Identification, Adaptation, Learning

The Science of Learning Models from Data

Edited by

Sergio Bittanti
Politecnico di Milano
Piazza Leonardo da Vinci 32
I-20133 Milano, Italy

Giorgio Picci
Università di Padova
Via Gradenigo 6/A
I-35131 Padova, Italy

Springer
Published in cooperation with NATO Scientific Affairs Division
Dedicated to the memory of E.J. Hannan (1921–1994)

Photograph by Peter Hall taken at the farewell to Geof Watson Princeton University, 1992
This book collects the lectures given at the NATO Advanced Study Institute *From Identification to Learning* held in Villa Olmo, Como, Italy, from August 22 to September 2, 1994.

The school was devoted to the themes of *Identification, Adaptation and Learning*, as they are currently understood in the Information and Control engineering community, their development in the last few decades, their inter-connections and their applications. These titles describe challenging, exciting and rapidly growing research areas which are of interest both to control and communication engineers and to statisticians and computer scientists.

In accordance with the general goals of the Institute, and notwithstanding the rather advanced level of the topics discussed, the presentations have been generally kept at a fairly tutorial level. For this reason this book should be valuable to a variety of researchers and to graduate students interested in the general area of Control, Signals and Information Processing. As the goal of the school was to explore a common methodological line of reading the issues, the flavor is quite interdisciplinary. We regard this as an original and valuable feature of this book.

During the two weeks of the school at Villa Olmo we have experienced a unique atmosphere and a most remarkable climate of interaction and communication between the outstanding experts gathered in Como for this occasion. It is remarkable that some of them hardly ever meet at conferences or at scientific meetings as their different fields have traditionally evolved along separate lines. The openness and active participation in discussions by both students and speakers, was a major point for the success of this Advanced Study Institute. The editors of this volume would like to thank all lecturers, and the remaining members of the Organizing Committee, S.K. Mitter and Jan C. Willems, for their helpful advice.

The superb local organization provided by the Centro di Cultura Scientificca Alessandro Volta deserves a primary acknowledgement, for it was a major factor for the smooth development of the school. Special thanks go to Manuela Troglio for her care in the general organizational aspects, and to Emanuela Salati for her kindness and patience in dealing with daily problems of students and teachers.
Last but not least, we would like to thank NATO for believing in this project and for the generous support of the Institute, and the Consiglio Nazionale delle Ricerche (CNR) of Italy which also provided financial funding. The general support of the Dipartimento di Elettronica e Informazione of the Politecnico di Milano is also gratefully acknowledged.

The final software layout of the book is due to Stefano Bertoncello, with the assistance of Marco Lovera.

March 1996

Sergio Bittanti and Giorgio Picci
Table of Contents

Geometric Methods for State Space Identification
Anders Lindquist and Giorgio Picci ........................................... 1

1. Introduction ........................................................................... 1
   1.1 Stationary Signals and the Statistical Theory of Model Building  4
   1.2 Input-Output Models .......................................................... 9

2. State Space Models of Stationary Processes ......................... 10

3. Spectral Factorization .......................................................... 15
   4. Spectral Factorization and the LMI ...................................... 20
      4.1 Ordering, (A,C) Pairs and Uniform Choice of Basis in X .... 25

5. Finite-Interval Realizations of a Stationary Process ............... 28
   5.1 Forward and Backward Kalman Filtering and the Family of
       Minimal Stationary Realizations of y .................................. 28
   5.2 Finite-Interval Realizations ................................................ 32

6. Estimation, Partial Realization and Balancing ....................... 37
   6.1 Positivity ......................................................................... 40
   6.2 The Hilbert Space of a Stationary Signal ............................ 43
   6.3 Identification Based on Finite Data .................................... 45
   6.4 The Partial Realization Problem ........................................ 46
   6.5 Partial Realization via SVD .............................................. 49
   6.6 Stochastic Balanced Realizations: the Stationary Setting ....... 51
   6.7 Stochastic Balanced Realizations: the Case of Finite Data ..... 56

