex divisions which may not be efficiently implemented on a DSP and it introduces a processing delay which is not acceptable in some applications. There are similarities between the second and the third methods. If in the second method (depicted in Fig.6.4) an extra filter identical to the filter for the desired signal is placed at the output of the adaptive filter, then the third method (depicted in Fig.6.5) results. The requirements of the filters for the two methods are different: the filter in the second method has to have a linear phase and flat response in the band of interest, but the filter of the third method does not have to. The third method (frequency-weighting) will be used in this chapter.

Human ears have high sensitivities in a narrow frequency band near 2 kHz which is responsible for almost all articulation in speech [4]. Using the frequency-weighting method, this requirement can be easily taken into account by assigning bigger weights

to the frequency points in the band. This is another advantage of the frequency-weighting method.

6.4 Identification by Adaptive FIR Filters

This section presents results on identification of the loudspeaker system by adaptive linear and nonlinear FIR filters using the measured data. For adaptive nonlinear FIR filters, the data length of $16k$ is not enough. The recorded signals were repeated once so that the length became $32k$. The recorded input signal was used as the input signal of an adaptive filter and the recorded output signal as the desired signal. For convenience, the input signal to the adaptive filters was delayed by 55 samples which is about the air-path delay. This is the case for all identification tests in the chapter.

Adaptive linear FIR filters with various orders were employed for a particular step size. The curve of MSE versus filter orders is shown in Fig.6.6 for the step size of 0.005. The MSE was calculated over $4k$ iterations (samples) to reduce fluctuations and the values shown in the figure were those at the $32k-th$ iteration. This step size was chosen so that the curve has a low minimum. The curve reaches its minimum of $-38.1dB$ for an order between 350 and 450. As the order increases beyond that, the MSE climbs due to mis-adjustment error. However, a high order filter with a properly chosen smaller step size is expected to further reduce MSE though more iterations are required. To confirm this, a step size of 0.0008 was used for the adaptive filter of order 700. It was run for three times using the set of $32k$ input-output samples. In the second and third runs. the starting point was the solution of the previous run. Then, the MSE

reached $-38.7dB$ at the end of the third run and it was smaller than the minimum of the curve in Fig.6.6. The impulse response and the transfer function are drawn in Figs.67 and 6.8 for the linear FIR filter obtained at the end of the third run. Comparing the impulse response and the transfer function with the ones in Figs.6.2 and 6.3, we see that the adaptive filter has identified the important features of the loudspeaker system. The mean square of the recorded system output signal was $-15dB$ and the best MSE by an adaptive linear FIR filter was about $-38.7dB$. This suggests that the nonlinearity is about 7 percent of the signal.

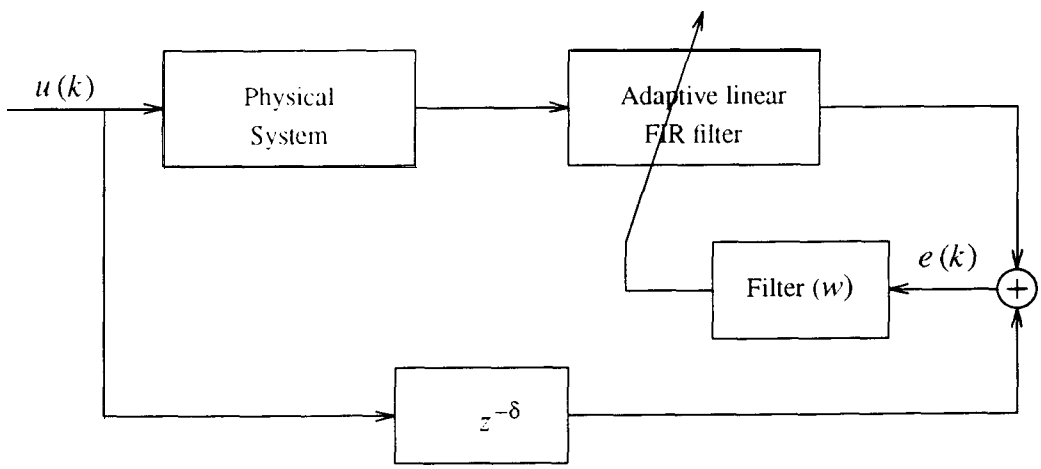


Fig.6.5 Inverse modeling with a filtered error (frequency weighting).

It was shown that an adaptive $700th$-order linear FIR filter with a step size of 0.0008 can model well the linearity of the system. This order and step size were used for the linear part of an adaptive nonlinear FIR filter in the following tests. An adaptive

quadratic filter *was* experimented with for different orders n_2. A sufficient order n_2, beyond which no improvement in MSE was observed, was found to be 400. After a few runs on the set of $32k$ input-output samples, the quadratic filter achieved an MSE of $-65dB$, a ***26dB*** improvement over that of the adaptive linear filter. A cubic filter with $n_1 = 700$ and $n_2 = 400$ was experimented with and it did not reduce MSE for any value of n_3 because the MSE was already very small.

