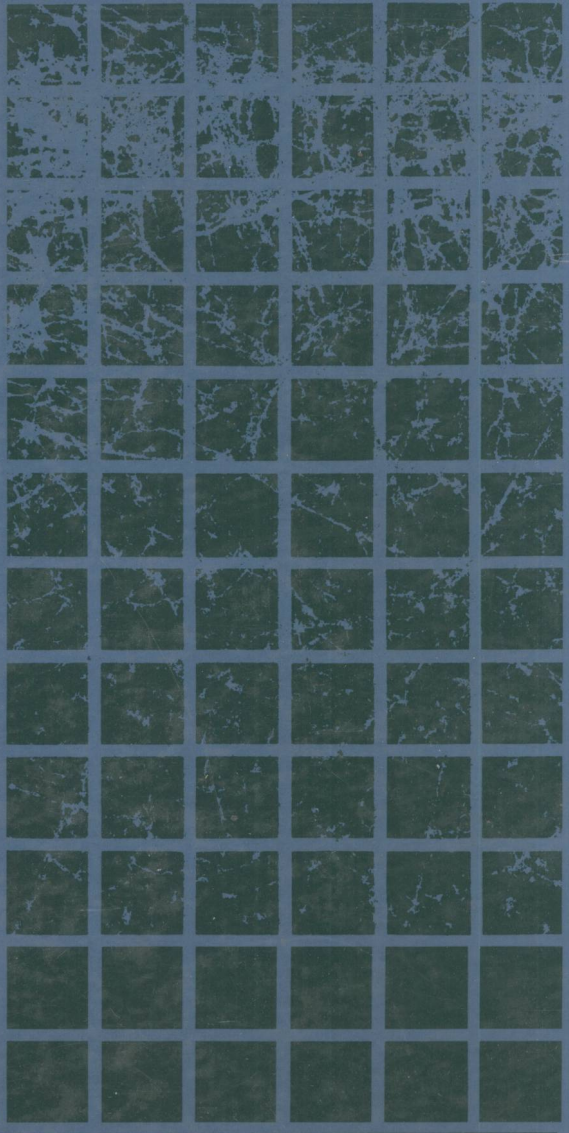
LENNART LJUNG



SYSTEM IDENTIFICATION

THEORY FOR THE USER

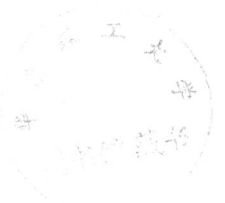
PRENTICE-HALL INFORMATION AND SYSTEM SCIENCES SERIES
Thomas Kailath, Series Editor

SYSTEM IDENTIFICATION:

Theory for the User

Lennart Ljung

University of Linköping
Sweden



Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Ljung, Lennart. (date)
System identification

Bibliography: p.
Includes index.
1. System identification. I. Title.
QA402.L59 1987 003 86-22712
ISBN 0-13-881640-9

Editorial/production supervision and
interior design: Gloria Jordan
Cover design: Ben Santora
Manufacturing buyer: S. Gordon Osbourne

A division of Simon & Schuster
Englewood Cliffs, New Jersey 07632

Printed in the United States of America

10 9 8 7 6 5 4 3 2

ISBN 0-13-881640-9 025

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall of Southeast Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

SYSTEM IDENTIFICATION:

Theory for the User

PRENTICE-HALL INFORMATION AND SYSTEM SCIENCES SERIES

Thomas Kailath, Editor

ANDERSON & MOORE	*Optimal Filtering*
ASTROM & WITTENMARK	*Computer-Controlled Systems: Theory and Design*
GOODWIN & SIN	*Adaptive Filtering, Prediction, and Control*
GRAY & DAVISSON	*Random Processes: A Mathematical Approach for Engineers*
HAYKIN	*Adaptive Filter Theory*
KAILATH	*Linear Systems*
KUMAR & VARAIYA	*Stochastic Systems: Estimation, Identification, and Adaptive Control*
KUN, WHITEHOUSE & KAILATH, EDS.	*VLSI and Modern Signal Processing*
LJUNG	*System Identification: Theory for the User*
MACOVSKI	*Medical Imaging Systems*
MELSO & SAGE	*An Introduction to Probability and Stochastic Processes*
SPILKER	*Digital Communications by Satellite*
WILLIAMS	*Designing Digital Filters*

