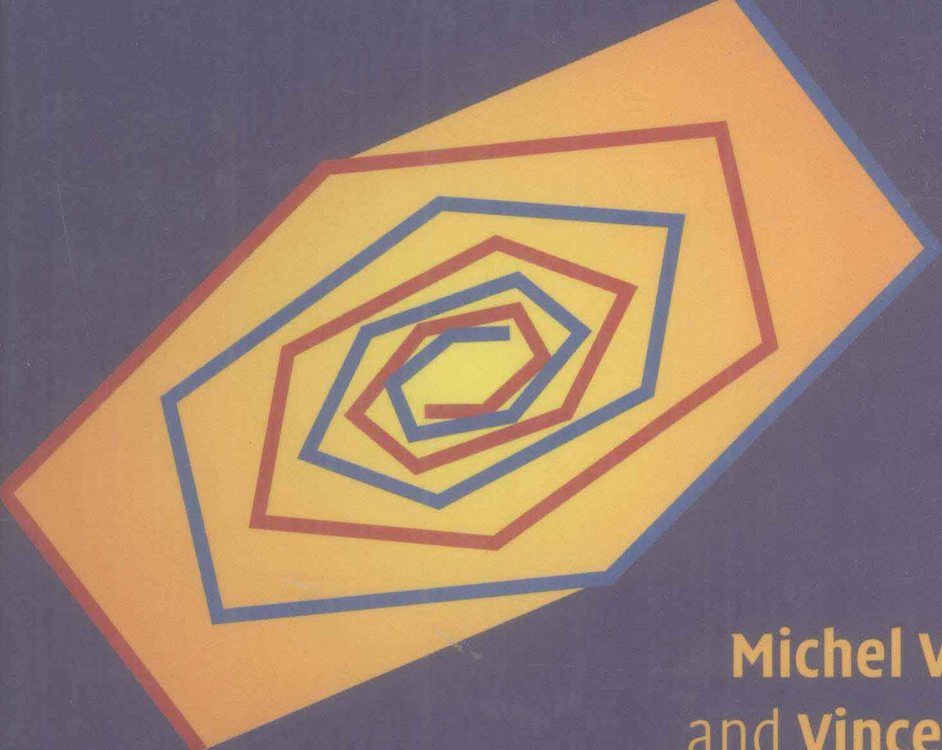

Michel Verhaegen
and Vincent Verdult

Filtering and System Identification

A Least Squares Approach

CAMBRIDGE

Filtering and System Identification

A Least Squares Approach

Michel Verhaegen
Technische Universiteit Delft, the Netherlands
Vincent Verdult
Technische Universiteit Delft, the Netherlands

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521875127

First published 2007

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN-13 978-0-521-87512-7 hardback

FILTERING AND SYSTEM IDENTIFICATION

Filtering and system identification are powerful techniques for building models of complex systems in communications, signal processing, control, and other engineering disciplines. This book discusses the design of reliable numerical methods to retrieve missing information in models derived using these techniques. Particular focus is placed on the least squares approach as applied to estimation problems of increasing complexity to retrieve missing information about a linear state-space model.

The authors start with key background topics including linear matrix algebra, signal transforms, linear system theory, and random variables. They then cover various estimation and identification methods in the state-space model. A broad range of filtering and system-identification problems are analyzed, starting with the Kalman filter and concluding with the estimation of a full model, noise statistics, and state estimator directly from the data. The final chapter on the system-identification cycle prepares the reader for tackling real-world problems.

With end-of-chapter exercises, MATLAB simulations and numerous illustrations, this book will appeal to graduate students and researchers in electrical, mechanical, and aerospace engineering. It is also a useful reference for practitioners. Additional resources for this title, including solutions for instructors, are available online at www.cambridge.org/9780521875127.

Michel Verhaegen is professor and co-director of the Delft Center for Systems and Control at the Delft University of Technology in the Netherlands. His current research involves applying new identification and controller design methodologies to industrial benchmarks, with particular focus on areas such as adaptive optics, active vibration control, and global chassis control.

