

Classification, Parameter Estimation and State Estimation

AN ENGINEERING APPROACH USING MATLAB

SECOND EDITION



BANGJUN LEI • GUANGZHU XU • MING FENG • YAOBIN ZOU
FERDINAND VAN DER HEIJDEN • DICK DE RIDDER • DAVID M. J. TAX

WILEY

**Classification, Parameter
Estimation and State Estimation**

Classification, Parameter Estimation and State Estimation

An Engineering Approach Using MATLAB

Second Edition

Bangjun Lei

Guangzhu Xu

Ming Feng

Yaobin Zou

Ferdinand van der Heijden

Dick de Ridder

David M. J. Tax

WILEY

This edition first published 2017
© 2017 John Wiley & Sons, Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Bangjun Lei, Dick de Ridder, David M. J. Tax, Ferdinand van der Heijden, Guangzhu Xu, Ming Feng, Yaobin Zou to be identified as the authors of this work has been asserted in accordance with law.

Registered Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

John Wiley & Sons, Ltd., The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty:

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This work's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Heijden, Ferdinand van der. | Lei, Bangjun, 1973– author. | Xu, Guangzhu, 1979– author. | Ming, Feng, 1957– author. | Zou, Yaobin, 1978– author. | Ridder, Dick de, 1971– author. | Tax, David M. J., 1973– author.

Title: Classification, parameter estimation, and state estimation : an engineering approach using MATLAB / Bangjun Lei, Guangzhu Xu, Ming Feng, Yaobin Zou, Ferdinand van der Heijden, Dick de Ridder, David M. J. Tax.

Description: Second edition. | Hoboken, NJ, USA : John Wiley & Sons, Inc., 2017. | Revised edition of: Classification, parameter estimation, and state estimation : an engineering approach using MATLAB / F. van der Heijden ... [et al.]. 2004. | Includes bibliographical references and index.

Identifiers: LCCN 2016059294 (print) | LCCN 2016059809 (ebook) | ISBN 9781119152439 (cloth) | ISBN 9781119152446 (pdf) | ISBN 9781119152453 (epub)

Subjects: LCSH: Engineering mathematics—Data processing. | MATLAB. | Measurement—Data processing. | Estimation theory—Data processing.

Classification: LCC TA331 .C53 2017 (print) | LCC TA331 (ebook) | DDC 681/.2—dc23

LC record available at <https://lcn.loc.gov/2016059294>

Cover Design: Wiley

Cover Images: neural network © maxuser/Shutterstock; digital circuit board

© Powderblue/Shutterstock

Set in 10/12pt WarnockPro by Aptara Inc., New Delhi, India

Printed in Great Britain by TJ International Ltd, Padstow, Cornwall

10 9 8 7 6 5 4 3 2 1

Contents

Preface *xi*

About the Companion Website *xv*

1	Introduction	1
1.1	The Scope of the Book	2
1.1.1	Classification	3
1.1.2	Parameter Estimation	4
1.1.3	State Estimation	5
1.1.4	Relations between the Subjects	7
1.2	Engineering	10
1.3	The Organization of the Book	12
1.4	Changes from First Edition	14
1.5	References	15
2	PRTools Introduction	17
2.1	Motivation	17
2.2	Essential Concepts	18
2.3	PRTools Organization Structure and Implementation	22
2.4	Some Details about PRTools	26
2.4.1	Datasets	26
2.4.2	Datafiles	30
2.4.3	Datafiles Help Information	31
2.4.4	Classifiers and Mappings	34
2.4.5	Mappings Help Information	36
2.4.6	How to Write Your Own Mapping	38
2.5	Selected Bibliography	42