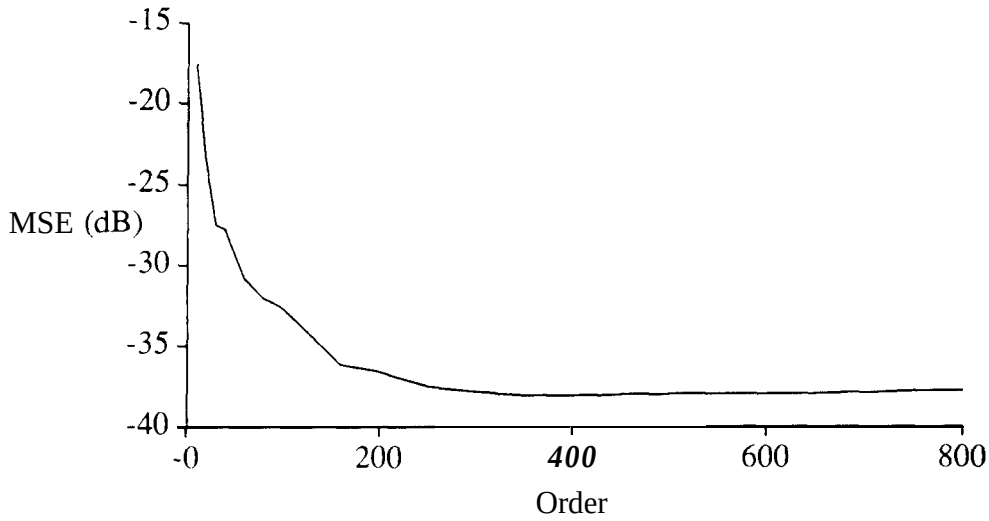


Fig.6.6 MSE versus order of an adaptive linear FIR filter.

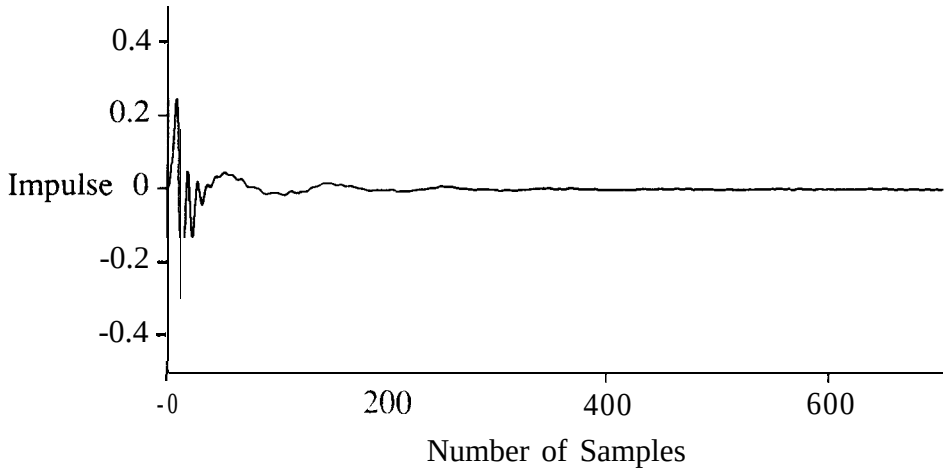


Fig.67 Impulse response of the adaptive linear FIR of order 700.

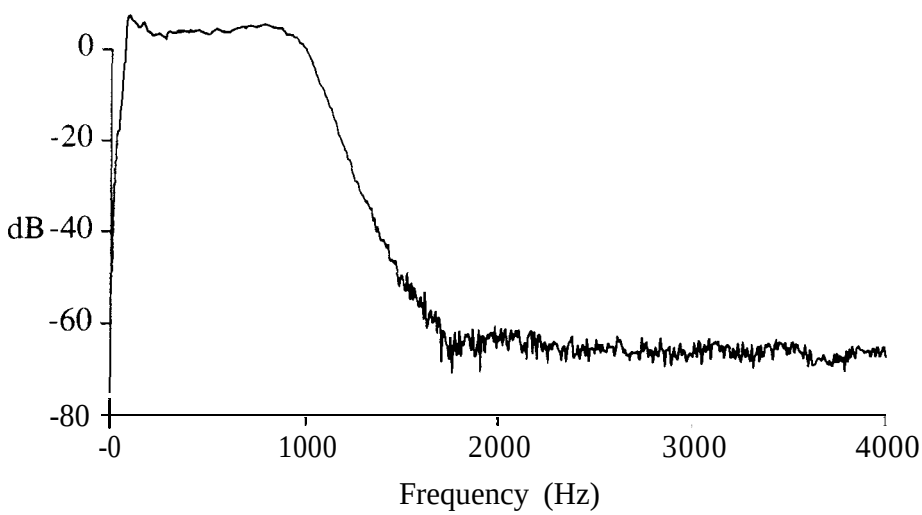


Fig.6.8 The transfer function of the adaptive linear FIR of order 700.

6.5 Identification by Adaptive IIR Filters

6.5.1 Adaptive Linear State-Space filter

The loudspeaker model presented in Chapter Four is of third order. Attempts were made to identify the loudspeaker system with an output-en-or adaptive linear filter based on this simple model. It was hoped that the dynamics of other parts, such as analogue filters, converters, and amplifiers, would be ignored by an adaptive filter. It was not surprising that an adaptive linear IIR filter based on either the form described in Equation (4.21) or third-order direct-form did not converge since order three is too low for this system.