To Ann-Kristin, Johan, and Arvid

PREFACE

System identification is a diverse field that can be presented in many different ways. The subtitle, *Theory for the User*, reflects the attitude of the present treatment. Yes, the book is about theory, but the focus is on theory that has direct consequences for the understanding and practical use of available techniques. My goal has been to give the reader a firm grip on basic principles so that he or she can confidently approach a practical problem, as well as the rich and sometimes confusing literature on the subject.

Stressing the utilitarian aspect of theory should not, I believe, be taken as an excuse for sloppy mathematics. Therefore, I have tried to develop the theory without cheating. The more technical parts have, however, been placed in appendixes or in asterisk-marked sections, so that the reluctant reader does not have to stumble through them. In fact, it is a redeeming feature of life that we are able to use many things without understanding every detail of them. This is true also of the theory of system identification. The practitioner who is looking for some quick advice should thus be able to proceed rapidly to Part III (User's Choices) by hopping through the summary sections of the earlier chapters.

The core material of the book should be suitable for a graduate-level course in system identification. As a prerequisite for such a course, it is natural, although not absolutely necessary, to require that the student should be somewhat familiar with dynamical systems and stochastic signals. The manuscript has been used as a text for system identification courses at Stanford University, the Massachusetts Institute of Technology, Yale University, the Australian National University and the Univer-

sities of Lund and Linköping. Course outlines, as well as a solutions manual for the problems, are available from the publisher.

For a course on system identification, the role of computer-based exercises should be stressed. Simulation sessions demonstrating how hidden properties of data are readily recovered by the techniques discussed in the book enhance the understanding and motivation of the material. In the problems labeled S, in Chapters 2 through 16, a basic interactive software package is outlined that should be possible to implement rather painlessly in a high-level environment. A PC-MATLAB version of this package is commercially available (see Ljung, 1986b). With such a package all basic techniques of this book can be illustrated and tested on real and simulated data.

The existing literature on system identification is indeed extensive and virtually impossible to cover in a bibliography. In this book I have tried to concentrate on recent and easily available references that I think are suitable for further study, as well as on some earlier works that reflect the roots of various techniques and results. Clearly, many other relevant references have been omitted.

Finally, some words about the structure of this book: The dependence among the different chapters is illustrated in Figure 1.13, which shows that some chapters are not necessary prerequisites for the following ones. Also, some portions contain material that is directed more toward the serious student of identification theory than to the user. These portions are put either in appendixes or in sections and subsections marked with an asterisk (*). While occasional references to this material may be encountered, it is safe to regard it as optional reading; the continuity will not be impaired if it is skipped.

The problem sections for each chapter have been organized into six groups of different problem types:

- *G problems:* These could be of General interest and it may be worthwhile to browse through them, even without intending to solve them.
- *E problems:* These are regular pencil-and-paper Exercises to check the basic techniques of the chapter.
- *T problems:* These are Theoretically oriented problems and typically more difficult than the E problems.
- *D problems:* In these problems the reader is asked to fill in technical Details that were glossed over in the text (a way to dump straightforward technicalities from the book into the solutions manual!).
- *S problems:* These develop the basic identification Software package mentioned earlier.
- *C problems:* These require a Computer. Clearly, with the software package at hand, the C problems can be complemented with a myriad of problems experimenting with identification methods and data. Such problems are not specifically listed, but the reader is encouraged to apply those techniques in an exploratory fashion.