3	Detection and Classification	43
3.1	Bayesian Classification	46
3.1.1	Uniform Cost Function and Minimum Error Rate	53
3.1.2	Normal Distributed Measurements; Linear and Quadratic Classifiers	56
3.2	Rejection	62
3.2.1	Minimum Error Rate Classification with Reject Option	63
3.3	Detection: The Two-Class Case	66
3.4	Selected Bibliography	74
	Exercises	74
4	Parameter Estimation	77
4.1	Bayesian Estimation	79
4.1.1	MMSE Estimation	86
4.1.2	MAP Estimation	87
4.1.3	The Gaussian Case with Linear Sensors	88
4.1.4	Maximum Likelihood Estimation	89
4.1.5	Unbiased Linear MMSE Estimation	91
4.2	Performance Estimators	94
4.2.1	Bias and Covariance	95
4.2.2	The Error Covariance of the Unbiased Linear MMSE Estimator	99
4.3	Data Fitting	100
4.3.1	Least Squares Fitting	101
4.3.2	Fitting Using a Robust Error Norm	104
4.3.3	Regression	107
4.4	Overview of the Family of Estimators	110
4.5	Selected Bibliography	111
	Exercises	112
5	State Estimation	115
5.1	A General Framework for Online Estimation	117
5.1.1	Models	117
5.1.2	Optimal Online Estimation	123
5.2	Infinite Discrete-Time State Variables	125
5.2.1	Optimal Online Estimation in Linear-Gaussian Systems	125
5.2.2	Suboptimal Solutions for Non-linear Systems	133

5.3	Finite Discrete-Time State Variables	147
5.3.1	Hidden Markov Models	148
5.3.2	Online State Estimation	152
5.3.3	Offline State Estimation	156
5.4	Mixed States and the Particle Filter	163
5.4.1	Importance Sampling	164
5.4.2	Resampling by Selection	166
5.4.3	The Condensation Algorithm	167
5.5	Genetic State Estimation	170
5.5.1	The Genetic Algorithm	170
5.5.2	Genetic State Estimation	176
5.5.3	Computational Issues	177
5.6	State Estimation in Practice	183
5.6.1	System Identification	185
5.6.2	Observability, Controllability and Stability	188
5.6.3	Computational Issues	193
5.6.4	Consistency Checks	196
5.7	Selected Bibliography	201
	Exercises	204
6	Supervised Learning	207
6.1	Training Sets	208
6.2	Parametric Learning	210
6.2.1	Gaussian Distribution, Mean Unknown	211
6.2.2	Gaussian Distribution, Covariance Matrix Unknown	212
6.2.3	Gaussian Distribution, Mean and Covariance Matrix Both Unknown	213
6.2.4	Estimation of the Prior Probabilities	215
6.2.5	Binary Measurements	216
6.3	Non-parametric Learning	217
6.3.1	Parzen Estimation and Histogramming	218
6.3.2	Nearest Neighbour Classification	223
6.3.3	Linear Discriminant Functions	230
6.3.4	The Support Vector Classifier	237
6.3.5	The Feedforward Neural Network	242
6.4	Adaptive Boosting – Adaboost	245
6.5	Convolutional Neural Networks (CNNs)	249
6.5.1	Convolutional Neural Network Structure	249
6.5.2	Computation and Training of CNNs	251

6.6	Empirical Evaluation	252
6.7	Selected Bibliography	257
	Exercises	257
7	Feature Extraction and Selection	259
7.1	Criteria for Selection and Extraction	261
7.1.1	Interclass/Intraclass Distance	262
7.1.2	Chernoff–Bhattacharyya Distance	267
7.1.3	Other Criteria	270
7.2	Feature Selection	272
7.2.1	Branch-and-Bound	273
7.2.2	Suboptimal Search	275
7.2.3	Several New Methods of Feature Selection	278
7.2.4	Implementation Issues	287
7.3	Linear Feature Extraction	288
7.3.1	Feature Extraction Based on the Bhattacharyya Distance with Gaussian Distributions	291
7.3.2	Feature Extraction Based on Inter/Intra Class Distance	296
7.4	References	300
	Exercises	300
8	Unsupervised Learning	303
8.1	Feature Reduction	304
8.1.1	Principal Component Analysis	304
8.1.2	Multidimensional Scaling	309
8.1.3	Kernel Principal Component Analysis	315
8.2	Clustering	320
8.2.1	Hierarchical Clustering	323
8.2.2	K-Means Clustering	327
8.2.3	Mixture of Gaussians	329
8.2.4	Mixture of probabilistic PCA	335
8.2.5	Self-Organizing Maps	336
8.2.6	Generative Topographic Mapping	342
8.3	References	345
	Exercises	346
9	Worked Out Examples	349
9.1	Example on Image Classification with PRTTools	349
9.1.1	Example on Image Classification	349