All the components except the loudspeaker in the system typically have flat magnitude response and linear phase response for low frequencies, say below $500Hz$. If an adaptive filter just identifies the low-frequency behavior of the system, the third-order model may be sufficient. The frequency weighting technique was used for such a test. The weighting function was realized by a third-order low-pass Butterworth filter with a cut-off frequency at $400Hz$. With this frequency-weighting, a third-order adaptive IIR filter still did not converge.

Then, an adaptive linear recursive state-space filter was experimented with for different orders. It had the direct form and was based on the output-error formulation. All elements of the input vector B were fixed to zero, except the last element which was fixed to one. The output vector C, the feedthrough coefficient d, and the last row of the feedback man-ix A were adapted. Frequency-weighting was not used.

Fig.6.9 shows the MSE versus order of an adaptive linear state-space filter for step sizes $\mu_A = 0.005$. $\mu_c = 0.01$, and $\mu_d = 0.01$. The MSE decreases as the order increases and the curve flattens when the order reaches 29.

An adaptive linear state-space filter achieved the best MSE of about $-35dB$, ***4dB*** worse than that of an adaptive FIR filter. The difference may be due to echoes. These echoes can be easily modeled by an adaptive linear FIR filter, but they are difficult to model with an adaptive IIR filter.

The impulse response and transfer function are plotted in Figs.6.10 and 6.11 for the 29*th*-order adaptive linear state-space filter. These plots are smoother than the corresponding ones of both the measured data and the adaptive linear FIR filter.

6.5.2 Adaptive Equation-Error Filter

An adaptive equation-error filter was used to identify the system. The curve of MSE versus order is drawn in Fig.6.12 for a step size μ=0.1. The transfer function of the 29*th*-order equation-error filter is shown in Fig.6.13. When an order was high (greater than 29), the minimum MSE achieved by an equation-error filter was always slightly worse than that of a state-space filter of the same order. This is probably due to the fact that noise and nonlinearities in the measurement data have bias effects on an equation-error filter [5].

6.5.3 Adaptive Backpropagation Cascade Filter

A backpropagation cascade filter was applied to identify the system. The curve of MSE versus order is plotted in Fig.6.14 for step sizes $\mu_{a_i} = 0.005$ and $\mu_{b_i} = 0.005$. The

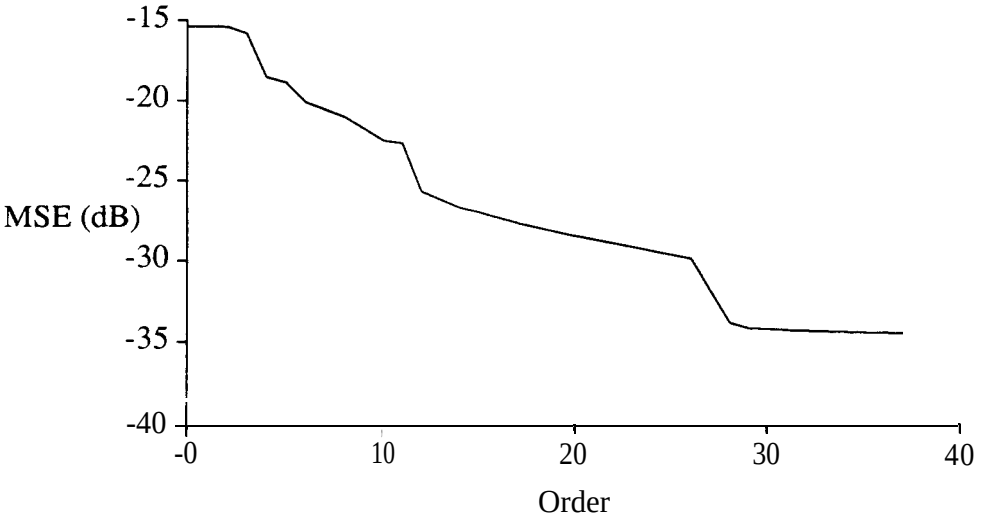


Fig.6.9 MSE versus order of an adaptive linear state-space filter.