ACKNOWLEDGMENTS

Any author of a technical book is indebted to the people who taught him the subject and to the people who made the writing possible. My interest in system identification goes back to my years as a graduate student at the Automatic Control Department in Lund. Professor Karl Johan Åström introduced me to the subject, and his serious attitude to research has always been a reference model for me. Since then I have worked with many other people who added to my knowledge of the subject. I thank, therefore, my previous coauthors (in alphabetical order) Anders Ahlén, Peter Caines, David Falconer, Farhat Fnaiech, Ben Friedlander, Michel Gevers, Keith Glover, Ivar Gustavsson, Tom Kailath, Stefan Ljung, Martin Morf, Ton van Overbeek, Jorma Rissanen, Torsten Söderström, Göte Solbrand, Eva Trulsson, Bo Wahlberg, Don Wiberg, and Zhen-Dong Yuan.

The book has developed from numerous seminars and several short courses that I have given on the subject world-wide. Comments from the seminar participants have been instrumental in my search for a suitable structure and framework for presenting the topic.

Several persons have read and used the manuscript in its various versions and given me new insights. First, I would like to mention: Michel Gevers, who taught from an early version and gave me invaluable help in revising the text; Robert Kosut and Arye Nehorai, who taught from the manuscript at Stanford and Yale, respectively; and Jan Holst, who lead a discussion group with it at Denmark's Technical University, and also gathered helpful remarks. I co-taught the course at MIT with Fred Schweppe, and his lectures as well as his comments, led to many clarifying

changes in the manuscript. Students in various courses also provided many useful comments. I mention in particular George Hart, Juan Lavalle, Ivan Mareels, Brett Ridgely, and Bo Wahlberg. Several colleagues were also kind enough to critique the manuscript. I am especially grateful to Hiro Akaike, Chris Byrnes, Peter Falb, Meir Feder, Gene Franklin, Claes Källström, David Ruppert, Torsten Söderström, Petre Stoica, and Peter Whittle.

Svante Gunnarsson and Stan Granath made the experiments described in Section 17.2, Bo Wahlberg contributed to the frequency-domain interpretations, and Alf Isaksson prepared Figure 14.4.

The preparation of the manuscript's many versions was impeccably coordinated and, to a large extent, also carried out by Ingegerd Stenlund. She had useful help from Ulla Salaneck and Karin Lönn. Marianne Anse-Lundberg expertly prepared all the illustrations. I deeply appreciate all their efforts.

Writing a book takes time, and I probably would not have been able to finish this one had I not had the privilege of sabbatical semesters. The first outline of this book was written during a sabbatical leave at Stanford University in 1980–1981. I wrote a first version of what turned out to be the last edition during a mini-sabbatical visit to the Australian National University in Canberra in 1984. The writing was completed during 1985–1986, the year I spent at MIT. I thank Tom Kailath, Brian Anderson and Sanjoy Mitter (and the U.S. Army Research Office under contract DAAG-29-84-K-005) for making these visits possible and for providing inspiring working conditions. My support from the Swedish National Board for Technical Development (STUF) has also been important.

Sabbatical or not, it was unavoidable that a lot of the writing (not to mention the thinking!) of the book had to be done on overtime. I thank my family, Ann-Kristin, Johan, and Arvid, for letting me use their time.

Lennart Ljung
Linköping, Sweden

OPERATORS AND NOTATIONAL CONVENTIONS

$\arg(z)$ = argument of the complex number z

$\arg\min f(x)$ = value of x that minimizes $f(x)$

$x_N \in AsF(n, m)$: sequence of random variables x_N converges in distribution to the F-distribution with n and m degrees of freedom

$x_N \in AsN(m, P)$: sequence of random variables x_N converges in distribution to the normal distribution with mean m and covariance matrix P; see (I.17)

$x_N \in As\chi^2(n)$: sequence of random variables x_N converges in distribution to the χ^2 distribution with n degrees of freedom

$\mathrm{Cov}(x)$ = covariance matrix of the random vector x; see (I.4)

$\det A$ = determinant of the matrix A

$\dim \theta$ = dimension (number of rows) of the column vector θ

$E\,x$ = mathematical expectation of the random vector x; see (I.3)

$\bar{E}\,x(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E\,x(t)$; see (2.60)

$\mathbf{O}(x)$ = ordo x: function tending to zero at the same rate as x

$\mathbf{o}(x)$ = small ordo x: function tending to zero faster than x

$x \in N(m, P)$: random variable x is normally distributed with mean m and covariance matrix P; see (I.6)