9.1.2	Example on Face Classification	354
9.1.3	Example on Silhouette Classification	357
9.2	Boston Housing Classification Problem	361
9.2.1	Dataset Description	361
9.2.2	Simple Classification Methods	363
9.2.3	Feature Extraction	365
9.2.4	Feature Selection	367
9.2.5	Complex Classifiers	368
9.2.6	Conclusions	371
9.3	Time-of-Flight Estimation of an Acoustic Tone Burst	372
9.3.1	Models of the Observed Waveform	374
9.3.2	Heuristic Methods for Determining the ToF	376
9.3.3	Curve Fitting	377
9.3.4	Matched Filtering	379
9.3.5	ML Estimation Using Covariance Models for the Reflections	380
9.3.6	Optimization and Evaluation	385
9.4	Online Level Estimation in a Hydraulic System	392
9.4.1	Linearized Kalman Filtering	394
9.4.2	Extended Kalman Filtering	397
9.4.3	Particle Filtering	398
9.4.4	Discussion	403
9.5	References	406

Appendix A: Topics Selected from Functional Analysis 407

Appendix B: Topics Selected from Linear Algebra and Matrix Theory 421

Appendix C: Probability Theory 437

Appendix D: Discrete-Time Dynamic Systems 453

Index 459

Preface

Information processing has always been an important factor in the development of human society and its role is still increasing. The inventions of advanced information devices paved the way for achievements in a diversity of fields like trade, navigation, agriculture, industry, transportation and communication. The term ‘information device’ refers here to systems for the sensing, acquisition, processing and outputting of information from the real world. Usually, they are measurement systems. Sensing and acquisition provide us with signals that bear a direct relation to some of the physical properties of the sensed object or process. Often, the information of interest is hidden in these signals. Signal processing is needed to reveal the information and to transform it into an explicit form. Further, in the past 10 years image processing (together with intelligent computer vision) has gone through rapid developments. There are substantial new developments on, for example, machine learning methods (such as Adaboost and its varieties, Deep learning etc.) and particle filtering like parameter estimation methods.

The three topics discussed in this book, classification, parameter estimation and state estimation, share a common factor in the sense that each topic provides the theory and methodology for the functional design of the signal processing part of an information device. The major distinction between the topics is the type of information that is outputted. In classification problems the output is discrete, that is a class, a label or a category. In estimation problems, it is a real-valued scalar or vector. Since these problems occur either in a static or in a dynamic setting, actually four different topics can be distinguished. The term state

estimation refers to the dynamic setting. It covers both discrete and real-valued cases (and sometimes even mixed cases).

The similarity between the topics allows one to use a generic methodology, that is Bayesian decision theory. Our aim is to present this material concisely and efficiently by an integrated treatment of similar topics. We present an overview of the core mathematical constructs and the many resulting techniques. By doing so, we hope that the reader recognizes the connections and the similarities between these constructs, but also becomes aware of the differences. For instance, the phenomenon of overfitting is a threat that ambushes all four cases. In a static classification problem it introduces large classification errors, but in the case of a dynamic state estimation it may be the cause of instable behaviour. Further, in this edition, we made some modifications to accommodate engineering requests on intelligent computer vision.

Our goal is to emphasize the engineering aspects of the matter. Instead of a purely theoretical and rigorous treatment, we aim for the acquirement of skills to bring theoretical solutions to practice. The models that are needed for the application of the Bayesian framework are often not available in practice. This brings in the paradigm of statistical inference, that is learning from examples. MATLAB[®]* is used as a vehicle to implement and to evaluate design concepts.

As alluded to above, the range of application areas is broad. Application fields are found within computer vision, mechanical engineering, electrical engineering, civil engineering, environmental engineering, process engineering, geo-informatics, bio-informatics, information technology, mechatronics, applied physics, and so on. The book is of interest to a range of users, from the first-year graduate-level student up to the experienced professional. The reader should have some background knowledge with respect to linear algebra, dynamic systems and probability theory. Most educational programmes offer courses on these topics as part of undergraduate education. The appendices contain reviews of the relevant material. Another target group

* MATLAB[®] is a registered trademark of The MathWorks, Inc. (<http://www.mathworks.com>).

is formed by the experienced engineers working in industrial development laboratories. The numerous examples of MATLAB[®] code allow these engineers to quickly prototype their designs.