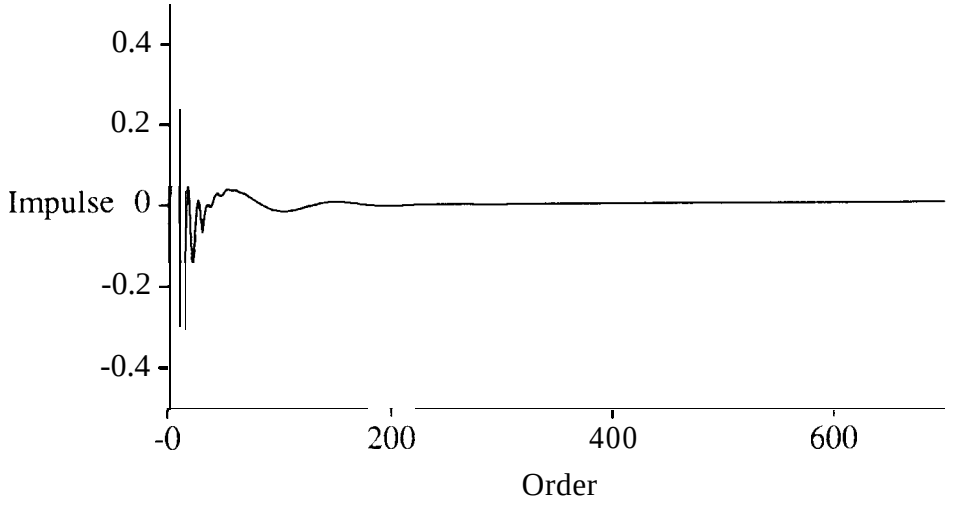


Fig.6.10 Impulse response of the 29th-order adaptive linear state-space filter,

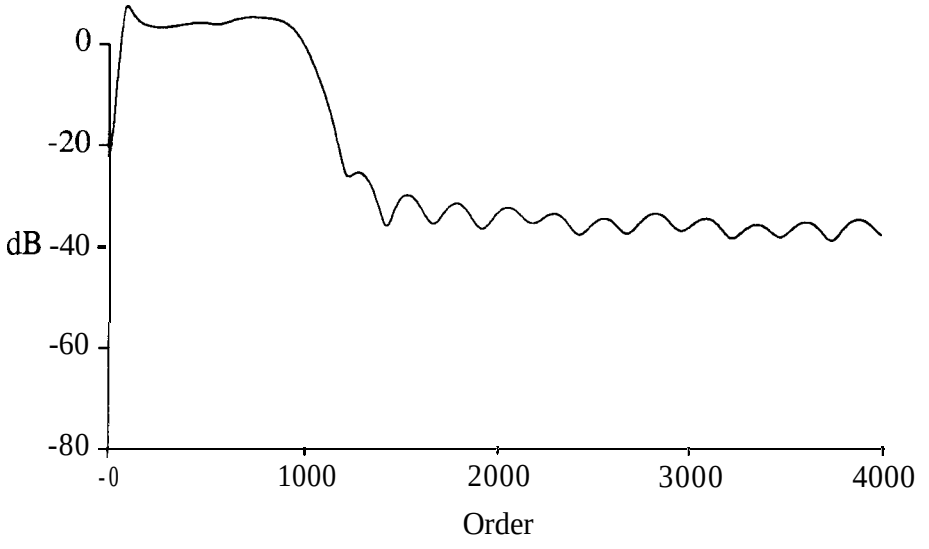


Fig.6.11 Transfer function of the 29th-order adaptive linear state-space filter.

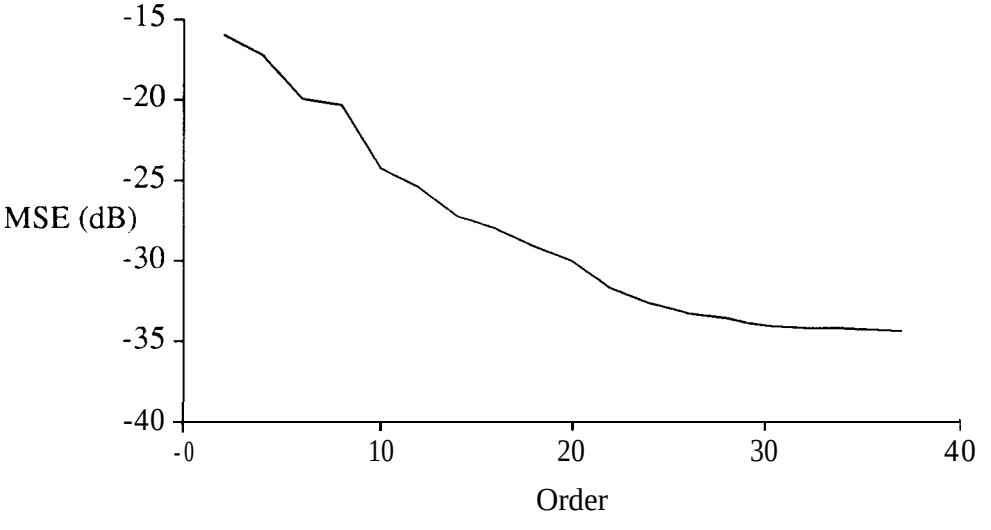


Fig.6.12 MSE versus order of an adaptive equation-error filter.

minimum MSE that it achieved was about -28 dB, 7 dB higher than that of the adaptive

linear state-space filter and equation-en-or filter. Different step sizes and starting points have been tried and the filter did not give better MSE results. The reason for this may be that the backpropagation algorithm is sensitive to the noise in the measured data (the nonlinearities in the data have the same effect as a noise). To see this, according to Fig.3.2 we can write

$$E_i(z) = E_i'(z) + \text{c; } \cdots C_m N \qquad i = 1,\ 2,\ \cdots, m \tag{6.8}$$

where N is the noise (including the nonlinear signal) and E_i' is the error signal without noise. The filter attempts to minimize the mean square of the true error signal plus a filtered noise.

The above results show that an adaptive linear IIR filter with a high order (about 30) was required to properly model the system. There are two major reasons for this. The first reason is that practical analogue systems have high orders. When we say that an analogue system is fourth-order, we just indicate the dominant dynamics and ignore those which are relatively insignificant. The second reason is that transformation from s domain to z domain is not order-preserving. It requires a digital system of an infinite order to represent precisely a first-order analogue system.

