2nd Edition

TIME SERIES ANALYSIS, IDENTIFICATION AND ADAPTIVE FILTERING

2nd Edition

TIME SERIES ANALYSIS, IDENTIFICATION AND ADAPTIVE FILTERING

ROBERT E. KRIEGER PUBLISHING COMPANY
MALABAR, FLORIDA
1989

Original Edition 1984
Second Edition 1989

Printed and Published by
ROBERT E. KRIEGER PUBLISHING CO.
KRIEGER DRIVE
MALABAR, FLORIDA 32950

Library of Congress Cataloging-in-Publication Data

Graupe, Daniel
Time series analysis: identification and adaptive filtering / Daniel Graupe.—2nd ed.
p. cm.
Includes bibliographies and indexes.
ISBN 0-89464-315-0 (alk. paper)
1. Time-series analysis. 2. Filters (Mathematics) I. Title.
QA280.G74 1989
519.5'5—dc19 88-14048
CIP

10 9 8 7 6 5 4 3 2

Preface to the Second Edition

The second edition of this text differs from the first edition in that it includes a new chapter on the identification of time series with inherently time varying parameters, where piece-wise stationarity cannot be assumed. This is the new Chapter 9. Consequently, the old chapters 9 to 12 become chapters 11 to 13, respectively. Also, a new subsection (10.7) was added to Chapter 10 (previously Chapter 9), to discuss artificial decision cells for certain adaptive decision problems with a "teacher." Furthermore, a second case study, dealing with the application of time series identification to the control of electrical neuromuscular stimulation of paraplegics, has been added to Appendix B, as were additional problems and computer programs for various chapters.

The second edition also includes corrections to typos and errors which unfortunately appeared in the first edition whose typesetting was not done with a word processor and for which the author deeply apologizes. The author is grateful to many colleagues and students for pointing out many of these typos and errors. Specific thanks are due to Professor Zehuan Liang of Academic Sinica, Beijing, China, who also translated the first edition to Chinese, to Mr. Alvin Moser of the University of Illinois at Chicago and to Mr. James Phillips of Motorola, Schaumburg, Illinois for bringing to the author's attention many typos. Finally, the author wishes to thank Dr. Wai-Kai Chen, Head of the Department of Electrical Engineering at the University of Illinois at Chicago, and to Dr. Paul Chung, Dean of the School of Engineering at the University of Illinois at Chicago, for providing him with an inspiring environment for his research and for their warm support throughout his work on this edition of the text.

Preface to the First Edition

. The advent of microprocessor technology has brought rigorous time series analysis from the theorist's library to industry, communications and medicine, *via* its applications to adaptive control, adaptive filtering and adaptive decision theory. This text presents a time-domain time series analysis applied to signal processing. Its emphasis is on situations of no *a priori* signal-parameter knowledge, of noisy measurements and of microcomputer short word-length, which are the most likely "real-world" situations. The book attempts to unify linear time series analysis, adaptive filtering, and adaptive control, as well as adaptive decision and signature-discrimination or diagnosis theory. Problems of convergence, convergence rate, bias and sensitivity to finite computer word-length. which are all essential to such analysis and applications, are carefully and rigorously discussed. The analysis is accompanied with a review of stochastic convergence theory, including an outline of 33 fundamental martingale and other convergence theorems. Considerations of real-time computerized applications of the theory to adaptive filtering of noise from information signals such as speech are discussed. Also, a practical microcomputer application of time-series analysis to controlling of prostheses for amputees and of electrical stimulation of paraplegics *via* myoelectric signal signature analysis, is presented in detail.

The text is based on a one semester, 3-credit, graduate level course given by the author at the Electrical Engineering Department of Illinois Institute of Technology, Chicago, Illinois. It is directed to graduate engineers and applied mathematicians in industry, and in research laboratories, who deal with problems of (adaptive) control, communications and (adaptive) signal processing in engineering, medicine, geology, and econometrics.