Vincent Verdult was an assistant professor in systems and control at the Delft University of Technology in the Netherlands, from 2001 to 2005, where his research focused on system identification for nonlinear state-space systems. He is currently working in the field of information theory.

Preface

This book is intended as a first-year graduate course for engineering students. It stresses the role of linear algebra and the least-squares problem in the field of filtering and system identification. The experience gained with this course at the Delft University of Technology and the University of Twente in the Netherlands has shown that the review of undergraduate study material from linear algebra, statistics, and system theory makes this course an ideal start to the graduate course program. More importantly, the geometric concepts from linear algebra and the central role of the least-squares problem stimulate students to understand how filtering and identification algorithms arise and also to start developing new ones. The course gives students the opportunity to see mathematics at work in solving engineering problems of practical relevance.

The course material can be covered in seven lectures:

(i) Lecture 1: Introduction and review of linear algebra (Chapters 1 and 2)
(ii) Lecture 2: Review of system theory and probability theory (Chapters 3 and 4)
(iii) Lecture 3: Kalman filtering (Chapter 5)
(iv) Lecture 4: Estimation of frequency-response functions (Chapter 6)
(v) Lecture 5: Estimation of the parameters in a state-space model (Chapters 7 and 8)
(vi) Lecture 6: Subspace model identification (Chapter 9)
(vii) Lecture 7: From theory to practice: the system-identification cycle (Chapter 10).

The authors are of the opinion that the transfer of knowledge is greatly improved when each lecture is followed by working classes in which the students do the exercises of the corresponding classes under the supervision of a tutor. During such working classes each student has the opportunity to ask individual questions about the course material covered. At the Delft University of Technology the course is concluded by a real-life case study in which the material covered in this book has to be applied to identify a mathematical model from measured input and output data.

The authors have used this book for teaching MSc students at Delft University of Technology and the University of Twente in the Netherlands. Students attending the course were from the departments of electrical, mechanical, and aerospace engineering, and also applied physics. Currently, this book is being used for an introductory course on filtering and identification that is part of the core of the MSc program Systems and Control offered by the Delft Center for Systems and Control (`http://www.dcsc.tudelft.nl`). Parts of this book have been used in the graduate teaching program of the Dutch Institute of Systems and Control (DISC). Parts of this book have also been used by Bernard Hanzon when he was a guest lecturer at the Technische Universität Wien in Austria, and by Jonas Sjöberg for undergraduate teaching at Chalmers University of Technology in Sweden.

The writing of this book stems from the attempt of the authors to make their students as enthusiastic about the field of filtering and system identification as they themselves are. Though these students have played a stimulating and central role in the creation of this book, its final format and quality has been achieved only through close interaction with scientist colleagues. The authors would like to acknowledge the following persons for their constructive and helpful comments on this book or parts thereof: Dietmar Bauer (Technische Universität Wien, Austria), Bernard Hanzon (University College Cork, Ireland), Gjerrit Meinsma (University of Twente, the Netherlands), Petko Petkov (Technical University of Sofia, Bulgaria), Phillip Regalia (Institut National des Télécommunications, France), Ali Sayed (University of California, Los Angeles, USA), Johan Schoukens (Free University of Brussels, Belgium), Jonas Sjöberg (Chalmers University of Technology, Sweden), and Rufus Fraanje (TU Delft).

Special thanks go to Niek Bergboer (Maastricht University, the Netherlands) for his major contributions in developing the Matlab software and guide for the identification methods described in the book. We finally would like to thank the PhD students Paolo Massioni and Justin Rice for help in proof reading and with the solution manual.