The book roughly consists of three parts. The first part, Chapter 2, presents an introduction to the PRTools used throughout this book. The second part, Chapters 3, 4 and 5, covers the theory with respect to classification and estimation problems in the static case, as well as the dynamic case. This part handles problems where it is assumed that accurate models, describing the physical processes, are available. The third part, Chapters 6 up to 8, deals with the more practical situation in which these models are not or only partly available. Either these models must be built using experimental data or these data must be used directly to train methods for estimation and classification. The final chapter presents three worked out problems. The selected bibliography has been kept short in order not to overwhelm the reader with an enormous list of references.

The material of the book can be covered by two semester courses. A possibility is to use Chapters 3, 4, 6, 7 and 8 for a one-semester course on Classification and Estimation. This course deals with the static case. An additional one-semester course handles the dynamic case, that is Optimal Dynamic Estimation, and would use Chapter 5. The prerequisites for Chapter 5 are mainly concentrated in Chapter 4. Therefore, it is recommended to include a review of Chapter 4 in the second course. Such a review will make the second course independent from the first one.

Each chapter is closed with a number of exercises. The mark at the end of each exercise indicates whether the exercise is considered easy ('0'), moderately difficult ('*') or difficult ('**'). Another possibility to acquire practical skills is offered by the projects that accompany the text. These projects are available at the companion website. A project is an extensive task to be undertaken by a group of students. The task is situated within a given theme, for instance, classification using supervised learning, unsupervised learning, parameter estimation, dynamic labelling and dynamic estimation. Each project consists of a set of instructions together with data that should be used to solve the problem.

The use of MATLAB[®] tools is an integrated part of the book. MATLAB[®] offers a number of standard toolboxes that are useful for parameter estimation, state estimation and data analysis. The standard software for classification and unsupervised learning is not complete and not well structured. This motivated us to develop the PRTools software for all classification tasks and related items. PRTools is a MATLAB[®] toolbox for pattern recognition. It is freely available for non-commercial purposes. The version used in the text is compatible with MATLAB[®] Version 5 and higher. It is available from <http://37steps.com>.

The authors keep an open mind for any suggestions and comments (which should be addressed to cpese@wiley.com). A list of errata and any other additional comments will be made available at the companion website.

Acknowledgements

We thank everyone who has made this book possible. Special thanks are given to Dr. Robert P. W. Duin for his contribution to the first version of this book and for allowing us to use PRTools and all materials on 37steps.com throughout this book. Thanks are also extended to Dr. Ela Pekalska for the courtesy of sharing documents of 37steps.com with us.

About the Companion Website

This book is accompanied by a companion website:



www.wiley.com/go/vanderheijden/classification_parameterestimation_stateestimation/

The website includes:

- Code and Datasets
- Tutorials
- Pictures
- Errata

1

Introduction

Engineering disciplines are those fields of research and development that attempt to create products and systems operating in, and dealing with, the real world. The number of disciplines is large, as is the range of scales that they typically operate in: from the very small scale of nanotechnology up to very large scales that span whole regions, for example water management systems, electric power distribution systems or even global systems (e.g. the global positioning system, GPS). The level of advancement in the fields also varies wildly, from emerging techniques (again, nanotechnology) to trusted techniques that have been applied for centuries (architecture, hydraulic works). Nonetheless, the disciplines share one important aspect: engineering aims at designing and manufacturing systems that interface with the world around them.

Systems designed by engineers are often meant to influence their environment: to manipulate it, to move it, to stabilize it, to please it, and so on. To enable such actuation, these systems need information, for example values of physical quantities describing their environments and possibly also describing themselves. Two types of information sources are available: *prior* knowledge and *empirical* knowledge. The latter is knowledge obtained by sensorial observation. Prior knowledge is the knowledge that was already there before a given observation became available (this does not imply that prior knowledge is obtained without any observation). The combination of prior knowledge and empirical knowledge leads to *posterior* knowledge.

The sensory subsystem of a system produces measurement signals. These signals carry the empirical knowledge. Often, the direct usage of these signals is not possible, or is inefficient. This can have several causes:

- The information in the signals is not represented in an explicit way. It is often hidden and only available in an indirect, encoded, form.
- Measurement signals always come with noise and other hard-to-predict disturbances.
- The information brought forth by posterior knowledge is more accurate and more complete than information brought forth by empirical knowledge alone. Hence, measurement signals should be used in combination with prior knowledge.

Measurement signals need processing in order to suppress the noise and to disclose the information required for the task at hand.

