Prentice Hall Information and System Sciences Series

Thomas Kailath, Editor

Anderson & Moore	*Optimal Control: Linear Quadratic Methods*
Anderson & Moore	*Optimal Filtering*
Åström & Wittenmark	*Computer-Controlled Systems: Theory and Design, second edition*
Basseville & Nikiforov	*Detection of Abrupt Changes: Theory and Application*
Boyd & Barratt	*Linear Controller Design: Limits of Performance*
Dickinson	*Systems: Analysis, Design & Computation*
Gardner	*Statistical Spectral Analysis: A Nonprobabilistic Theory*
Goodwin & Sin	*Adaptive Filtering, Prediction, and Control*
Gray & Davisson	*Random Processes: A Mathematical Approach for Engineers*
Grewal & Andrews	*Kalman Filtering: Theory and Practice*
Haykin	*Adaptive Filter Theory*
Haykin, ed.	*Blind Deconvolution*
Jain	*Fundamentals of Digital Image Processing*
Jamshidi, Tarokh, & Shafai	*Computer-Aided Analysis and Design of Linear Control Systems*
Johansson	*System Modeling & Identification*
Johnson	*Lectures on Adaptive Parameter Estimation*
Kailath	*Linear Systems*
Kumar & Varaiya	*Stochastic Systems*
Kung	*Digital Neural Networks*
Kung	*VLSI Array Processors*
Kung, Whitehouse, & Kailath, Eds.	*VLSI and Modern Signal Processing*
Kwakernaak & Sivan	*Signals & Systems*
Landau	*System Identification and Control Design Using P.I.M. + Software*
Ljung	*System Identification: Theory for the User*
Ljung & Glad	*Modeling of Dynamic Systems*
Macovski	*Medical Imaging Systems*
Melsa & Sage	*An Introduction to Probability and Stochastic Processes*
Middleton & Goodwin	*Digital Control & Estimation*
Narendra & Annaswamy	*Stable Adaptive Systems*
Porat	*Digital Processing of Random Signals: Theory & Methods*
Rugh	*Linear System Theory*
Sastry & Bodson	*Adaptive Control: Stability, Convergence, and Robustness*
Soliman & Srinath	*Continuous and Discrete Signals and Systems*
Spilker	*Digital Communications by Satellite*
Williams	*Designing Digital Filters*

Modeling of Dynamic Systems

Lennart Ljung
Torkel Glad

P T R Prentice Hall
Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Ljung, Lennart.
Modeling of dynamic systems / Lennart Ljung, Torkel Glad.
p. cm. -- (Prentice-Hall information and system sciences series)
Includes index.
ISBN 0-13-597097-0
1. Mathematical models. 2. Computer simulation. I. Glad, Torkel. II. Title. III. Series.
QA401.L58 1994
620'.001'185--dc20 94-862
CIP

Editorial/production supervision: *Dit Mosco*
Cover design: *Design Solutions*
Manufacturing manager: *Alexis Heydt*
Acquisitions editor: *Karen Gettman*

©1994 by P T R Prentice Hall
Prentice-Hall, Inc.
A Paramount Communications Company
Englewood Cliffs, New Jersey 07632

The publisher offers discounts on this book when ordered in bulk quantities. For more information, contact:

Corporate Sales Department
PTR Prentice Hall
113 Sylvan Avenue
Englewood Cliffs, NJ 07632

Phone: 201-592-2863
Fax: 201-592-2249

Originally published in Swedish as *Modellbygge och simulering* by Studentlitteratur, Lund, Sweden

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-597097-0

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*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

Contents

I Models 9

1 Systems and Models 13
- 1.1 Systems and Experiments 13
- 1.2 What Is a Model? 14
- 1.3 Models and Simulation 15
- 1.4 How to Build Models 16
- 1.5 How to Verify Models 17
- 1.6 Different Types of Mathematical Models 19
- 1.7 The Book in Summary 21

2 Examples of Models 23
- 2.1 Introduction 23
- 2.2 An Ecological System 23
- 2.3 A Flow System 27
- 2.4 An Economic System 29
- 2.5 Conclusions 32

3 Models for Systems and Signals 33
- 3.1 Types of Models 33
- 3.2 Input, Output, and Disturbance Signals 36
- 3.3 Differential Equations 40
- 3.4 The Concept of State and State-space Models 43
- 3.5 Stationary Solutions, Static Relationships, and Linearization 46
- 3.6 Disturbances in Dynamic Models 53
- 3.7 Description of Signals in the Time Domain 58
- 3.8 Description of Signals in the Frequency Domain 64