Notation and symbols

$\mathbb{Z}$	the set of integers
$\mathbb{N}$	the set of positive integers
$\mathbb{C}$	the set of complex numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the set of real-valued n-dimensional vectors
$\mathbb{R}^{m \times n}$	the set of real-valued m by n matrices
∞	infinity
Re	real part
Im	imaginary part
$\in$	belongs to
$=$	equal
$\approx$	approximately equal
$\square$	end of proof
$\otimes$	Kronecker product
I_n	the $n \times n$ identity matrix
$[A]_{i,j}$	the (i, j)th entry of the matrix A
$A(i, :)$	the ith row of the matrix A
$A(:, i)$	the ith column of the matrix A
A^{T}	the transpose of the matrix A
A^{-1}	the inverse of the matrix A
$A^{1/2}$	the symmetric positive-definite square root of the matrix A
$\mathrm{diag}(a_1, a_2, \ldots, a_n)$	an $n \times n$ diagonal matrix whose (i, i)th entry is a_i
$\det(A)$	the determinant of the matrix A
$\mathrm{range}(A)$	the column space of the matrix A
$\mathrm{rank}(A)$	the rank of the matrix A
$\mathrm{trace}(A)$	the trace of the matrix A

$\mathrm{vec}(A)$	a vector constructed by stacking the columns of the matrix A on top of each other
$\|A\|_2$	the 2-norm of the matrix A
$\|A\|_{\mathrm{F}}$	the Frobenius norm of the matrix A
$[x]_i$	the ith entry of the vector x
$\|x\|_2$	the 2-norm of the vector x
lim	limit
min	minimum
max	maximum
sup	supremum (least upper bound)
$E[\,\cdot\,]$	statistical expected value
$\delta(t)$	Dirac delta function (Definition 3.8 on page 53)
$\Delta(k)$	unit pulse function (Definition 3.3 on page 44)
$s(k)$	unit step function (Definition 3.4 on page 44)
$X \sim (m, \sigma^2)$	Gaussian random variable X with mean m and variance σ^2

List of abbreviations

ARX	Auto-Regressive with eXogeneous input
ARMAX	Auto-Regressive Moving Average with eXogeneous input
BIBO	Bounded Input, Bounded Output
BJ	Box–Jenkins
CDF	Cumulative Distribution Function
DARE	Discrete Algebraic Ricatti Equation
DFT	Discrete Fourier Transform
DTFT	Discrete-Time Fourier Transform
ETFE	Empirical Transfer-Function Estimate
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FRF	Frequency-Response Function
IID	Independent, Identically Distributed
IIR	Infinite Impulse Response
LTI	Linear Time-Invariant
LTV	Linear Time-Varying
MIMO	Multiple Input, Multiple Output
MOESP	Multivariable Output-Error State-sPace
N4SID	Numerical algorithm for Subspace IDentification
PDF	Probability Density Function
PEM	Prediction-Error Method
PI	Past Inputs
PO	Past Outputs
OE	Output-Error
RMS	Root Mean Square
SISO	Single Input, Single Output
SRCF	Square-Root Covariance Filter
SVD	Singular-Value Decomposition
WSS	Wide-Sense Stationary

Contents

1
Introduction

Making observations through the senses of the environment around us is a natural activity of living species. The information acquired is diverse, consisting for example of sound signals and images. The information is processed and used to make a particular model of the environment that is applicable to the situation at hand. This act of model building based on observations is embedded in our human nature and plays an important role in daily decision making.

Model building through observations also plays a very important role in many branches of science. Despite the importance of making observations through our senses, scientific observations are often made via measurement instruments or sensors. The measurement data that these sensors acquire often need to be processed to judge or validate the experiment, or to obtain more information on conducting the experiment. Data are often used to build a mathematical model that describes the dynamical properties of the experiment. System-identification methods are systematic methods that can be used to build mathematical models from measured data. One important use of such mathematical models is in predicting model quantities by filtering acquired measurements.

A milestone in the history of filtering and system identification is the method of least squares developed just before 1800 by Johann Carl Friedrich Gauss (1777–1855). The use of least squares in filtering and identification is a recurring theme in this book. What follows is a brief sketch of the historical context that characterized the early development of the least-squares method. It is based on an overview given by Bühler (1981).

