Theory and Practice of Recursive Identification

Lennart Ljung
Torsten Soderstrom

Theory and Practice of Recursive Identification

Lennart Ljung
Torsten Söderström

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Times New Roman by Asco Trade Typesetting Ltd., Hong Kong and printed and bound by Halliday Lithograph in the United States of America.

Library of Congress Cataloging in Publication Data

Ljung, Lennart.
Theory and practice of recursive identification.
(The MIT Press series in signal processing, optimization, and control; 4)
Includes bibliographies and index.
1. System identification. I. Söderström, Torsten. II. Title. III. Series.
QA402.L6 1983 003 82-17928
ISBN 0-262-12095-X

Series Foreword

The fields of signal processing, optimization, and control stand as well-developed disciplines with solid theoretical and methodological foundations. While the development of each of these fields is of great importance, many future problems will require the combined efforts of researchers in all of the disciplines. Among these challenges are the analysis, design, and optimization of large and complex systems, the effective utilization of the capabilities provided by recent developments in digital technology for the design of high-performance control and signal-processing systems, and the application of systems concepts to a variety of applications such as transportation systems, seismic signal processing, and data communication networks.

This series serves several purposes. It not only includes books at the leading edge of research in each field but also emphasizes theoretical research, analytical techniques, and applications that merit the attention of workers in all disciplines. In this way the series should help acquaint researches in each field with other perspectives and techniques and provide cornerstones for the development of new research areas within each discipline and across the boundaries.

Lennart Ljung and Torsten Söderström's book *Theory and Practice of Recursive Identification* stands as a major addition to the literature on system identification and parameter estimation. This topic is a natural one for this series, as the problem of parameter estimation is of great importance in both the fields of signal processing and control. Furthermore, interest in this subject is on the increase, as the availability of inexpensive but computationally powerful digital processors has made feasible the use of advanced and complex adaptive algorithms in a wide variety of applications in which they had not been used or even considered in the past. Consequently Ljung and Söderström's book is a most timely one.

As the authors point out in their preface, the field of recursive identification is filled with a multitude of approaches, perspectives, and techniques whose interrelationships and relative merits are difficult to sort out. As a consequence it has become a decidedly nontrivial task for a newcomer to the field or a nonspecialist to extract the fundamental concepts of recursive identification or to gain enough intuition about a particular technique to be able to use it effectively in practice. For this reason Ljung and Söderström's book is a welcome contribution, as its primary aim is to present a coherent picture of recursive identification.

In doing this the authors have done an outstanding job of constructing and describing a unified framework which not only exposes the crucial issues in the choice and design of an on-line identification algorithm but also provides the reader with a natural and simple frame of reference for understanding the similarities and differences among the many approaches to recursive identification. Furthermore, thanks to careful organization, the authors have produced a book which should have broad appeal. For graduate students and nonspecialists it provides an excellent introduction to the subject. For those primarily interested in using identification algorithms in practice it provides a thorough treatment of the critical aspects of and tradeoffs involved in algorithm design, as well as "user's summaries" which identify those points in each chapter that are of most importance to the practitioner. For the more theoretically inclined, there is a detailed development of convergence analysis for recursive algorithms. And finally, for all who have an interest in identification, be it peripheral or principal, this book should prove to be a valuable reference for many years.

Alan S. Willsky

Preface

The field of recursive identification has been called a "fiddler's paradise" (Åström and Eykhoff, 1971), and it is still often viewed as a long and confusing list of methods and tricks. Though the description was no doubt accurate at the time Åström and Eykhoff's survey was written, we believe that the time has now come to challenge this opinion by providing a comprehensive yet coherent treatment of the field. This has been our motivation for writing this book.

Coherence and unification in the field of recursive identification is not immediate. One reason is that methods and algorithms have been developed in different areas with different applications in mind. The term "recursive identification" is taken from the control literature. In statistical literature the field is usually called "sequential parameter estimation," and in signal processing the methods are known as "adaptive algorithms."

Within these areas, algorithms have been developed and analyzed over the last 30 years. Recently there has been a noticably increased interest in the field from practitioners and industrial "users." This is due to the construction of more complex systems, where adaptive techniques (adaptive control, adaptive signal processing) may be useful or necessary, and, of course, to the availability of microprocessors for the easy implementation of more advanced algorithms. As a consequence, material on recursive identification should be included in undergraduate and graduate courses. With this in mind, the series editor, Alan Willsky, has encouraged us to make this book accessible to a broad audience. This objective perhaps conflicts with our ambition to give a comprehensive treatment of recursive identification. We have tried to solve this conflict by providing bypasses around the more technical portions of the book (see sections 1.4 and 4.1). We have also included a more "leisurely" introduction to the field in chapter 2.