Re z = real part of the complex number z

$\mathcal{R}(f)$ = range of the function f = the set of values that $f(x)$ may assume

$\mathbf{R}^d$ = Euclidian d-dimensional space

$x = \text{sol}\{f(x) = 0\}$: x is the solution (or set of solutions) to the equation $f(x) = 0$

$\text{tr}(A)$ = trace (the sum of the diagonal elements) of the matrix A

Var (x) = variance of the random variable x

A^{-1} = inverse of the matrix A

A^T = transpose of the matrix A

A^{-T} = transpose of the inverse of the matrix A

$\bar{z}$ = complex conjugate of the complex number z

(superscript * is not used to denote transpose and complex conjugate: it is used only as a distinguishing superscript)

$y_s^t = \{y(s), y(s+1), \ldots, y(t)\}$

$y^t = \{y(1), y(2), \ldots, y(t)\}$

$U_N(\omega)$ = Fourier transform of u^N; see (2.37)

$R_v(\tau) = \bar{E}v(t)v^T(t-\tau)$; see (2.61)

$R_{sw}(\tau) = \bar{E}s(t)w^T(t-\tau)$; see (2.62)

$\Phi_v(\omega)$ = spectrum of v = Fourier transform of $R_v(\tau)$; see (2.63)

$\Phi_{sw}(\omega)$ = cross spectrum between s and w = Fourier transform of $R_{sw}(\tau)$; see (2.64)

$\hat{R}_s^N(\tau) = \frac{1}{N}\sum_{t=1}^{N} s(t)s^T(t-\tau)$; see (6.10)

$\hat{\Phi}_u^N(\omega)$ = estimate of the spectrum of u based on u^N; see (6.48)

$\hat{v}(t|t-1)$ = prediction of $v(t)$ based on v^{t-1}

$\frac{d}{d\theta}V(\theta)$ = gradient of $V(\theta)$ with respect to θ: a column vector of dimension dim θ if V is scalar valued

$V'(\theta)$ = gradient of V with respect to its argument

$\ell'_\varepsilon(\varepsilon,\theta)$ = partial derivative of ℓ with respect to ε

δ_{ij} = Kronecker's delta: zero unless $i = j$

$\delta(k) = \delta_{k0}$

$\mathcal{B}(\theta_0,\varepsilon) = \varepsilon$ neighborhood of θ_0: $\{\theta \| \theta - \theta_0| < \varepsilon\}$

$\triangleq$ = the left side is defined by the right side

$|\cdot|$ = (Euclidian) norm of a vector

$\|\cdot\|$ = (Frobenius) norm of a matrix (see 2.89)

SYMBOLS USED IN TEXT

This list contains symbols that have some global use. Some of the symbols may have another local meaning.

$D_{\mathcal{M}}$ = set of values over which θ ranges in a model structure. See (4.119)

D_c = set into which the θ-estimate converges. See (8.23)

$e(t)$ = disturbance at time t; usually $\{e(t), t = 1, 2, \dots\}$ is white noise (a sequence of independent random variables with zero mean values and variance λ)

$e_0(t)$ = "true" driving disturbance acting on a given system $\mathcal{S}$; see (8.2)

$f_e(x), f_e(x,\theta)$ = probability density function of the random variable e; see (I.2) and (4.4)

$G(q)$ = transfer function from u to y; see (2.20)

$G(q,\theta)$ = transfer function in a model structure, corresponding to the parameter value θ; see (4.4)

$G_0(q)$ = "true" transfer function from u to y for a given system; see (8.7)

$\hat{G}_N(q)$ = estimate of $G(q)$ based on Z^N

$G^*(q)$ = limiting estimate of $G(q)$; see (8.68)

$\tilde{G}_N(q)$ = difference $\hat{G}_N(q) - G_0(q)$; see (8.15)

$\mathcal{G}$ = set of transfer functions obtained in a given structure; see (8.44)

