Adaptive Signal Processing

Theory and Applications

S. Thomas Alexander

Adaptive Signal Processing

Theory and Applications

S. Thomas Alexander

With 42 Illustrations

Springer-Verlag New York Berlin Heidelberg
London Paris Tokyo

S. Thomas Alexander
Department of Electrical and
Computer Engineering
North Carolina State University
Raleigh, NC 27695-7911
U.S.A.

Series Editor

David Gries
Department of Computer Science
Cornell University
Upson Hall
Ithaca, NY 14853
U.S.A.

Library of Congress Cataloging in Publication Data
Alexander, S. Thomas.
Adaptive signal processing.
(Texts and monographs in computer science)
Bibliography: p.
Includes index.
1. Adaptive signal processing. I. Title.
II. Series.
TK5102.5.A424 1986 621.38'043 86-13956

Typeset by Asco Trade Typesetting Ltd., Hong Kong.
Printed and bound by R.R. Donnelley & Sons, Harrisonburg, Virginia.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-96380-4 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-96380-4 Springer-Verlag Berlin Heidelberg New York

Preface

The creation of the text really began in 1976 with the author being involved with a group of researchers at Stanford University and the Naval Ocean Systems Center, San Diego. At that time, adaptive techniques were more laboratory (and mental) curiosities than the accepted and pervasive categories of signal processing that they have become. Over the last 10 years, adaptive filters have become standard components in telephony, data communications, and signal detection and tracking systems. Their use and consumer acceptance will undoubtedly only increase in the future.

The mathematical principles underlying adaptive signal processing were initially fascinating and were my first experience in seeing applied mathematics work for a paycheck. Since that time, the application of even more advanced mathematical techniques have kept the area of adaptive signal processing as exciting as those initial days. The text seeks to be a bridge between the open literature in the professional journals, which is usually quite concentrated, concise, and advanced, and the graduate classroom and research environment where underlying principles are often more important.

In that spirit, this text will be most beneficially used as an introductory tool for anyone interested in learning the fascinating field of adaptive signal processing. Most of the intended audience will be seniors and graduate students in electrical engineering or computer science, although the practicing engineer "gearing up" to work on product development using adaptive techniques will also find the text useful. An understanding of linear systems, digital signal processing, and matrix algebra approximately equivalent to that of an undergraduate electrical engineering curriculum is adequate. The text has been used for a graduate course in adaptive signal processing at North Carolina State University, in which students from a wide variety of backgrounds have actively participated.

The main distinction between this text and others that have appeared on the subject is the inclusion in this text of the timely subject of vector space approaches to fast adaptive filtering. This is currently one of the most active areas of research in signal processing, but the mathematical sophistication required to understand the open literature in the area has been formidable. This text develops the vector space approach through the liberal use of geometrical analogies, which encompasses Chapters 9–11. In so doing, the vector space approach becomes actually very easy to understand and possesses a great deal of simple elegance. After completing these chapters, the reader will be well prepared to tackle some of the more specific research problems associated with fast adaptive techniques. The book can be an effective text for a one-semester course in adaptive signal processing or as a reference for the researcher (academic or industrial) to absorb material at his own pace. Problems that have survived the classroom experience are included at the end of the chapters.

This text is approximately the same process by which I became familiar with the different areas of adaptive signal processing. Much of the original material was literally "back of the envelope" information from hearing conference talks and informal discussions. Other portions had their beginning as notes scribbled in the margin of books or papers when something finally jelled.

As with any book, the contributions of many people over many years were instrumental to the entire process. Specifically, I would like to thank the following: Bob Plemmons of North Carolina State for sharing his mastery of linear algebra; Lloyd Griffiths of USC for long runs during which philosophies were discussed; Ed Satorius of JPL and Joel Trussell of North Carolina State for consistently providing honest and, therefore, valuable technical evaluation and discussion on both this text and the Big Picture; John Cioffi of Stanford for his contributions to my own understanding of geometrical approaches and fast adaptive techniques; Nino Masnari, Chairman of the Electrical and Computer Engineering Department at North Carolina State, for helping to foster the professional environment that allowed the time and resources to develop this text; and the editorial staff at Springer-Verlag for providing the assistance and encouragement I needed to successfully complete the manuscript. Additionally, for their unsung efforts in reading and debugging the original drafts and homework problems, I would like to thank the following at North Carolina State: Gary Ybarra, Glenda Poston, Zong Rhee, Daehoon Kim, and Randy Avent. Special thanks are also extended to Peggy Ball, Liz Story, George Winston, and Red Ryder.