6.5.4 **ANRSS Filter**

Finally, an ANRSS filter was employed to identify the system. A nonlinear loudspeaker model was derived in Chapter Four and an ANRSS filter could be based on this model. However, the experiments presented above have shown that this low-order model was not good enough. It is difficult to answer the question: what structure should be used for an ANRSS filter? The form that we tried was the nonlinear direct-form.

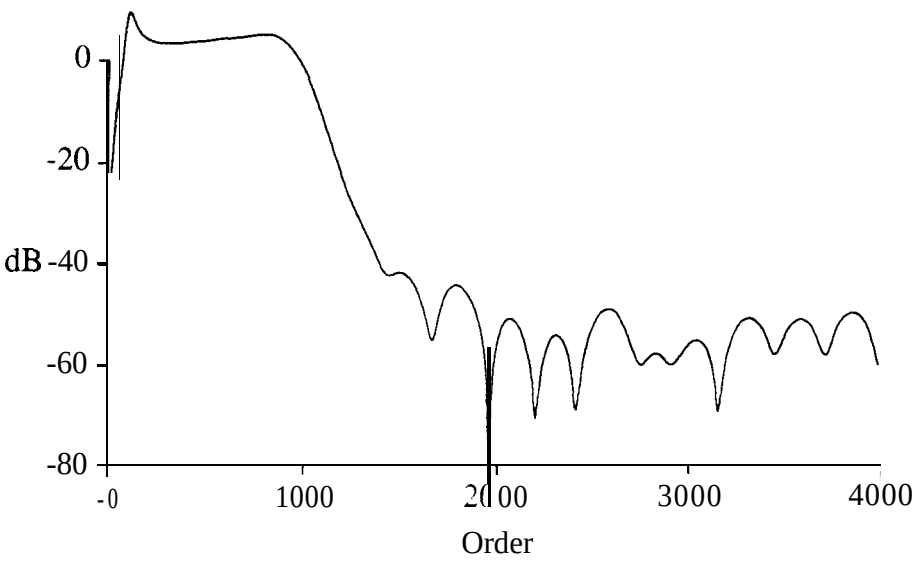


Fig.6.13 Transfer function of the 29th-order adaptive equation-error filter.

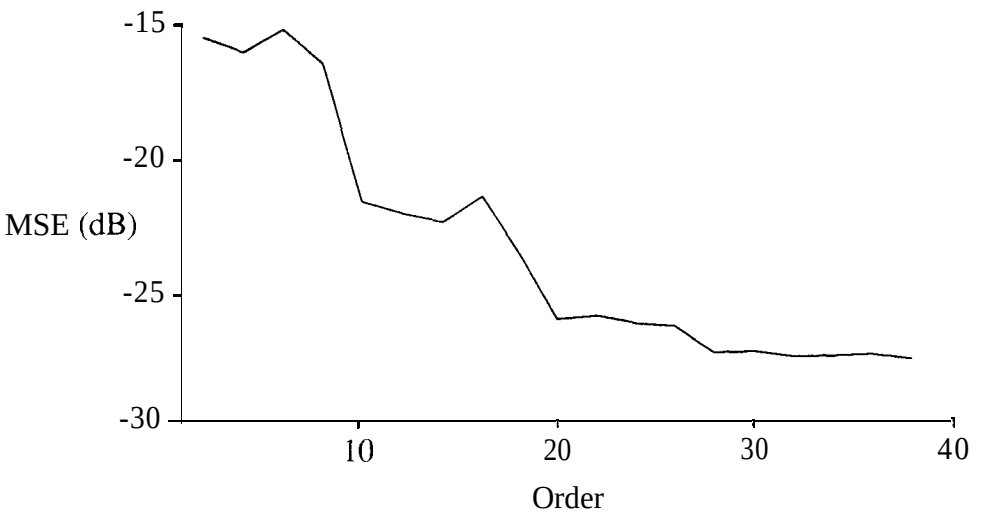


Fig.6.14 MSE versus order of an adaptive backpropagation filter.

Nonlinear terms were only quadratic (cross-product terms) and appeared only on the input side of the state variable $x_n(k+1)$, namely

$$\mathbf{g} = (\,0\;0\;\cdots\;g_n\,)^T \tag{6.9}$$

where $g_n = \sum_{i=0}^{n}\sum_{j=i}^{n} p_{ij}x_i(k)x_j(k)$ and x_0 is the input signal u. The input vector **B** was fixed to be $(0\;0\;\ldots\;1)^T$. The following parameters were adapted: the last row of the feedback matrix **A,** the output vector **C**, the feedthrough coefficient d, and the nonlinear coefficients p_{ij}.