The manuscript of this book has been tested as a text for a first-year graduate course on identification at Stanford University, and for a course for users at Lawrence Livermore Laboratory. For use as a text, the appendixes should be excluded. Depending on whether the emphasis of the course is theory or practice, further reductions in chapters 4–6 could be considered. In a course oriented to practice, chapter 4 could be read according to the "sufficiency path" described in figure 4.1. In a theory course, chapter 5 could be used for illustration, and the algorithms in chapter 6 could be omitted. We have not included exercises in the material. The natural way of getting familiar with recursive identification is to implement and simulate different algorithms; such programming problems are more valuable than formal paper-and-pencil exercises.

As remarked above, the existing literature in the field is extensive. Any attempt to make the reference list comprehensive would therefore be a formidable task. Instead, we have mostly confined ourselves to what appear to be original references and to "further reading" of more detailed accounts of various problems.

Lennart Ljung
Division of Automatic Control
Department of Electrical Engineering
Linköping University, Linköping, Sweden

Torsten Söderström
Department of Automatic Control and Systems Analysis
Institute of Technology
Uppsala University, Uppsala, Sweden

Acknowledgments

Initial conditions play an important role for most systems. We are happy to acknowledge the excellent initial conditions provided for us by Professor Karl Johan Åström and his group at the Lund Institute of Technology in Sweden. They have had a lasting influence on our research in general and on the development of this book in particular.

The structure of the book has emerged from numerous seminars and some short courses on the subject given by the first author around the world. The comments and reactions by the participants in these seminars and courses have helped us greatly.

A large number of researchers have helped us with useful comments on the manuscript in different stages of completion. In the first place we must mention the editor of the series, Dr. Alan Willsky, who has provided most detailed and useful comments throughout the work. We are also grateful to many of our colleagues and friends for helping us with numerous important and clarifying comments. In particular we would like to mention G. Bierman, T. Bohlin, J. Candy, B. Egardt, A. Nehorai, P. Hagander, M. Morf, G. Olsson, U. V. Reddy, J. Salz, C. Samson, S. Shah, B. Widrow, and B. Wittenmark for important and helpful comments.

Dr. F. Soong has provided figure 5.12, which is reproduced from his thesis.

The completion of the book took seven years after it was first planned. Chances are that it would never have been finished if the first author had not had the privilege to spend a sabbatical year at Stanford University in 1980–1981. He would like to thank Professor Thomas Kailath for making this possible and for providing inspiring working conditions.

The manuscript has gone through several versions. They have been expertly typed by Mrs. Ingegerd Stenlund, Miss Ingrid Ringård, and Mrs. Margareta Hallenberg. Our deadlines have been met only because of these persons' willingness to work overtime. We express our sincere appreciation for their help.

We are also indebted to Mrs. Marianne Anse-Lundberg who skilfully and quickly prepared the illustrations.

Finally, we express our gratitude to Dr. Ivar Gustavsson, who took part in the research behind the book and the planning of it. He was an intended coauthor during most of the work. We regret that his obligations in industry finally prevented him from completing his contribution. However, we are sure that he is performing a more important task: To apply and use the methods in the real world–not merely write about them.

Symbols, Abbreviations, and Notational Conventions

Symbols

AsN	asymptotic normal distribution
d	dimension of the parameter vector θ
D_c	set of parameter vectors describing the convergence points
$D_{\mathscr{M}}$	set of parameter vectors describing the model set
D_s	set of parameter vectors decribing models with stable predictors
$\bar{E}$	$\bar{E}f(t) = \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} Ef(k)$ where E = expectation operator
$e(t)$	white noise (a sequence of independent random variables)
F, G, H	matrices for state-space models
$g_{\mathscr{M}}(\cdot, \cdot, \cdot)$	predictor function
$K_i(t)$	reflection coefficient (see section 6.4)
$L(t)$	gain in algorithm
$l(\cdot, \cdot, \cdot)$	loss function
$\mathscr{M}$	model set, model structure
$\mathscr{M}(\theta)$	model corresponding to the parameter vector θ
n	model order
N	run length (see chapter 5)
$N(m, P)$	normal (Gaussian) distribution of mean value m and covariance matrix P
$O(x)$	$O(x)/x$ bounded when $x \to 0$
$o(x)$	$o(x)/x \to 0$ when $x \to 0$
$P(t)$	$\bar{R}^{-1}(t)$
p	dimension of the output vector $y(t)$
$p(\theta \mid z^*)$	posterior probability density function
$\mathbf{R}^n$	Euclidean n-dimensional space
R_1, R_2, R_{12}	covariance matrices