Finally, the city of Boston and the season of Winter had a lot to do with the whole process.

Raleigh, NC S.T. ALEXANDER

Contents

CHAPTER 1
Introduction

1.1 Signal Processing in Unknown Environments

Many everyday problems encountered in communications and signal processing involve removing noise and distortion due to physical processes that are time varying or unknown or possibly both. These types of processes represent some of the most difficult problems in transmitting and receiving information. The area of adaptive signal processing techniques provides one approach for removing distortion in communications, as well as extracting information about unknown physical processes. A short consideration of some of these problems shows that distortion is often present regardless of whether the communication is conversation between people or data between physical devices.

For example, a common problem in long distance telephone communications is the creation of echoes due to impedance mismatches on the network. This has an extremely annoying effect on the persons using the telephone link, and can degrade the quality of communication such that the conversation is rendered unsatisfactory.

Another example is that of computers exchanging data over physical communications channels. Many channels are well conditioned and deliver the original transmitted pulses undistorted to the receiver. However, many channels are poorly conditioned and distort the received digital pulses to such a degree that data decision devices would make far too many errors to provide a useful service.

While there are numerous other examples, these two are sufficient to illustrate some of the main reasons for needing adaptive signal processors. In the first example above, the impedance mismatch is usually unknown. That is, the sheer number of local telephone lines that must be accessed effectively prohibits the impedance for any one local line to be accurately matched to the long distance link. Even if the resources were available to match the impedance of each local line to the long distance link, there is still the problem that due to aging, inaccurate component values, moisture, etc., the impedance of each local line may be time varying. Therefore, attempts to build a single processor that has the flexibility of addressing all these time-varying and unknown phenomena require adaptivity in the processor. Such adaptive signal processing devices are widely in use now and are known as adaptive echo cancellers.

In the second example above concerning data transmission, the unknown and/or time-varying segments is the communications channel itself. For example, with the proliferation of mobile radios, there has emerged the possibility of transmitting and receiving data from highly mobile stations to a central computer database. Consider the case in which the transmitter is mobile (i.e., located in a car). Since the data is encoded and sent over the atmospheric radio channel, the propagation path between the transmitter and fixed receiver is changing, sometimes quite rapidly. Therefore, such considerations as data symbol timing, power of received signal, and propagation loss, for example, are time varying and unknown to the system. Once again, designing a fixed parameter system to handle the wide range of *possible* values of these parameters could render the system performance unacceptable for certain specifically encountered situations.

The common element in each of the preceding problems, and indeed in most of the applications of adaptive signal processing, is that some element of the problem is unknown and must therefore be learned, or some component of the system is changing in an unknown manner and therefore must be tracked. Quite frequently, both of these problems are resident in the applications of adaptive signal processing.

1.2 Two Examples

Two general examples were discussed in the previous section. This section will discuss two additional examples in more detail and provide a mathematical framework for adaptive signal processing. These two applications examples will be used throughout the text to illustrate new concepts and to investigate the performance characteristics of various adaptive methods. The first example is known as systems identification and the second will be referred to as linear prediction.

Systems identification

The first example concerns systems identification, which is used quite frequently in controls and communications work. Consider the case of Figure 1.1, in which it is desired to learn the structure of an unknown system from a

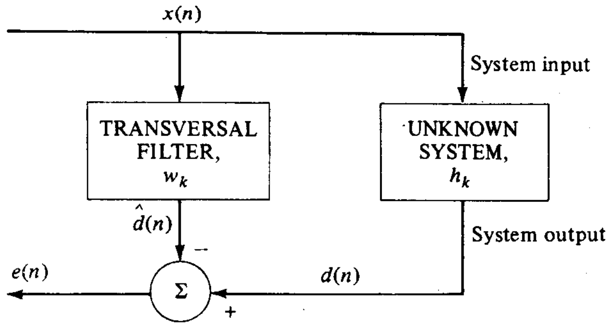

Figure 1.1 Identification of unknown system, h_k, using transversal filter, w_k.

knowledge of its input $x(n)$ and output $d(n)$. For example, it may be necessary to determine if any of the system parameters are approaching critical values. Although there are several way of quantifying the "knowledge of a system," the system impulse response, h_k, is often used. From elementary linear systems,

$$d(n) = \sum_{k=0}^{n} h_k x(n-k), \tag{1.2.1}$$

where both the input signal and the impulse response have been assumed to be casual. In this problem, the true impulse response h_k is unknown and must be obtained.