A background in the fundamentals of probability and statistics and in matrix theory and linear control theory is assumed. However, the text attempts to provide background material and complementary appendices for making the book accessible to graduate engineers whose background in these areas is rusty or vague. It is the author's hope that readers from industry, from research establishments and from universities will find the book useful.

Thirteen representative computer programs and several sub-versions thereof are appended to the text, in form of exact printouts of both the programs and their subsequent results. These cover all the major results of the text, so that the reader can re-run the programs to satisfy him/herself that the theory does indeed "work." Since the programs and the results are given as photo-copies of what was actually executed and run (on widely accessible computers), the reader can be confident that they are free of printing or other errors, as is so difficult to guarantee in other text material.

This text emanated from several runs of a graduate course at IIT, whose earlier versions I taught at Colorado State University, and during my sabbatical leave at the University of Notre Dame and at the University of California, Berkeley.

I am most grateful to my students in these courses. Their active participation, discussions and questions greatly helped me in shaping the text into its present form. In particular, I thank several of these students, who participated in my 1981 and 1983 EE 539 course at IIT, on the subject matter of this book, and who computed and ran several programs and examples given in the book. Those are, in alphabetical order: Joseph Bentzman, Farid El-Wailly, Peter Holterman, Henry Kazecki, Larry Paarmann and Steve Ruzinsky.

I am very indebted to Dr. Atsuhiko Noda of Tokyo Institute of Technology, to Dr. Eli Fogel of the Draper Labs., Cambridge, Mass., to Dr. Javad Salahi of Bell Labs, Holmdel, N.J. and to Mr. John Grosspietsch of IIT, for their careful and critical reading of the text, and for their excellent criticisms and detailed comments, which greatly helped to improve the text.

John Grosspietsch was of particular help in performing many of the derivations concerning the various lattice algorithms. It is he and Mr. Stavros Basseas of IIT who helped me by running many comparative computations of the lattice algorithms.

The financial support from the Applied Mathematics Program of ONR and the encouragement of Dr. Stuart Brodsky, its former director, to various of the author's research projects in areas of time series analysis and adaptive systems, is greatfully acknowledged.

The case study of Appendix B is the outcome of research formerly supported by NSF under the direction and encouragement of Norman Caplan, Program

Director, Automation Bioengineering, Computer and Systems Engineering at NSF.

I wish to thank Dr. Thomas Martin, President of IIT, Dr. Andre G. Vacroux, Dean of Engineering, and Dr. Shi-Kuo Chang, Chairman of the Electrical Engineering Department for the inspiring academic environment provided for me at IIT during this writing.

Special thanks are due to my sons, Henny Menahem Graupe and Pelleg Pinhas Graupe for editing and arranging the indices of the book. Henny also wrote several test programs for the text, whereas my son Oren sacrificed computer-games time for this purpose.

The interest and support of Mr. Robert E. Krieger, my publisher went far beyond what one would expect of a publisher.

It is with the greatest pleasure that I thank Ms. Betty Nessinger, Ms. Debbie Waddy and Ms. Barbara Skubiszewski for their tireless and devoted typing of this book, through its many drafts, and Ms. Mary Bishop for her careful type-setting of the book. Their patience with me and with my carelessness was beyond comprehension.

Last but not least, my endless thanks to my parents, for their inspiration and encouragement throughout this writing. Without the continual support and encouragement of my wife, and the patience and forgiveness of my children, this book would have never been written.

Contents

Chapter 1

Introduction, Goal and Outline

1.1 PRELIMINARY COMMENTS

In this text we concentrate on the analysis of discrete time series (namely, discrete sets of observations or measurements that evolve in time), when their properties or parameters are *a-priori* unknown. This lack of prior knowledge is a very realistic situation in signal processing, prediction and control in engineering, economics or in medicine. We shall concentrate on stochastic (probabilistic) time series rather than on deterministic ones, since the stochastic problem is the one of interest in the real world, and this is where the problems and the challanges challanges lie.