7. The "Subspace Methods" Identification Algorithm of Van Overschee
   and DeMoor ........................................................................ 56
   7.1 Choosing Bases in the Predictor Spaces ............................... 57
   7.2 Skipping some Redundant Steps ........................................ 62
   7.3 The Least Squares Implementation ..................................... 63
   7.4 Use of the SVD ............................................................... 64

Parameter Estimation of Multivariable Systems Using
Balanced Realizations
J. M. Maciejowski ...................................................................... 70

1. Introduction .......................................................................... 70
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Problem Setting</td>
<td>71</td>
</tr>
<tr>
<td>3. Identifiable Parametrizations</td>
<td>76</td>
</tr>
<tr>
<td>4. Balanced Parametrization</td>
<td>78</td>
</tr>
<tr>
<td>5. Some Useful Classes of Models</td>
<td>81</td>
</tr>
<tr>
<td>5.1 Minimum-Phase Models</td>
<td>81</td>
</tr>
<tr>
<td>5.2 Positive-Real Models</td>
<td>83</td>
</tr>
<tr>
<td>6. Outline of Parameter Estimation</td>
<td>86</td>
</tr>
<tr>
<td>7. Gradient Calculations</td>
<td>88</td>
</tr>
<tr>
<td>8. Finding an Initial Model</td>
<td>91</td>
</tr>
<tr>
<td>8.1 Available Methods</td>
<td>91</td>
</tr>
<tr>
<td>8.2 Realization Methods</td>
<td>92</td>
</tr>
<tr>
<td>8.3 Subspace Methods</td>
<td>98</td>
</tr>
<tr>
<td>8.4 Guaranteeing Stability</td>
<td>104</td>
</tr>
<tr>
<td>8.5 Estimating $n$</td>
<td>106</td>
</tr>
<tr>
<td>9. Examples</td>
<td>107</td>
</tr>
<tr>
<td>9.1 Distillation Column</td>
<td>107</td>
</tr>
<tr>
<td>9.2 Industrial Dryer</td>
<td>111</td>
</tr>
<tr>
<td>9.3 Sea Wave Spectrum</td>
<td>112</td>
</tr>
<tr>
<td>10. Conclusions</td>
<td>114</td>
</tr>
</tbody>
</table>

**Balanced Canonical Forms**
Raimund J. Ober

1. Introduction                                                        | 120  |
2. Lyapunov Balanced Realizations and Model Reduction                  | 126  |
3. A Lyapunov Balanced Canonical Form for Stable Continuous-Time Systems | 129  |
4. L-Characteristic, LQG-Balanced Canonical Form and Model Reduction for Minimal Systems | 146  |
5. Characteristics, Canonical Forms and Model Reduction for Bounded-Real and Positive-Real Systems | 162  |
6. Concluding Remarks                                                  | 179  |

**From Data to State Model**
Paolo Rapisarda and Jan C. Willems

1. Introduction                                                        | 184  |
2. Background                                                          | 187  |
2.1 Discrete Time Systems                                              | 187  |
2.2 Latent Variables                                                   | 191  |
2.3 State Models                                                       | 193  |
2.4 Existence and Uniqueness of State Space Models                     | 195  |
2.5 Input/State/Output, Output Nulling, and Driving Variable Representations | 196  |
2.6 Recapitulation                                                     | 199  |
3. From Difference Equation to State Models                            | 199  |
5. Nonlinear Algorithms .................................................. 275
  5.1 Two-Stage Nonlinear Algorithm .................................. 275
  5.2 Convex and Concave Windows ...................................... 277
  5.3 Frequency Domain Analysis ........................................ 277
  5.4 Trapezoidal Window .................................................. 278