The form of (1.2.1) suggests the following approach to "learning" the h_k. Assume the h_k form a finite impulse response of no more than N samples in duration, counting h_0. Then,

$$d(n) = \sum_{k=0}^{N-1} h_k x(n-k). \tag{1.2.2}$$

This is the true system output $d(n)$. Using the form suggested by (1.2.2), a prediction of $d(n)$, denoted as $\hat{d}(n)$, may be made using a set of filter coefficients w_k:

$$\hat{d}(n) = \sum_{k=0}^{N-1} w_k x(n-k). \tag{1.2.3}$$

Strictly speaking, (1.2.3) is an estimation of the signal $d(n)$. However, much of the current literature refers to a form such as (1.2.3) as a prediction of $d(n)$, and this terminology will be used in this book.

If each chosen w_k is "close" to each true h_k, then the prediction error,

$$e(n) = d(n) - \hat{d}(n) \tag{1.2.4}$$

should be small in magnitude. In systems identification, the rationale is that if $\hat{d}(n) \approx d(n)$, then $w_k \approx h_k$. Therefore, minimizing some measure of $e(n)$ should force the individual w_k to approach the individual h_k, thus identifying the

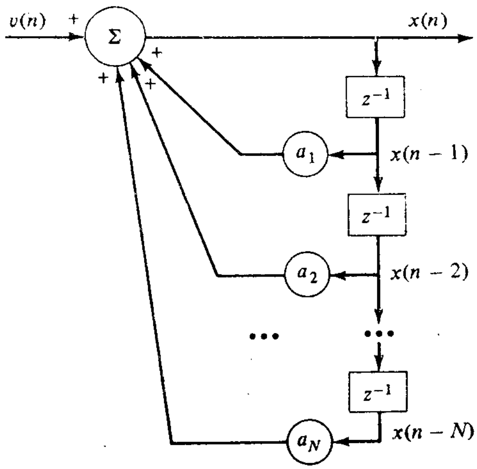

Figure 1.2 Information signal $x(n)$ produced by Nth order autoregressive (AR) process.

system. In Chapter 2, the mean square error will be seen to be a natural measure that is amenable to mathematical analysis, as well.

Linear prediction

The second example is very common in speech analysis and telecommunications. Consider Figure 1.2, which shows a model of a process that generates an information signal. The output of the model is the current sample $x(n)$, and it is easy to see that $x(n)$ is given by

$$x(n) = \sum_{k=1}^{N} a_k x(n-k) + v(n), \tag{1.2.5}$$

where the a_k are the unknown system parameters and $v(n)$ is an unknown random excitation sequence. In statistical literature, (1.2.5) is called an Nth order autoregressive (AR) process. The model structure in (1.2.5) is analogous to a signal produced through a series of reflecting/transmitting media. This is a very good model for speech produced in the vocal tract, seismic signals propagating through a layered earth, and certain types of electromagnetic reflections in radar applications.

The form of the model in (1.2.5) is an approach similar to the preceding example for "learning" the generating parameters of the unknown system. That is, form a linear prediction of $x(n)$ based upon the N most recent $x(n-1)$, ..., $x(n-N)$:

$$\hat{x}(n) = \sum_{k=1}^{N} w_k x(n-k). \tag{1.2.6}$$

Since the excitation sequence $v(n)$ is unknown, it clearly cannot be used in the prediction (1.2.6). The prediction error then becomes

$$e(n) = x(n) - \hat{x}(n). \qquad (1.2.7)$$

In this text, predicting a signal sample based upon previous values of the same signal will be called the purely linear prediction problem or, more simply, the linear prediction problem. Since the excitation sequence $v(n)$ is unknown, it clearly cannot be used in the prediction (1.2.6).

Once more, the rationale is that if the mean square of $e(n)$ is small, then the w_k of the filter will be forced to the a_k of the signal model. Sometimes in speech communications problems, the object is to identify and transmit the a_k. Other telecommunications applications might transmit information about $e(n)$. These applications will be discussed in more detail as they naturally arise during the course of the text.