1.1 The Scope of the Book

In a sense, classification and estimation deal with the same problem: given the measurement signals from the environment, how can the information that is needed for a system to operate in the real world be inferred? In other words, how should the measurements from a sensory system be processed in order to bring maximal information in an explicit and usable form? This is the main topic of this book.

Good processing of the measurement signals is possible only if some knowledge and understanding of the environment and the sensory system is present. Modelling certain aspects of that environment – like objects, physical processes or events – is a necessary task for the engineer. However, straightforward modelling is not always possible. Although the physical sciences provide ever deeper insight into nature, some systems are still only partially understood; just think of the weather. Even if systems are well understood, modelling them exhaustively may be beyond our current capabilities (i.e. computer power) or beyond the scope of the application. In such cases, approximate general models,



Figure 1.1 Licence plate recognition: a classification problem with noisy measurements.

but adapted to the system at hand, can be applied. The development of such models is also a topic of this book.

1.1.1 Classification

The title of the book already indicates the three main subtopics it will cover: classification, parameter estimation and state estimation. In classification, one tries to assign a class label to an object, a physical process or an event. Figure 1.1 illustrates the concept. In a speeding detector, the sensors are a radar speed detector and a high-resolution camera, placed in a box beside a road. When the radar detects a car approaching at too high a velocity (a parameter estimation problem), the camera is signalled to acquire an image of the car. The system should then recognize the licence plate, so that the driver of the car can be fined for the speeding violation. The system should be robust to differences in car model, illumination, weather circumstances, etc., so some pre-processing is necessary: locating the licence plate in the image, segmenting the individual characters and converting it into a binary image. The problem then breaks down to a number of individual classification problems. For each of the locations on the license plate, the input consists of a binary image of a character, normalized for size, skew/rotation and intensity. The desired output is the label of the true character, that is one of 'A', 'B', ..., 'Z', '0', ..., '9'.

Detection is a special case of classification. Here, only two class labels are available, for example 'yes' and 'no'. An example is a quality control system that approves the products of a manufacturer or refuses them. A second problem closely related to classification is identification: the act of proving that an object-under-test and a second object that is previously seen are the same. Usually, there is a large database of previously seen objects to choose from. An example is biometric identification, for example fingerprint recognition or face recognition. A third problem that can be solved by classification-like techniques is retrieval from a database, for example finding an image in an image database by specifying image features.

1.1.2 Parameter Estimation

In parameter estimation, one tries to derive a parametric description for an object, a physical process or an event. For example, in a beacon-based position measurement system (Figure 1.2), the goal is to find the position of an object, for example a ship or a mobile robot. In the two-dimensional case, two beacons with known reference positions suffice. The sensory system provides two measurements: the distances from the beacons to the object, r_1 and r_2 . Since the position of the object involves two parameters, the estimation seems to boil down to solving two equations with two unknowns. However,

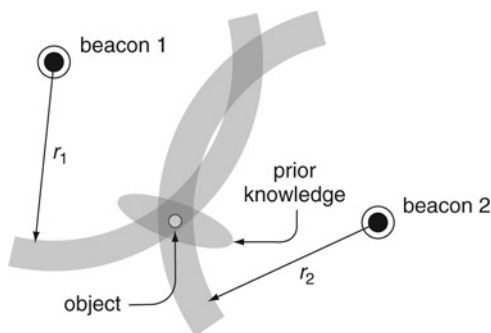


Figure 1.2 Position measurement: a parameter estimation problem handling uncertainties.

the situation is more complex because measurements always come with uncertainties. Usually, the application not only requires an estimate of the parameters but also an assessment of the uncertainty of that estimate. The situation is even more complicated because some prior knowledge about the position must be used to resolve the ambiguity of the solution. The prior knowledge can also be used to reduce the uncertainty of the final estimate.

In order to improve the accuracy of the estimate the engineer can increase the number of (independent) measurements to obtain an overdetermined system of equations. In order to reduce the cost of the sensory system, the engineer can also decrease the number of measurements, leaving us with fewer measurements than parameters. The system of equations is then underdetermined, but estimation is still possible if enough prior knowledge exists or if the parameters are related to each other (possibly in a statistical sense). In either case, the engineer is interested in the uncertainty of the estimate.