6. Engineering Applications ............................................. 279

System Identification with Information Theoretic Criteria
A.A. Stoorvogel and J.H. van Schuppen ............................. 289
1. Introduction ............................................................. 289
2. Problem Formulation ................................................... 290
3. Approximation with Mutual Information ......................... 292
   3.1 Mutual Information .................................................. 292
   3.2 A Parameter Estimation Problem ................................ 294
   3.3 Relation of Mutual Information, $H_\infty$ Entropy, and LEQG Cost 295
   3.4 Parameter Estimation with an Exponential-of-Quadratic Cost 296
   3.5 Parameter Estimation with $H_\infty$ Entropy ................. 298
4. Approximation with Likelihood and Divergence ................. 300
   4.1 Approximation with the Likelihood Function ............... 300
   4.2 Divergence .......................................................... 301
   4.3 Relation of Likelihood Function and Divergence ............ 302
   4.4 Approximation with Divergence ................................ 304
   4.5 Parameter Estimation by Divergence Minimization ......... 306
5. Concluding Remarks .................................................. 308

A. Concepts from Probability and the Theory of Stochastic Processes 310
   A.1 Probability Concepts .............................................. 310
   A.2 Gaussian Random Variables ..................................... 310
   A.3 Concepts from the Theory of Stochastic Processes ......... 312

B. Concepts from System Theory ....................................... 313

C. Concepts from Information Theory ................................ 314

D. Information Measures of Gaussian Random Variables .......... 317

E. Information Measures of Stationary Gaussian Processes ...... 320

F. LEQG Optimal Stochastic Control .................................. 327

G. H-Infinity Control with an Entropy Criterion .................. 332

Least Squares Based Self-Tuning Control Systems
Sergio Bittanti and Marco Campi .................................... 339
1. Introduction ............................................................ 339
2. Self-Tuning Adaptive Control ....................................... 341
   2.1 Basic Self-Tuning Concepts ..................................... 341
   2.2 Mathematical Framework ........................................... 342
   3.1 Imaginary and Asymptotic Imaginary Systems ............... 345
   3.2 Adaptive Stabilization ............................................ 346
On Neural Network Model Structures in System Identification
L. Ljung, J. Sjöberg, and H. Hjalmarsson

1. Introduction and Summary
   1.1 What is the Problem?
   1.2 Black Boxes
   1.3 Nonlinear Black Box Models
   1.4 Estimating \( \hat{g}_N \)
   1.5 Properties of the Estimated Model
   1.6 Basis Functions
   1.7 What Is the Neural Network Identification Approach?
   1.8 Why Have Neural Networks Attracted So Much Interest?
   1.9 Related Approaches

2. The Problem
   2.1 Inferring Relationships from Data
   2.2 Prior Assumptions
   2.3 Function Classes

3. Some General Estimation Results

4. The Bias/Variance Trade-Off

5. Neural Nets
   5.1 Feedforward Neural Nets
   5.2 Recurrent Neural Nets

6. Algorithmic Aspects
   6.1 Search Directions
   6.2 Back-Propagation: Calculation of the Gradient
   6.3 Implicit Regularization
   6.4 Off-line and On-line Algorithms
   6.5 Local Minima

7. Adaptive Methods
   7.1 Adaptive Basis Function Expansion
   7.2 The “Curse” of Dimensionality
   7.3 Methods to Avoid the “Curse”
8. Specific Properties of NN Structures .......................................................... 388
9. Models of Dynamical Systems Based on Neural Networks ......................... 389
  9.1 A Review of Linear Black Box Models ..................................................... 389
  9.2 Choice of Regressors for Neural Network Models .................................... 391
  9.3 Neural Network Dynamic Models ............................................................ 392
  9.4 Some Other Structural Questions ............................................................ 393
  9.5 The Identification Procedure ................................................................. 393

An Overview of Computational Learning Theory and Its Applications to Neural Network Training
M. Vidyasagar .................................................. 400
1. Introduction ......................................................... 400
2. Problem Formulation .................................................. 401
3. Summary of Known Results .................................................. 403
4. Families of Measures with a Nonempty Interior ..................................... 407
5. Totally Bounded Families of Measures ............................................. 409
6. Two Sufficient Conditions .................................................. 414
7. Conclusions ......................................................... 416
Appendix: A Counterexample .................................................. 417