In either of the preceding examples, the filter coefficients w_k may be constant if the unknown process is at least statistically stationary. This will be referred to as the fixed filter case. However, if the system to be identified in (1.2.1) or the information process in (1.2.5) has parameters that change in an unknown manner, then adaptivity of the filter coefficients is necessary. These two examples of systems identification and linear prediction will be referred to frequently throughout the text in more detail. For additional applications of adaptive filtering, as well as other approaches to the subject of adaptive signal processing, the reader is directed to references [1] through [9].

1.3 Outline of the Text

One purpose of this text is to develop in a cohesive, structured approach some of the more useful and promising adaptive signal processing techniques. This approach will be designed to accurately display the common foundations of these methods, but it will also illuminate the differences between the candidate adaptive approaches. This should allow the systems designer the tools for selecting the appropriate approach to the problem at hand, given the engineering constraints of memory, speed, and cost.

Another purpose is to provide the reader with a well-founded physical and geometrical understanding of the adaptive signal processing methods. It is sometimes tempting to launch immediately into mathematical derivations without providing the often necessary physical understanding of the adaptive processes. However, this text strives to demonstrate physical or geometrical interpretation whenever possible, thus providing another tool for understanding adaptive methods. To this end, a general outline of the text is as follows.

Since the concept of minimum mean square error (MSE) is prerequisite to understanding much of modern adaptive signal processing, Chapter 2 develops a foundation for determining minimum MSE filters. Necessary probabilistic and statistical concepts are also introduced as they are needed in this chapter. In addition, the very important normal equations are derived in Chapter 2. The concept of a "bowl-shaped" error surface is developed, as well as the geometrical analogy of locating the minimum of this surface.

Having thus formulated the minimum MSE problem, Chapter 3 then explores one very important method of solving the normal equations that result in the very important application of linear prediction filtering. This leads to the Durbin recursion, which in turn leads to the lattice filter structure. This lattice structure will be seen to be very useful in many approaches to linear prediction and adaptive signal processing and provides an alternative to the transversal filter implementation.

Another important iterative approach for solving the normal equations is developed in Chapter 4. This is the gradient-based technique known as the method of steepest descent. A derivation of the convergence properties of this method is presented, as well as the valuable geometrical analogy of "finding the bottom of the bowl."

Chapter 5 then makes the transition to the least mean squares (LMS) algorithm, which computes an approximation to the method of steepest descent. The analytical convergence properties of the LMS algorithm are developed in detail in this chapter. Additionally, the differences between LMS and steepest descent are discussed in detail, as well as considerations for using the LMS algorithm in actual systems.

Two application examples, known as systems identification and linear prediction, are developed in Chapters 2–5. Chapter 6 then provides some additional applications of adaptive filtering, which give more insight into actual systems usage. The applications of Chapter 6 are all done using the popular LMS algorithm.

Chapter 7 then discusses some adaptive approaches based upon Durbin's algorithm that converge more rapidly than LMS for most applications. Known collectively as gradient-based lattice techniques, these approaches create sets of orthogonal signals from the acquired data signal, which are then used in efficient updating algorithms.

The modern area of recursive least squares (RLS) adaptive filters is next introduced in Chapter 8. The specific method of Chapter 8 is the regular RLS method, which lays the foundation for understanding the extremely rapidly converging techniques of modern least squares filtering. The RLS method investigated is different from the gradient-based methods examined thus far, in that it computes the optimal least squares prediction at every point in time. Gradient-based methods, such as LMS, are only optimal at convergence. This added performance capability is not without cost, however, since the regular RLS requires substantially more computation than LMS.

However, there are newly derived methods of reducing the required computations in RLS filters, which are collectively known as fast RLS techniques. Chapter 9 recasts the RLS problem by structuring it as a minimization problem in a Hilbert space, which has some very beneficial geometrical interpretations. This leads to the very powerful vector space approach to fast adaptive filters, which can be applied to either the fast lattice or fast transversal implementations.

Chapter 10 then applies these vector space concepts to the derivation of the fast least squares lattice (LSL) filter for linear prediction. The geometrical interpretations of the LSL are emphasized in this chapter.

Chapter 11 then applies the geometrical concepts of Chapter 9 and derives the least squares fast transversal filter (FTF), which is the transversal counterpart of the LSL implementation in Chapter 10. The FTF has the fewest arithmetic operations per time update of any least squares method derived to date. In Chapters 10 and 11, it will be seen that the LSL and FTF approaches usually converge much more rapidly than the gradient-based adaptive filters. Since their computational complexity is comparable to LMS, they are indeed worthy of consideration for many modern applications.