The tests presented above showed that a $29th$-order linear state-space filter can model properly the linear part of the system. To get a good starting point, the $29th$-order linear state-space filter was performed for $32k$ iterations twice. The starting point of the second run was the solution of the first run. Then, the solution of second run provided initial values for the linear coefficients of the ANRSS filter. The ANRSS filter gave an MSE of $-35.5dB$ after a few runs on the set of $32k$ input-output samples. For a fair comparison, the $29th$-order linear state-space filter was performed a third time based on the solution of the second run mentioned above. The last two runs made little improvements in MSE and the best MSE achieved by the linear filter was $-34.6dB$, $0.9dB$ worse than that of the ANRSS filter. Since the MSE was computed over ***4k*** consecutive iterations, the fluctuations in estimates of MSE must be much smaller than $0.9dB$. This was confirmed by simulations. Estimates of MSE were computed from errors of ***4k*** consecutive iterations using the final filter obtained from the third run of the 29-th order adaptive linear state-space filter mentioned above. The filter was not adapted in the test. The standard deviation of the estimates was $0.023dB$, much smaller

than $0.9dB$. The same simulation was performed using the final filter obtained by the ANRSS filter discussed above. The standard deviation of the MSE estimates was 0.028, much smaller than $0.9dB$ too.

It is important to note that the ANRSS filter performed better than the linear filter. There might be two reasons that the improvement was not so significant. First, the effective number of independent samples was only $8k$ and it hardly provided enough information for an ANRSS filter to converge. Secondly, the model used for the ANRSS filter may not be a good choice.

6.6 Linearization

This section presents results on application of the pre-distortion technique to the loudspeaker system. The loudspeaker system was identified by an adaptive nonlinear FIR filter with orders $n_1 = 700$ and $n_2 = 400$. Then, the physical loudspeaker in Fig.4.7 was replaced by this extracted loudspeaker model.

Inverse modeling of the speaker was performed by an adaptive linear FIR filter with an order of 300. The frequency-weighted inverse modeling method was used. The weighting function is shown in Fig.6.16, which is a bandpass filter. The adaptive inverse linear filter obtained a reduction of 21 dB in the mean square of the filtered error after convergence. The transfer function obtained is plotted in Fig.6.17. The product (in dB) of the inverse transfer function and the forward transfer function (Fig.6.8) is shown in Fig.6.18. It is clear that the inverse function is an inversion of the loudspeaker transfer function in the band specified by the weighting function of

Fig.6.16.

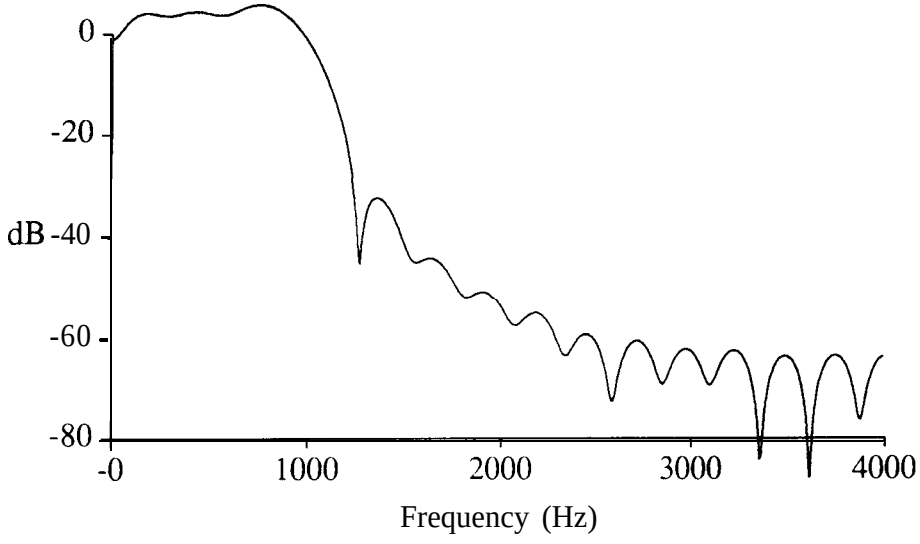


Fig.6.15 Transfer function of the 29th-order adaptive backpropagation filter.

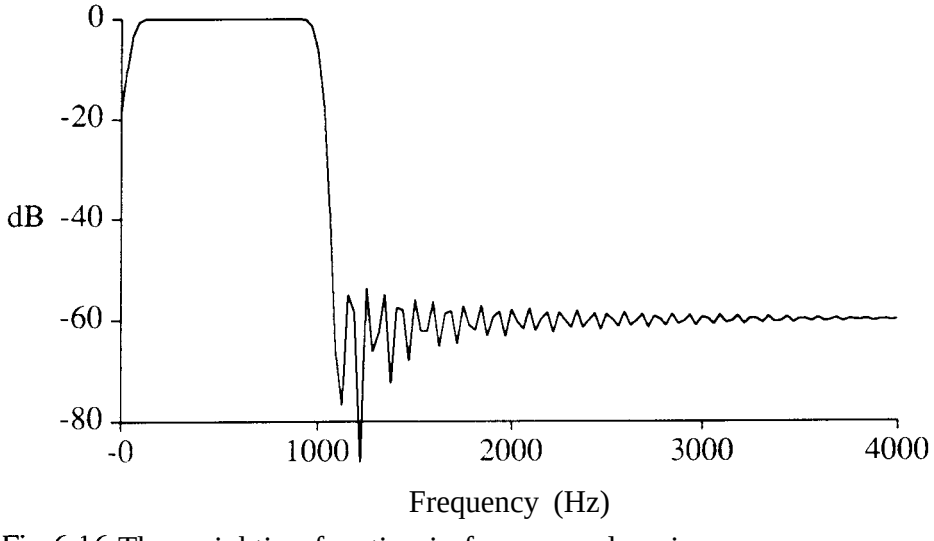


Fig.6.16 The weighting function in frequency domain.