3.9 Links between Continuous Time and Discrete Time Models 71
3.10 Appendix 78

II Physical Modeling **79**

4 Principles of Physical Modeling **83**
4.1 The Phases of Modeling 83
4.2 An Example: Modeling the Head Box of a Paper Machine 85
4.3 Phase 1: Structuring the Problem 86
4.4 Phase 2: Setting up the Basic Equations 91
4.5 Phase 3: Forming the State-space Model 95
4.6 Simplified Models 97
4.7 Conclusions 104

5 Some Basic Relationships in Physics **107**
5.1 Introduction 107
5.2 Electrical Circuits 107
5.3 Mechanical Translation 110
5.4 Mechanical Rotation 112
5.5 Flow Systems 114
5.6 Thermal Systems 119
5.7 Some Observations 120
5.8 Conclusions 121

6 Bond Graphs **123**
6.1 Efforts and Flows 123
6.2 Junctions 126
6.3 Simple Bond Graphs 131
6.4 Transformers and Gyrators 135
6.5 Systems with Mixed Physical Variables 138
6.6 Causality: Signals between Subsystems 139
6.7 State Equations from Bond Graphs 148
6.8 Ill-posed Modeling Problems and Bond Graphs 152
6.9 Controlled Elements 154
6.10 Further Remarks 161
6.11 Conclusions 163

6.12 Appendix 164

7 Computer-aided Modeling 169

7.1 Computer Algebra and Its Applications to Modeling . . 170
7.2 Analytical Solutions 172
7.3 Algebraic Modeling 173
7.4 An Automatic Translation of Bond Graphs to Equations 178
7.5 Conclusions 184

III Identification 187

8 Estimating Transient Response, Spectra, and Frequency Functions 191

8.1 Experiments for the Structuring Phase: Transient Analysis 191
8.2 Correlation Analysis 195
8.3 Frequency Analysis 199
8.4 Fourier Analysis 206
8.5 Estimation of Signal Spectra 209
8.6 Estimating Transfer Functions Using Spectral Analysis . 218
8.7 Summary 222
8.8 Appendix 222

9 Parameter Estimation in Dynamic Models 227

9.1 Tailor-made Models 228
9.2 Linear, Ready-made Models 231
9.3 Fitting Parameterized Models to Data 236
9.4 Model Properties 243
9.5 Summary 252
9.6 Appendix 253

10 System Identification as a Tool for Model Building 259

10.1 Program Packages for Identification 260
10.2 Design of Identification Experiments 262
10.3 Posttreatment of Data 269
10.4 Choice of Model Structure 274
10.5 Model Validation 283

10.6 An Example 285
10.7 Conclusions: The Possibilities and Limitations of Identification 289

IV Simulation and Model Use 293

11 Simulation 297
11.1 Short Review 297
11.2 Scaling 298
11.3 Block Diagrams 301
11.4 Connecting Subsystems 306
11.5 Simulation Languages 313
11.6 Numeric Methods 318
11.7 Simulators 327
11.8 Summary 332

12 Model Validation and Model Use 333
12.1 Model Validation 333
12.2 Domain of Validity of the Model 336
12.3 Remaining Critical of the Model 337
12.4 Use of Several Models 338

A Linear Systems. Description and Properties 341
A.1 Time Continuous Systems 341
A.2 Time Discrete Models 343
A.3 Connections between Time Continuous and Time Discrete Models 345

B Linearization 347
B.1 Continuous Time Models 347
B.2 Discrete Time Models 349

C Signal Spectra 351
C.1 Time Continuous Deterministic Signal with Finite Energy 351
C.2 Sampled Deterministic Signals with Finite Energy . . . 352
C.3 Connections between the Continuous and the Sampled Signal 352
C.4 Signals with Infinite Energy 353

C.5 Stochastic Processes 354

Preface

More and more engineering work relies on mathematical models of the studied object. It is thus important to master the art of model construction of real processes. Two kinds of knowledge necessary for model construction can be discerned. One is the actual knowledge and insights of the process's way of functioning and its properties. The other is the knowledge of how these facts can be transferred into a useful model. We can call these areas of knowledge the *domain expert's* and the *knowledge engineer's*, respectively.

This is a book about the knowledge engineer's role in the modeling. The book treats methods of transferring physical facts, more intuitive insights, and information in measured signals into useful mathematical models. It also deals with how to use such models in simulation applications, which play a more and more important role in the contemporary engineer's work.

The material has been used in different courses in the engineering education at Linköping Institute of Technology and Chalmers Institute of Technology in Sweden. It has also been used in several courses for practicing engineers. The reader is assumed to have some background in signals and systems as well as in elementary physics and statistics.