References

1. S. Haykin, *Introduction to Adaptive Filters*, Macmillan, New York, 1984.
2. R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*, Wiley-Interscience, New York, 1980.
3. F. Hsu and A.A. Giordano, *Least Squares Signal Processing*, John Wiley & Sons, New York, 1984.
4. B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
5. J.G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1984.
6. S.J. Orfanidis, *Optimum Signal Processing*, Macmillan, New York, 1985.
7. G.C. Goodwin and K.S. Sin, *Adaptive Filtering, Prediction, and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
8. M.L. Honig and D.G. Messerschmitt, *Adaptive Filters*, Kluwer Academic Publishers, Hingham, MA, 1984.
9. B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

CHAPTER 2

The Mean Square Error (MSE) Performance Criteria

2.1 Introduction

Adaptive signal processing algorithms generally attempt to optimize a performance measure that is a function of the unknown parameters to be identified. The most pervasive of these performance measures are based upon squared prediction errors, although the specific prediction error used in adaptation often depends upon the particular algorithm. Two broad categories of adaptive signal processing methods are: (1) stochastic and (2) exact. The latter category refers to adaptive filters based upon the actual or exact data signals acquired. The recursive least squares techniques comprising Chapters 8–11 are examples of these exact techniques, and investigation of those techniques will be deferred until the later chapters.

The former category of adaptive techniques known collectively as stochastic methods are based upon derivations using the statistical properties of the data signals. The primary statistical measure used is the ensemble average, or mean, of a squared prediction error function, and this has evolved into widespread use of the mean squared prediction error as a performance measure. Often this is shortened to simply the mean square error (MSE).

Many of the properties of minimum MSE filters and the MSE surface are derived using basic linear algebra techniques, such as eigenvalue and eigenvector analysis. Excellent texts on linear algebra at the introductory level are those by Anton [1] and Strang [2]. Another text that is particularly strong in geometrical interpretations is by Moore [3]. At the somewhat more advanced level is the text by Noble and Daniel [4], which is written largely from an engineering and physical science standpoint. As a result, the development of eigenvalues and eigenvectors is superb and gives excellent physical analogies. A more expanded treatment of physical considerations is given in an earlier edition by Noble [5]. The book by Bellman [6] is a somewhat more mathematically oriented text, but is still very good for the engineering student and contains a number of worked examples. For the active researcher in mean square prediction, filtering, and estimation, the text by Fadeev and Fadeeva [7] offers insight into both the theoretical and computational aspects of matrix and eigensystem analysis. Additionally, Golub and Van Loan [17] is an excellent text for advanced matrix analysis and computational considerations.

Linear prediction techniques using the MSE criteria have been applied extensively in the speech research community. Many of the speech applications, as well as early general theoretical work, are contained in the book by Markel and Gray [8]. An excellent journal article that displays the flexibility of linear prediction to diverse applications is that by Makhoul [9]. Since the autocorrelation function plays such an important role in linear prediction and MSE analysis, the paper by Markel and Gray [10] analyzes in detail its impact. The text by Rabiner and Schafer [11] develops approximations to the autocorrelation function and applies them to the problem of linear prediction of speech. Finally, the excellent text by Jayant and Noll [12] illustrates the use of random process theory in many techniques chosen from the areas of speech and image coding.

Minimization of the MSE is the objective of many currently used adaptive methods, such as the least mean square (LMS) algorithm and the gradient lattice method. These techniques are the topics of Chapters 4–7. In the current chapter, the basics of minimizing the MSE using the techniques of linear prediction filtering are introduced. Section 2.2 develops the concept of a quadratic error surface, which has the simple geometrical property of a single, or global, minimum. Section 2.3 then discusses some additional properties of the error surface, which will be useful and important for relating the error surface to physical phenomena. Section 2.4 then derives the relation known as the normal equations, which defines the location of this global minimum. Section 2.5 then concludes with a discussion of some of the geometrical properties of the error surface, which aids in understanding the dynamic properties of the adaptive methods.

2.2 Mean Square Error (MSE) and MSE Surface

A rationale for the MSE as a performance measure is perhaps best illustrated by example. In the general case, it is sometimes desired to predict the current sample of one signal, $d(n)$, using samples of a second signal, $x(n)$. An example of this case was the systems identification application of Figure 1.1. In this text, $d(n)$ will be called the desired signal and $x(n)$ will be called the data signal, since it is desired to predict $d(n)$ using $x(n)$. Ideally, the prediction filter output