At the time Gauss first developed the least-squares method, he did not consider it very important. The first publication on the least-squares method was published by Adrien-Marie Legendre (1752–1833) in 1806, when Gauss had already clearly and frequently used the method much earlier. Gauss motivated and derived the method of least squares substantially in the papers *Theoria combinationis observationum erroribus minimis obnoxiae* I and II of 1821 and 1823. Part I is devoted to the theory and Part II contains applications, mostly to problems from astronomy. In Part I he developed a probability theory for accidental errors (*Zufallsfehler*). Here Gauss defined a (probability distribution) function $\phi(x)$ for the error in the observation x. On the basis of this function, the product $\phi(x)\mathrm{d}x$ is the probability that the error falls within the interval between x and $x+\mathrm{d}x$. The function $\phi(x)$ had to satisfy the normalization condition

$$\int_{-\infty}^{\infty} \phi(x)\mathrm{d}x = 1.$$

The decisive requirement postulated by Gauss is that the integral

$$\int_{-\infty}^{\infty} x^2\phi(x)\mathrm{d}x$$

attains a minimum. The selection of the square of the error as the most suitable weight is why this method is called the method of least squares. This selection was doubted by Pierre-Simon Laplace (1749–1827), who had earlier tried to use the absolute value of the error. Computationally the choice of the square is superior to Laplace's original method.

After the development of the basic theory of the least-squares method, Gauss had to find a suitable function $\phi(x)$. At this point Gauss introduced, after some heuristics, the Gaussian distribution

$$\phi(x) = \frac{1}{\pi}\mathrm{e}^{-x^2}$$

as a "natural" way in which errors of observation occur. Gauss never mentioned in his papers statistical distribution functions different from the Gaussian one. He was caught in his own success; the applications to which he applied his theory did not stimulate him to look for other distribution functions. The least-squares method was, at the beginning of the nineteenth century, his indispensable theoretical tool in experimental research; and he saw it as the most important witness to the connection between mathematics and Nature.

Still today, the ramifications of the least-squares method in mathematical modeling are tremendous and any book on this topic has to narrow itself down to a restrictive class of problems. In this introductory textbook on system identification we focus mainly on the identification of linear state-space models from measured data sequences of inputs and outputs of the engineering system that we want to model. Though this focused approach may at first seem to rule out major contributions in the field of system identification, the contrary is the case. It will be shown in the book that the state-space approach chosen is capable of treating many existing identification methods for estimating the parameters in a difference equation as special cases. Examples are given for the widely used ARX and ARMAX models (Ljung, 1999).

The central goal of the book is to help the reader discover how the linear least-squares method can solve, or help in solving, different variants of the linear state-space model-identification problem. The linear least-squares method can be formulated as a deterministic parameter-optimization problem of the form

$$\min_x \mu^{\mathrm{T}} \mu \quad \text{subject to } y = Fx + \mu, \tag{1.1}$$

with the vector $y \in \mathbb{R}^N$ and the matrix $F \in \mathbb{R}^{N \times n}$ given and with $x \in \mathbb{R}^n$ the vector of unknown parameters to be determined. The solution of this optimization problem is the subject of a large number of textbooks. Although its analytic solution can be given in a proof of only a few lines, these textbooks analyze the least-squares solution from different perspectives. Examples are the statistical interpretation of the solution under various assumptions on the entries of the matrix F and the perturbation vector μ, or the numerical solution in a computationally efficient manner by exploiting structure in the matrix F. For an advanced study of the least-squares problem and its applications in many signal-processing problems, we refer to the book of Kailath *et al.* (2000).

The main course of this book is preceded by three introductory chapters. In Chapter 2 a refreshment survey of matrix linear algebra is given. Chapter 3 gives a brief overview of signal transforms and linear system theory for deterministic signals and systems. Chapter 4 treats random variables and random signals. Understanding the system-identification methods discussed in this book depends on a profound mastering of the background material presented in these three chapters.

Often, the starting point of identifying a dynamical model is the determination of a predictor. Therefore, in Chapter 5, we first study the