Well known examples of such sotchastic time series range from seismic data, to fluctuations of stock prices with time. Further examples are the myoelectric (EMG) signal as it appears on an oscilloscope, the electro-encephalographic signal or even the speech signal as displayed as a function of time on an oscilloscope in terms of a transducer (microphone) output-voltage fluctuations with time.

The stochastic form of time series does not necessarily mean that the time series are of inherent stochastic nature or origin. One could argue that speech is deterministic in origin. We seem to know and to determine which words we utter. However, even forgetting about environmental noise accompanying the speech, the random nature of the speech signal itself as viewed on an oscilloscope is obvious, especially in cases of unvoiced speech ("sh", "f", "s" sounds etc., at the higher range of speech frequencies, say above 1500 Hz). This is why no two utterances of even the same word are exactly identical in their waveform even on an ordinary oscilloscope screen. Again, when one purchases IBM shares one usually would think a lot about such a purchase. However, the accumulative effect of thousands of buyers and sellers of such shares every day yields the price variation of these shares with time which has a stochastic appearance. The stochastic analysis is thus a macroscopic or global picture of a possibly deterministic, microscopic phenomena.

One could carry these arguments furhter and further, right to the foundations of stochastic processes and of quantum mechanics [1.1], [1.2], to argue that one may deterministically *generate* or simulate very accurately *random processes* by means of *iterative mixing* transformations applied to deterministic (non-random) functions. Hence, the *boundaries between a stochastic process and a deterministic* one seem to be at best obscure.* Without going into detail that is beyond the scope of this text, let us say, that time series are concerned with processes that, whether deterministic in origin or not, appear to be stochastic and lend themselves to a stochastic analysis. Furthermore, the stochastic analysis is much simpler and faster for purposes of prediction, of decision making, of filtering of relevant information from irrelevant one ("noise"), of detection, etc., than is the delving into the deterministic origins, if indeed we can get there. Hence, in relating a myoelectric signal to a certain limb function, it is far simpler to identify a few stochastic time series parameters than to study billions of neurons in the brain. Again, to determine the trend of stock prices, it is faster to look at the time series of the price trends, then to interview every buyer or seller, as it is to analyze the behavior of a multitude of molecules or particles and their collision then apply basic laws of physics to each particle and to all its interactions.

Now, why discrete time?

The *discrete time* form of time series is convenient for two major reasons: First this avoids difficulties in analysis that appear in continuous time stochastic analysis. These difficulties are due to problems that arise in applying classical integration theory to random variables. Roughly speaking, classical differentiation and integration imply continuity or predictability over short intervals, whereas randomness implies the opposite. To accommodate for these difficulties, one must employ stochastic integration theory [1.3], to complicate the analysis considerably at best. In discrete time analysis, integrals are replaced by discrete sums, and differentials are replaced by differences. Therefore, continuity, the

*If you flip a coin and know exactly and can adjust the initial conditions and every detail of finger location, movement, etc., then this may be a deterministic action. Still, if that is true, prediction will, of course, be very laborious. Again if you generate a sequence by printing out $y(t) = \sin(t)$ for $t = \frac{17}{24} k^2$, with $k = 1,2,3,\ldots$ then this is certainly very deterministic but the result is a random-like sequence (see Prob. 1.1). This shows the power of even most simple mixing of deterministic functions.

source of the difficulties, is no more assumed, to simplify the analysis. Secondly, any meaningful analysis of time series is presently solely based on digital computation. The digital computer is a discrete machine. Hence, a discrete time analysis is only natural, as long as sampling rates are appropriate.