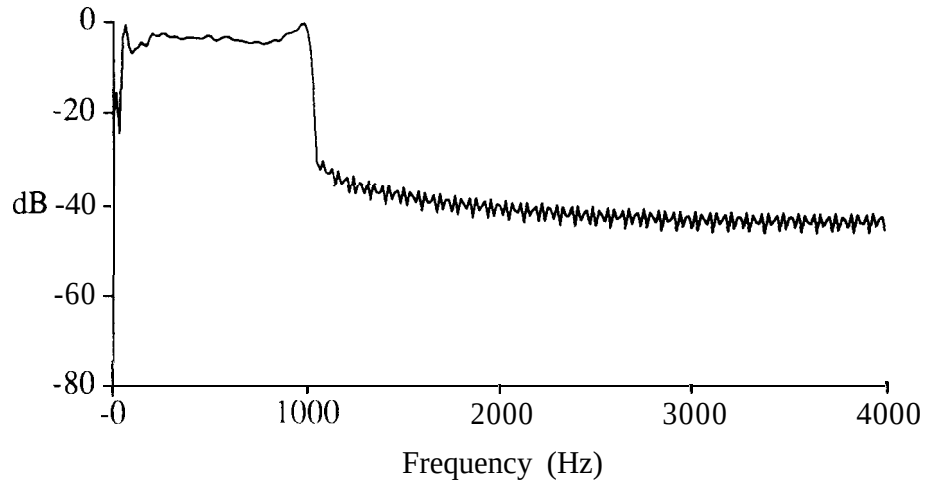


Fig.6.17 The transfer function of the adaptive linear inverse filter.

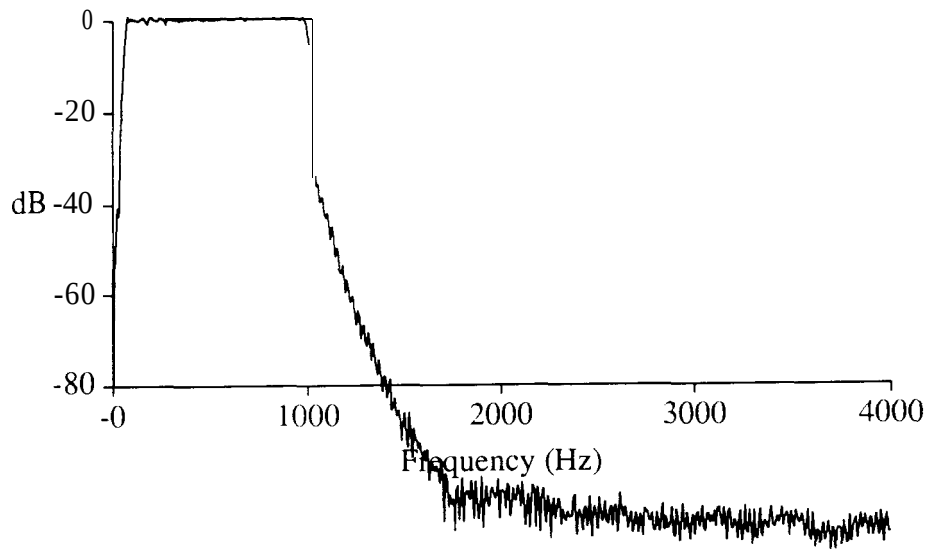


Fig.6.18 The product (in dB) of the transfer functions of the linear inverse filter (Fig.6.17) and the 700*th* linear filter (Fig.6.8).

The nonlinear pre-processor was constructed with the inverse linear operator obtained by the inverse filter and the nonlinear operator (nonlinear part) of the extracted loudspeaker model. The recorded input signal was used so that a proper ratio of linearity and nonlinearity was maintained. The pre-distortion technique enhanced the ratio of linearity to nonlinearity from $22dB$ to $36dB$, an improvement of $14dB$, which is substantial in practice. In this example, the reduction in distortion is smaller than those examples in Chapter Four because the inverse transfer obtained in this example was not as good.

4.7 Summary

This chapter is concerned with tests of adaptive filters on measurements of a loudspeaker system. Solutions for some practical issues. such as inversion of a transfer function with high attenuation regions, have been proposed. For comparison and preparation, existing techniques (adaptive linear FIR, nonlinear FIR, linear state-space, and equation-error filters) have been successfully used to identify the system. The results on the adaptive linear IIR filters and nonlinear FIR filters are valuable themselves since these adaptive filters have received extensive theoretical and simulation studies, but few reports are on practical applications. Then, a backpropagation cascade filter and an ANRSS filter were used in attempts to model the system. The backpropagation cascade filter did not reach the global minimum. This was probably due to its sensitivity to noise. A direct-form ANRSS filter was employed to identify the system and it made the MSE smaller than that of an adaptive linear state-space filter. The

improvement might be more significant if better data and/or a better filter structure are available. The pre-distortion technique was applied to linearize the model extracted from the measured data. The nonlinear distortion was reduced by about $14dB$, which is substantial.