1.1.3 State Estimation

In state estimation, one tries to do either of the following – either assigning a class label or deriving a parametric (real-valued) description – but for processes that vary in time or space. There is a fundamental difference between the problems of classification and parameter estimation, on the one hand, and state estimation, on the other hand. This is the ordering in time (or space) in state estimation, which is absent from classification and parameter estimation. When no ordering in the data is assumed, the data can be processed in any order. In time series, ordering in time is essential for the process. This results in a fundamental difference in the treatment of the data.

In the discrete case, the states have discrete values (classes or labels) that are usually drawn from a finite set. An example of such a set is the alarm stages in a safety system (e.g. ‘safe’, ‘pre-alarm’, ‘red alert’, etc.). Other examples of discrete state estimation are speech recognition, printed or handwritten text recognition and the recognition of the operating modes of a machine.

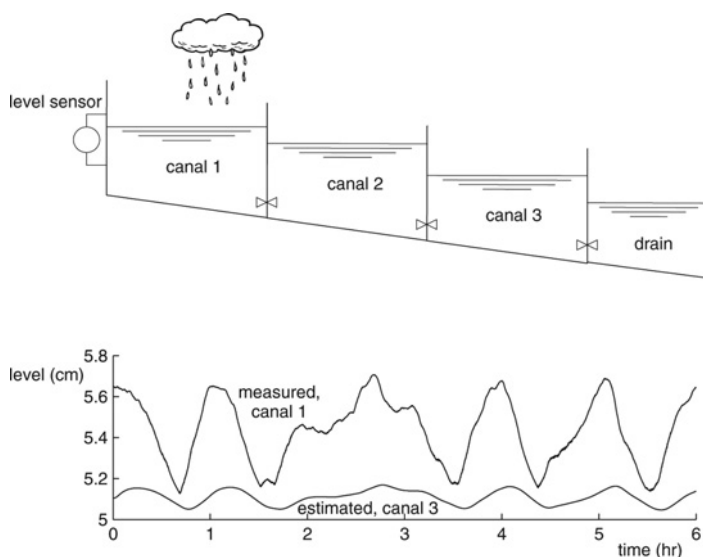


Figure 1.3 Assessment of water levels in a water management system: a state estimation problem (the data are obtained from a scale model).

An example of real-valued state estimation is the water management system of a region. Using a few level sensors and an adequate dynamical model of the water system, a state estimator is able to assess the water levels even at locations without level sensors. Short-term prediction of the levels is also possible. Figure 1.3 gives a view of a simple water management system of a single canal consisting of three linearly connected compartments. The compartments are filled by the precipitation in the surroundings of the canal. This occurs randomly but with a seasonal influence. The canal drains its water into a river. The measurement of the level in one compartment enables the estimation of the levels in all three compartments. For that, a dynamic model is used that describes the relations between flows and levels. Figure 1.3 shows an estimate of the level of the third compartment using measurements of the level in the first compartment. Prediction of the level in the third compartment is possible due to the causality of the process and the delay between the levels in the compartments.

1.1.4 Relations between the Subjects

The reader who is familiar with one or more of the three subjects might wonder why they are treated in one book. The three subjects share the following factors:

- In all cases, the engineer designs an instrument, that is a system whose task is to extract information about a real-world object, a physical process or an event.
- For that purpose, the instrument will be provided with a sensory subsystem that produces measurement signals. In all cases, these signals are represented by vectors (with fixed dimension) or sequences of vectors.
- The measurement vectors must be processed to reveal the information that is required for the task at hand.
- All three subjects rely on the availability of models describing the object/physical process/event and of models describing the sensory system.
- Modelling is an important part of the design stage. The suitability of the applied model is directly related to the performance of the resulting classifier/estimator.

Since the nature of the questions raised in the three subjects is similar, the analysis of all three cases can be done using the same framework. This allows an economical treatment of the subjects. The framework that will be used is a probabilistic one. In all three cases, the strategy will be to formulate the posterior knowledge in terms of a conditional probability (density) function:

$$P(\text{quantities of interest measurements available})$$

This so-called posterior probability combines the prior knowledge with the empirical knowledge by using Bayes' theorem for conditional probabilities. As discussed above, the framework is generic for all three cases. Of course, the elaboration of this principle for the three cases leads to different solutions because the nature of the 'quantities of interest' differs.