1.2 PURPOSE AND SCOPE

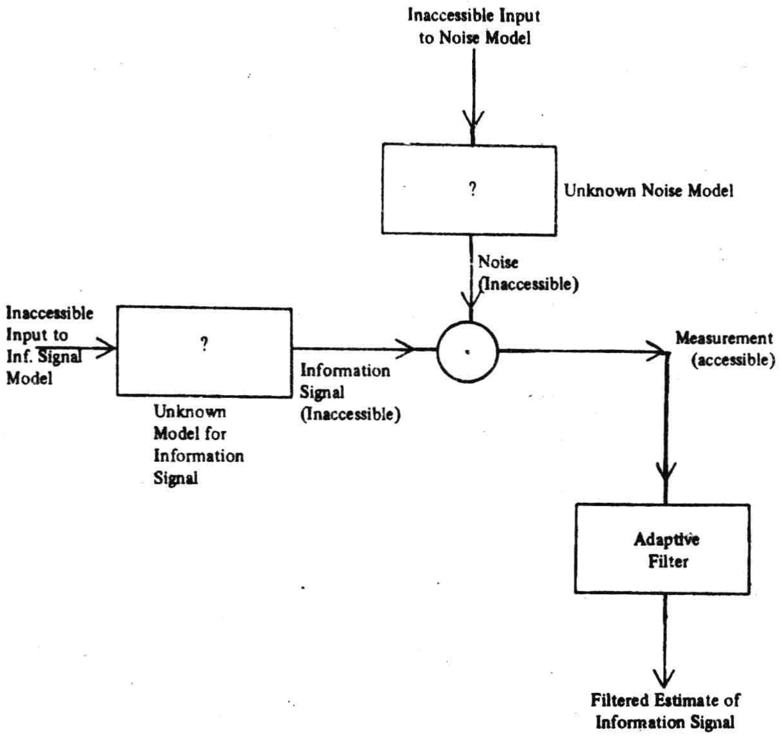
The purpose of this text is not to serve as a text for theory of time series. This book aims at providing tools for rigorously and efficiently employing time series analysis for the retrieval of information imbedded in noise (as is the usual real-world situation) when no prior knowledge of mathematical structure, of the system's or signal's parameter structure, or of the noise structure is available. This retrieval procedure is known as *adaptive filtering* of *a-priori* unknown data from *a-priori* unknown noisy background (see Fig. 1.1)

At this point we feel that some further comments are in order on terminology regarding adaptive filtering. For the purpose of this text we consider that an *adaptive filter* is a filter that serves to filter an information signal from an irrelevant (noise) signal, where the parameters of the information signal or of the noise signal or *of both* are *a-priori unknown*, and (both) are not necessarily white noise. No access to the information signal alone or to the noise signal alone is assumed.

In the most general case where *neither the parameters of the information signal nor those of the noise are a-priori known*, then some prior feature-information (other than parameter knowledge) must be available regarding the parameters of signal or of noise in order that adaptive filtering can be rigorously performed (this feature information may be, for example, in terms of model order, of rate of parameter variation, etc.)

We thus consider situations where only an information signal is considered (which is accessible) and whose parameters require identification, to be a *time series identification problem* and *not* an adaptive filtering problem (as is often assumed in the literature). We consider a situation where information signal and noise exist, but where their parameters are known, to be a *non-adaptive filtering* problem. Filtering when either signal parameters or noise parameters are known, is considered below as an *adaptive filtering problem with partial knowledge*.

This text derives time series analysis and identification theory specifically for the purposes of developing an adaptive filtering theory that is powerful enough under the above situation of lack of prior parameter knowledge. Hence, to be consistent with such a *degree of ignorance*, our analysis must drop assumptions that may be mathematically convenient but unrealistic in *our* real world of ignorance (and where adaptive filtering is of so much need), such as that the signals are Gaussian or that they are stable. If an adaptive filter (controller) loses control in the face of a fault that causes instability, then what is our effort good for? Therefore,



COMMENT: In this generalized schematic of adaptive filtering, both the information-signal and the noise have unknown mathematical models (unknown statistics). Certain adaptive filters, however, assume knowledge of the mathematical model of one or the other, as discussed in Chapter 11 below, whereas others require only some general structure information

Fig. 1.1 Unknown noise.

this is not a classical text on time series, but is concentrates on what is useful *vis-à-vis* adaptive filtering. Since such a degree of ignorance requires parameter identification before everything, a major effort of the text is directed towards identification. Hence again, we cannot assume *stability* or *invertibility* or *Gaussianity*, as we cannot do elsewhere in the text.