References

[1] K.J. Astrom and B. Wittenmark. *"Adaptive Control,"* Don Mills, Ontario: Addison-Wesley Publishing Company, 1989.

[2] J. Kuriyama and Y. Furukawa, "Adaptive Loudspeaker System," *J. Audio Engineering Society*, Vol.37, pp.919-926, Nov. 1989.

[3] C. Ouslis, W.M. Snelgrove, and A. Sedra, "A Filter Designer's Filter Design Aid: Filtor X," *Pi-oc. of International Symposium on Circuits and Systems*, pp.376-*379*, Singapore, June 199 1.

[4] D. Davis and C. Davis. "Application of Speech Intelligibility to Sound Reinforcement," *J. Audio Engineering Society, Vol.37*, pp. 1002- 10 19. Nov. 1989.

[5] J.J. Shynk, "Adaptive IIR Filtering," *IEEE ASSP Magazine, pp.4 - 21*, April 1989.

[6] K.B. Benson (eds.), *Audio Engineering Handbook*, Toronto: McGraw-Hill Book Company, 1988.

[7] G.E.P. Box. W.G. Hunter. J.S. Hunter. "Statistics for Experimenters - An Introduction to Design, Data Analysis, Modeling Building," Toronto: John Wiley & Sons, 1978.

Chapter Seven

Conclusions and Suggestions for Future Work

7.1 Conclusions

The research work in this dissertation has advanced the state-of-the-art of both linear and nonlinear adaptive filters.

A novel technique of backpropagation of desired signal was developed. It was then applied to an adaptive linear cascade IIR filter, resulting in a structure with guaranteed stability and high computational efficiency. It was shown that the equation-error formulation is a special case of the idea of backpropagating the desired signal.

Three adaptive linearization schemes were devised for weakly nonlinear systems using adaptive FIR filters. The first scheme cancels nonlinearity at the output of a nonlinear physical system to be linearized. The second scheme post-distorts the output signal of a physical system and the third scheme pre-distorts the input signal. The scheme of linearization by cancellation at the output is able to accomplish virtually perfect linearization if the adaptive nonlinear filter gives a perfect identification of the physical system. Reduction in nonlinearities by other two schemes depends on the weakness of the nonlinearities: the weaker, the more reduction. Simulations have demonstrated that excellent linearization can be achieved by these schemes.

ANRSS filters have been developed, which are especially attractive for those applications with long memories where adaptive nonlinear FIR filters are too expensive

to use. The filters are recursive and thus generally have an infinite impulse response. To facilitate further their application in real-time signal processing, two efficient methods have been presented which significantly reduce computation for gradients. Several guidelines for maintaining filter stability were described. It was shown that the convergence speed depends on the eigenvalue spread of the correlation matrix of the gradient signals. Simulations on a first-order example showed that compared with an adaptive nonlinear FIR filter, an ANRSS filter needed only 0.4% of its computation, converged 13 times faster, and achieved a perfect matching of the system. Simulations on the ANRSS filters have shown the approximate stochastic-gradient method performed as well as the stochastic-gradient method in the case of the weak nonlinearities.

Measurement data were obtained on a loudspeaker system. Adaptive linear FIR, linear state-space, equation-error, nonlinear FIR, backpropagation cascade, and ANRSS filters were applied to identify the loudspeaker system using the data. Although adaptive nonlinear FIR and linear state-space filters are existing techniques, the results obtained are interesting since few results were previously reported on practical applications. A backpropagation cascade filter did not reach its global minimum probably because of its sensitivity to noise. An ANRSS filter made an improvement in MSE over an adaptive linear state-space filter. The improvement should be more significant if a better choice of filter structure and a better set of measurement data are available. The pre-distortion technique was applied to the loudspeaker model extracted from the measured data and a significant reduction of nonlinearity was achieved.

7.2 Suggestions for Future Work

Performance of the backpropagation cascade IIR filter may be further investigated.

The linearization scheme with a post-processor may be applied to equalization of a nonlinear data channel. where a linear equalizer can not perform well due to the presence of nonlinearities. The channel can be linearized first using the post-distortion scheme so that a linear equalizer is able to function better.

The stability problem of an ANRSS filter should be investigated further and efficient stability monitoring techniques may be developed. The hyperstability concept may be used to devise an adaptive nonlinear filter with guaranteed stability and global convergence for a nonlinear direct-form filter, which is of IIR type and is a special case of ANRSS filters.

Further, it is necessary to understand which set of parameters of an ANRSS filter should be adjusted for a given problem so that the adaptation can be successful. If all the elements of the feedback matrix A, the output coefficient vector **C**, the input coefficient vector B. and the nonlinear coefficient vector **p** are adjusted to minimize mean square of the output error, the adaptation may not succeed and error may not be minimized because there is a redundancy in terms of degrees of freedom in adaptation. Application of the ANRSS filters to practical problems should be further investigated and implementation of the filters should be experimented on real-time signal processing systems.

New measurements can be performed on a loudspeaker system with a signal generator capable of generating a long sequence of independent random signals. Then, an ANRSS filter can be applied to identify the loudspeaker system.