The second similarity between the topics is their reliance on models. It is assumed that the constitution of the object/physical process/event (including the sensory system) can be captured by a mathematical model. Unfortunately, the physical structures responsible for generating the objects/process/events are often

unknown, or at least partly unknown. Consequently, the model is also, at least partly, unknown. Sometimes, some functional form of the model is assumed, but the free parameters still have to be determined. In any case, empirical data are needed in order to establish the model, to tune the classifier/estimator-under-development and also to evaluate the design. Obviously, the training/evaluation data should be obtained from the process we are interested in.

In fact, all three subjects share the same key issue related to modelling, namely the selection of the appropriate generalization level. The empirical data are only an example of a set of possible measurements. If too much weight is given to the data at hand, the risk of overfitting occurs. The resulting model will depend too much on the accidental peculiarities (or noise) of the data. On the other hand, if too little weight is given, nothing will be learned and the model completely relies on the prior knowledge. The right balance between these opposite sides depends on the statistical significance of the data. Obviously, the size of the data is an important factor. However, the statistical significance also holds a relation with dimensionality.

Many of the mathematical techniques for modelling, tuning, training and evaluation can be shared between the three subjects. Estimation procedures used in classification can also be used in parameter estimation or state estimation, with just minor modifications. For instance, probability density estimation can be used for classification purposes and also for estimation. Data-fitting techniques are applied in both classification and estimation problems. Techniques for statistical inference can also be shared. Of course, there are also differences between the three subjects. For instance, the modelling of dynamic systems, usually called *system identification*, involves aspects that are typical for dynamic systems (i.e. determination of the order of the system, finding an appropriate functional structure of the model). However, when it finally comes to finding the right parameters of the dynamic model, the techniques from parameter estimation apply again.

Figure 1.4 shows an overview of the relations between the topics. Classification and parameter estimation share a common foundation indicated by 'Bayes'. In combination with models for dynamic systems (with random inputs), the techniques for

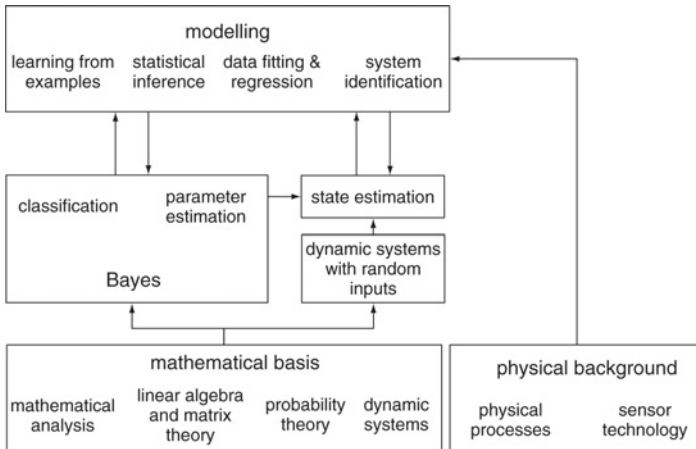


Figure 1.4 Relations between the subjects.

classification and parameter estimation find their application in processes that proceed in time, that is state estimation. All this is built on a mathematical basis with selected topics from mathematical analysis (dealing with abstract vector spaces, metric spaces and operators), linear algebra and probability theory. As such, classification and estimation are not tied to a specific application. The engineer, who is involved in a specific application, should add the individual characteristics of that application by means of the models and prior knowledge. Thus, apart from the ability to handle empirical data, the engineer must also have some knowledge of the physical background related to the application at hand and to the sensor technology being used.

All three subjects are mature research areas and many overview books have been written. Naturally, by combining the three subjects into one book, it cannot be avoided that some details are left out. However, the discussion above shows that the three subjects are close enough to justify one integrated book covering these areas.

The combination of the three topics into one book also introduces some additional challenges if only because of the differences in terminology used in the three fields. This is, for instance, reflected in the difference in the term used for ‘measurements’.

In classification theory, the term ‘features’ is frequently used as a replacement for ‘measurements’. The number of measurements is called the ‘dimension’, but in classification theory the term ‘dimensionality’ is often used.¹ The same remark holds true for notations. For instance, in classification theory the measurements are often denoted by \mathbf{x} . In state estimation, two notations are in vogue: either \mathbf{y} or \mathbf{z} (MATLAB[®] uses \mathbf{y} , but we chose \mathbf{z}). In all cases we tried to be as consistent as possible.