$H(q), H(q,\theta), H_0(q), \hat{H}_N(q), H^*(q), \tilde{H}_N(q), \mathcal{H}$: analogous to G but for the transfer function from e to y

$L(q)$ = prefilter for the prediction errors; see (7.10)

$\ell(\varepsilon), \ell(\varepsilon,\theta), \ell(\varepsilon,t,\theta)$ = norm for the prediction errors used in the criterion; see (7.11), (7.16), (7.18)

$\mathcal{M}$ = model structure (a mapping from a parameter space to a set of models); see (4.119)

$\mathcal{M}(\theta)$ = particular model corresponding to the parameter value θ; see (4.119)

$\mathcal{M}^*$ = set of models (usually generated as the range of a model structure); see (4.115) and page 93

P_θ = asymptotic covariance matrix of θ; see (9.11)

q, q^{-1} = forward and backward shift operators; see (2.15)

$\mathcal{S}$ = "the true system"; see (8.7)

$T(q) = [G(q)\ H(q)]$; see (4.106)

$T(q,\theta), T_0(q), \hat{T}_N(q), \tilde{T}_N(q)$ = analogous to G and H

$u(t)$ = input variable at time t

$V_N(\theta, Z^N)$ = criterion function to be minimized; see (7.11)

$\bar{V}(\theta)$ = limit of criterion function; see (8.28)

$v(t)$ = disturbance variable at time t

$w(t)$ = usually a disturbance variable at time t; the precise meaning varies with the local context

$x(t)$ = state vector at time t; dimension = n

$y(t)$ = output variable at time t

$\hat{y}(t|\theta)$ = predicted output at time t using a model $\mathcal{M}(\theta)$ and based on Z^{t-1}; see (4.6)

$z(t) = [y(t)\ u(t)]^T$; see (4.110)

$Z^N = \{u(0), y(0), \ldots, u(N), y(N)\}$

$\varepsilon(t,\theta)$ = prediction error $y(t) - \hat{y}(t|\theta)$

λ = used to denote variance; also, in Chapter 11, the forgetting factor; see (11.6), (11.63)

θ = vector used to parametrize models; dimension = d; see (4.4), (4.5), (5.33)

$\hat{\theta}_N$, θ_0, θ^*, $\bar{\theta}_N$ = analogous to G

$\varphi(t)$ = regression vector at time t; see (4.11) and (5.34)

$\chi_0(t) = [u(t)\ e_0(t)]^T$; see (8.14)

$\psi(t,\theta)$ = gradient of $\hat{y}(t|\theta)$ with respect to θ; a d-dimensional column vector; see (4.118c)

$\zeta(t)$, $\zeta(t,\theta)$ = "the correlation vector" (instruments); see (7.96)

$\mathbf{T}'(q,\theta)$ = gradient of $T(q,\theta)$ with respect to θ (a $d \times 2$ matrix); see (4.122)

ABBREVIATIONS AND ACRONYMS

ARARX: See Table 4.1

ARMA: AutoRegressive Moving Average (see Table 4.1)

ARMAX: AutoRegressive Moving Average with eXternal input (see Table 4.1)

ARX: AutoRegressive with eXternal input (see Table 4.1)

BJ: Box–Jenkins model structure (see Table 4.1)

ETFE: Empirical Transfer Function Estimate; see (6.24)

FIR: Finite Impulse Response model (see Table 4.1)

IV: Instrumental variables (see Section 7.6)

LS: Least Squares (see Section 7.3)

ML: Maximum Likelihood (see Section 7.4)

MSE: Mean Square Error

OE: Output error model structure (see Table 4.1)
PDF: Probability Density Function
PEM: Prediction-Error Method (see Section 7.2)
PLR: PseudoLinear Regression (see Section 7.5)
RIV: Recursive IV (see Section 11.3)
RLS: Recursive LS (see Section 11.2)
RPEM: Recursive PEM (see Section 11.4)
RPLR: Recursive PLR (see Section 11.5)
SISO: Single Input Single Output
w.p.: with probability
w.p. 1: with probability one; see (I.15)
w.r.t.: with respect to

CONTENTS