Just-in-Time Learning and Estimation
George Cybenko .................................................. 423
1. Introduction ......................................................... 423
2. Global Models ....................................................... 425
3. Local Models ......................................................... 427
4. Just-In-Time Models .................................................. 428
  4.1 Analysis of Just-In-Time Models .................................................. 429
  4.2 Discussion ......................................................... 432

Wavelets in Identification
A. Benveniste, A. Juditsky, B. Delyon, Q. Zhang, and P-Y. Glorennec 435
1. Introduction, Motivations, Basic Problems ............................................ 435
  1.1 Two Application Examples .................................................. 436
2. Basic Mathematical Problems .................................................. 438
3. Classical Methods of Nonlinear System Identification:
   Linear Nonparametric Estimators .................................................. 440
  3.1 Projection Estimates as an Example of Linear Nonparametric Estimators .. 441
  3.2 Choice of Model Order, Bandwidth, or Binwidth:
      the Generalized Cross Validation – GCV – Method ............................. 442
4. Performance Analysis of the Nonparametric Estimators .......................... 443
  4.1 Lower Bounds for Best Achievable Performance .................................. 444
  4.2 Discussion ......................................................... 446
Table of Contents

5. Nonlinear Estimates ............................................. 447

6. Wavelets: What They Are, and Their Use in
   Approximating Functions ........................................ 449
   6.1 The Continuous Wavelet Transform ......................... 449
   6.2 The Discrete Wavelet Transform:
       Orthonormal Bases of Wavelets and Extensions ........... 450
   6.3 Wavelets and Besov Spaces .................................. 453

7. Wavelets: Their Use in Nonparametric Estimation ............ 456

8. A Wavelet Network for Practical System Identification ........ 458
   8.1 The Wavelet Network and Its Structure .................... 459
   8.2 Constructing the Wavelet Library $W$ ....................... 459
   8.3 Selecting Best Wavelet Regressors ......................... 460

9. Fuzzy Models: Expressing Prior Information in
   Nonlinear Nonparametric Models ............................... 462

10. Experimental Results ............................................. 464
    10.1 Modelling the Gas Turbine System ......................... 464
    10.2 Modelling the Hydraulic Actuator of the Robot Arm ....... 465

11. Discussion and Conclusions ..................................... 470

Fuzzy Logic Modelling and Control

P. Albertos ......................................................... 479

1. Motivation ....................................................... 479

2. Fuzzy Logic Basic Concepts .................................... 481
   2.1 Fuzzy Logic Variables ....................................... 482
   2.2 Fuzzy Logic Operations ..................................... 485
   2.3 Approximated Reasoning ..................................... 486
   2.4 Defuzzification .............................................. 490
   2.5 Fuzzy Systems ................................................ 492

3. Fuzzy Logic Controller Structure ............................... 494
   3.1 Fuzzifier .................................................... 495
   3.2 Fuzzy Operator .............................................. 495
   3.3 Defuzzification .............................................. 497

4. FLC Analysis ..................................................... 497
   4.1 Fuzzy Systems Approximation Properties .................... 499

5. Fuzzy Logic Controllers Design ................................ 500
   5.1 Experimental FLC Design .................................... 501
   5.2 FLC from Linear Ones ........................................ 502
   5.3 FLC Supervision ............................................. 503

6. Process Fuzzy Modelling ........................................ 504

7. Cement Kiln Control ............................................. 505
   7.1 Process Description .......................................... 506
   7.2 Control Structure ............................................ 507
Searching for the Best: Stochastic Approximation, Simulated Annealing and Related Procedures
Georg Pflug .................................................. 514

1. Introduction ................................................. 514
2. Deterministic Iterations, Fixpoints and Convergence ........ 515
3. Stochastic Iterations, Stationary Distributions and Convergence ... 517
4. Simulated Annealing and Global Optimization .................. 522
5. Stochastic Approximation .................................. 524
6. Robustness of Dynamical Systems .......................... 534
7. Parallelization ............................................. 537
   7.1 Stochastic Ordering .................................... 538
8. Response Surface Methods .................................. 543
9. Bandit Processes .......................................... 544

List of Contributors ........................................... 550