1.2 Engineering

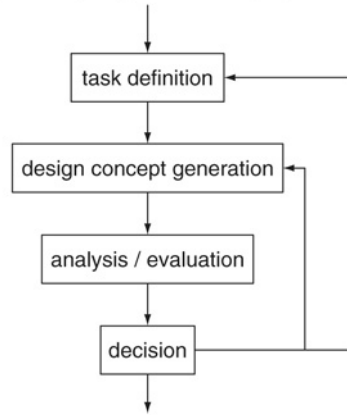
The top-down design of an instrument always starts with some primary need. Before starting with the design, the engineer has only a global view of the system of interest. The actual need is known only at a high and abstract level. The design process then proceeds through a number of stages during which progressively more detailed knowledge becomes available and the system parts of the instrument are described at lower and more concrete levels. At each stage, the engineer has to make design decisions. Such decisions must be based on explicitly defined evaluation criteria. The procedure, the elementary design step, is shown in Figure 1.5. It is used iteratively at the different levels and for the different system parts.

An elementary design step typically consists of collecting and organizing knowledge about the design issue of that stage, followed by an explicit formulation of the involved task. The next step is to associate the design issue with an evaluation criterion. The criterion expresses the suitability of a design concept related to the given task, but also other aspects can be involved, such as cost of manufacturing, computational cost or throughput.

¹ Our definition complies with the mathematical definition of ‘dimension’, i.e. the maximal number of independent vectors in a vector space. In MATLAB[®] the term ‘dimension’ refers to an index of a multidimensional array as in phrases like: ‘the first dimension of a matrix is the row index’ and ‘the number of dimensions of a matrix is two’. The number of elements along a row is the ‘row dimension’ or ‘row length’. In MATLAB[®] the term ‘dimensionality’ is the same as the ‘number of dimensions’.

Figure 1.5 An elementary step in the design process (Finkelstein and Finkelstein, 1994).

from preceding stage of the design process



to next stage of the design process

Usually, there are a number of possible design concepts to select from. Each concept is subjected to an analysis and an evaluation, possibly based on some experimentation. Next, the engineer decides which design concept is most appropriate. If none of the possible concepts are acceptable, the designer steps back to an earlier stage to alter the selections that have been made there.

One of the first tasks of the engineer is to identify the actual need that the instrument must fulfil. The outcome of this design step is a description of the functionality, for example a list of preliminary specifications, operating characteristics, environmental conditions, wishes with respect to user interface and exterior design. The next steps deal with the principles and methods that are appropriate to fulfil the needs, that is the *internal functional structure* of the instrument. At this level, the system under design is broken down into a number of functional components. Each component is considered as a subsystem whose input/output relations are mathematically defined. Questions related to the actual construction, realization of the functions, housing, etc., are later concerns.

The functional structure of an instrument can be divided roughly into sensing, processing and outputting (displaying,

recording). This book focuses entirely on the design steps related to *processing*. It provides:

- Knowledge about various methods to fulfil the processing tasks of the instrument. This is needed in order to generate a number of different design concepts.
- Knowledge about how to evaluate the various methods. This is needed in order to select the best design concept.
- A tool for the experimental evaluation of the design concepts.

The book does not address the topic ‘sensor technology’. For this, many good textbooks already exist, for instance see Regtien *et al.* (2004) and Brignell and White (1996). Nevertheless, the sensory system does have a large impact on the required processing. For our purpose, it suffices to consider the sensory subsystem at an abstract functional level such that it can be described by a mathematical model.

1.3 The Organization of the Book

Chapter 2 focuses on the introduction of PRTools designed by Robert P.W.Duin. PRTools is a pattern recognition toolbox for MATLAB[®] freely available for non-commercial use. The pattern recognition routines and support functions offered by PRTools represent a basic set covering largely the area of statistical pattern recognition. In this book, except for additional notes, all examples are based on PRTools5.

The second part of the book, containing Chapters 3, 4 and 5, considers each of the three topics – classification, parameter estimation and state estimation – at a theoretical level. Assuming that appropriate models of the objects, physical process or events, and of the sensory system are available, these three tasks are well defined and can be discussed rigorously. This facilitates the development of a mathematical theory for these topics.

The third part of the book, Chapters 6 to 9, discusses all kinds of issues related to the deployment of the theory. As mentioned in Section 1.1, a key issue is modelling. Empirical data should be combined with prior knowledge about the physical process underlying the problem at hand, and about the sensory system used. For classification problems, the empirical data are often