Again, when we deal with non-stationarity, we assume instability or fast variations of parameter values, where the parameters themselves (and not just the signal, say, the AR parameters) are stochastic time series, as is most often ignored in the literature but as exists in the real world.

Now, what is adaptive filtering for? We already mentioned separation of relevant information from "noise," say, of speech from noise. The scope of use of adaptive filtering is of course far beyond this. Adaptive filtering is applicable to *prediction* of an information-signal of unknown parameters (unknown mathematical model), when received in (usually) unknown measurement noise. It extends to *detection and decision* on the existance of a certain class of signals (of unknown parameters) some of which may be contaminated by unknown noises. An example of this situation is the decision that the presently generated myoelectric signal is to activate one (particular) of six different joint movements of an artificial limb of an amputee or of a powered orthotic device for a partly paralyzed person, as in Appendix B below [1.4]. Another class of applications is to *controlling* a system (a process, a plant) or to *navigating* a system (of unknown parameters) on the basis of system's data imbedded in measurement noise of unknown parameters—see Fig. 1.2. This type of control or navigation problem is a very common control situation in noisy environments. Adaptive filtering plays a major role if that control problem is extended to become a (self-) adaptive control problem, where in addition to noisy measruements in *a-priori* unknown noise we have also that the process or system to be controlled is unknown, or of time varying parameters, which again is a very realistic situation.

The solutions of adaptive decision, classification and detection problems with unknown parameters as above, and of adaptive control in unknown noise and with unknown parameters, depend largely, if not solely, on solving the time series identification and the adaptive filtering problems imbedded in them. Hence, this text will consider them in the framework of time series analysis and adaptive filtering.

All the above have in common the filtering of signal or information from noise (in the adaptive decision problem) or from irrelevant information with different degrees of parameters uncertainties, to be dealt with *via* time series analysis methods developed in the text. The filtering problem may thus be for noise removal purposes, for control, for classification or for detection purposes.

Considering identification, and since for adaptive filtering and control in real time, the fastest possible adaptation is required, it follows that fast identification is essential. This implies *convergence* of one identifier *at the fastest possible rate,* namely that adequate identification is achieved for the least amount of information (that one need not wait for too many data points for obtaining a good parameter estimate). This in turn calls for *Least Squares* (LS) identifiers which can be shown to possess the fastest possible convergence rates as is discussed in

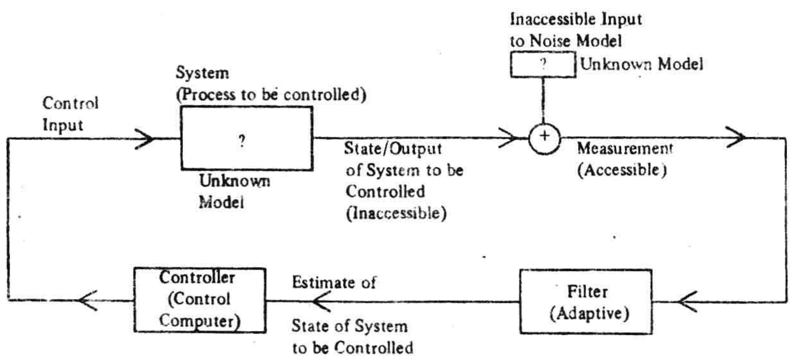


Fig. 1.2 Schematic of generalized structure of an adaptive control system (to control unknown system in unknown noise).