$R(t)$	Hessian approximation in Gauss-Newton algorithm
$\bar{R}(t)$	$tR(t)$
R_D	matrix corresponding to $R(t)$ in the associated d.e.
r	dimension of the input vector $u(t)$
$r(t)$	scalar factor for stochastic gradient algorithm
$\mathscr{S}$	true system
$S(q^{-1})$	prefilter of data
$^{\mathrm{T}}$	transpose
$T(q^{-1})$	prefilter of data
tr	trace (of a matrix)
t	time variable (integer-valued)
$u(t)$	input signal (column vector of dimension r)
$V_t(\theta)$	loss function at time t
$w(t)$	white noise (a sequence of independent random variables)
$\mathscr{X}$	experimental condition
$y(t)$	output signal (column vector of dimension p)
$y_F(t)$	filtered output
$\hat{y}(t)$	predictor using running estimate
$\hat{y}(t \mid \theta)$	as above using fixed model parameter θ
$z(t)$	measurements $(y^{\mathrm{T}}(t)\ \ u^{\mathrm{T}}(t))^{\mathrm{T}}$ at time t
z^N	data set made up of $z(1), \ldots, z(N)$
$\bar{\beta}(N, t)$	forgetting profile defined by (2.115), (2.117)
$\beta(N, t)$	forgetting profile defined by (2.128)
$\gamma(t)$	gain (sequence)
$\delta_{t,s}$	Kronecker's delta
$\varepsilon(t)$	prediction error using running estimate
$\varepsilon(t, \theta)$	as above using fixed model parameter θ
$\bar{\varepsilon}(t)$	residual (posterior prediction error)

$\zeta(t)$	vector of instrumental variables used as gradient approximation in the IV method
$\zeta(t, \theta)$	as above for fixed model parameters θ
$\eta(t)$	general gradient approximation using running estimate (in actual algorithms replaced by one of φ, ψ, or ζ)
$\eta(t, \theta)$	as above for fixed model parameters θ
θ	parameter vector of unknown coefficients (column vector of dimension d)
$\hat{\theta}(t)$	recursive estimate of θ based on data up to time t
θ_D	vector corresponding to $\hat{\theta}(t)$ in the associated d.e.
θ_0	true value of the parameter vector θ
θ^*	limit value of $\hat{\theta}(t)$
$\hat{\theta}_t$	off-line estimate of θ based on data up to time t
Λ_0	covariance matrix ($p \times p$-matrix) of prediction errors
$\hat{\Lambda}(t)$	estimated covariance matrix of prediction errors
λ	forgetting factor
$\xi(t)$	state vector in prediction and gradient calculations using running estimate
$\xi(t, \theta)$	as above for fixed model parameters θ
$\varphi(t)$	vector formed from observed data using a running estimate (the gradient approximation used in PLR)
$\varphi(t, \theta)$	as above for fixed model parameters θ
$\varphi_F(t)$	vector of filtered and lagged data
$\psi(t)$	gradient of the predictions computed using running estimate ($d \times p$-matrix)
$\psi(t, \theta)$	as above for fixed model parameter θ
$\triangleq$	defined as
$:=$	assignment operator
$[x]_D$	projection of x into D

Abbreviations

AR	autoregressive
ARMA	autoregressive moving average
ARMAX	autoregressive moving average with exogenous variables
d.e.	differential equation
EKF	extended Kalman filter
ELS	extended least squares
GLS	generalized least squares
IV	instrumental variable
LS	least squares
MA	moving average
MLE	maximum likelihood estimate
PLR	pseudolinear regression
PRBS	pseudorandom binary sequence
RGLS	recursive generalized least squares
RIV	recursive instrumental variable
RLS	recursive least squares
RML	recursive maximum likelihood
RPE	recursive prediction error
RPEM	recursive prediction error method
SISO	single-input/single output
w.p.1	with probability one
w.r.t.	with respect to

Notational Conventions

$H^{-1}(q^{-1})$	$[H(q^{-1})]^{-1}$
$\varphi^{T}(t)$	$[\varphi(t)]^{T}$

A^{-T}	$[A^{-1}]^T$
l_θ	first derivative of l with respect to θ
$l_{\theta\theta}$	second derivative of l with respect to θ
V'	first derivative of V with respect to its argument

To
Ann-Kristin, Johan, Arvid
and
Marianne, Johanna, Andreas

Contents