Several persons have helped us with the work in this book. Dr. Krister Gerdin has supplied the material in Example 8.1. Examples 8.3 and 8.4 are based on Professor Karl Johan Åström's report "Frequency Response Analysis, Report 7504," Lund Institute of Technology while the head box example in Chapter 4 is described in his report *Lecture Notes on Paper Machine Control – Head Box Flow Dynamics and Control*, Department of Automatic Control, Lund Institute of Technology, March 1972. He has been kind enough to let us use this material.

Professors Bo Egardt and Gustaf Olsson gave valuable points of view on the manuscript. Almost all our co-workers in the Division of Automatic Control at Linköping Institute of Technology, as well as the students taking our courses, have given valuable advice and viewpoints that have lead to several revisions.

Ulla Salaneck translated the book into English and also produced it in LaTeX. Ingegerd Stenlund also did a substantial part of the preparatory work. Leif Andersson, Patrik Lägermo, and Björn Ottersten solved the last technical problems in transferring the manuscript into a final ps-file. We thank all of them for their help.

Part I

Models

To Describe Reality with Models

Constructing models for a slice of reality and studying their properties is really what science is about. The models – "the hypotheses," "the laws of nature," "the paradigms" – can be of a more or less formal character, but they all have the fundamental property that they try to link observations to some pattern.

In Chapter 1 we will describe the roles that models of dynamical systems play; Chapter 2 gives a number of examples of models from different areas. In Chapter 3 the necessary, formal mathematical background to handle models and systems is given.

Chapter 1

Systems and Models

1.1 Systems and Experiments

The concept of *system* can be defined in several different ways. Here we will use it to denote an object or a collection of objects whose properties we want to study. With such a broad definition most things in our environment will become systems:

Example 1.1
The solar system, a paper machine, an evergreen forest, a capacitor with a resistor, and so on, are all examples of systems. □

It is typical for human activity and curiosity to seek answers to many questions about various system properties.

Example 1.2
For the solar system a typical question is: When will the next solar eclipse occur? For the paper machine one might wonder: How shall I adjust all the valves so that good quality paper is produced? Concerning the capacitor and the resistor the question can be: What will happen if I connect them? □

Many questions of this kind can be answered by *experimentation.* Connect the capacitor and the resistor and observe what happens! A main activity for the natural sciences over several centuries has been to ask appropriate questions about system properties and answer them by experimentation.

The experimental method is based on a sound scientific principle, but it has its limitations. It is sometimes inappropriate or impossible to carry out an experiment. The reason might be one of the following:

- It is too expensive: Arbitrarily testing different valve positions on the paper machine would produce unsellable paper during the tests.
- It is too dangerous: Training nuclear plant operators in how to react in dangerous situations on a real nuclear plant would be inappropriate.
- The system does not (yet) exist: When designing a new airplane one wants to test the effect of different wing shapes on the aerodynamical properties.

Note especially the last point. It represents a very common actual situation. What possibilities do we then have to answer such questions? Well, first we have to make a *model* of the system.

1.2 What Is a Model?

Loosely put, a model of a system is a tool we use to answer questions about the system without having to do an experiment. In this way we use models in everyday life all the time. It is, for example, a model of a person's behavior to say that he is "kind." This model helps us to answer the question of how he will react if we ask him for a favor. We also have models for technical systems that are based on intuition and experience "in the back of our heads." We call such models *mental models.* To learn to drive a car, for example, consists partly of developing a mental model of the car's driving properties. The operator's picture of how an industrial process reacts on different actions is also a mental model developed by training and experience.

Another kind of model is a *verbal model;* the behavior of a system under different conditions is described in words; *If* the bank rate goes up, *then* the unemployment rate will rise. Expert systems are examples of formalized verbal models. It is important to separate verbal and mental models. We use a mental model of the bicycle dynamics when we ride a bike. It is not easy to convert it to a verbal model.

In addition to the mental and verbal models there are models that try to imitate the system. (The word "model" is derived from Latin and originally means mold or pattern.) This can simply mean *physical models*, like the ones architects and boat builders use to test the system's (the house and boat, respectively) esthetic and hydrodynamical properties, respectively.

But the models that we will work with in this book are of a fourth kind: *Mathematical models.* By this we mean that the relationships between quantities (distances, currents, flows, unemployment, and so on) that can be observed in the system are described as mathematical relations in the model. Most laws of nature are mathematical models in this sense.

Example 1.3
For the system "a mass point," Newton's law of motion gives a relationship between force and acceleration.

For the system "a resistor," Ohm's law describes the connection between current and voltage. □

The laws of nature deal, in general, with simple and often ideal systems. For realistic systems the relationships between the variables can be much more complicated.

1.3 Models and Simulation

Assume now that for different reasons the experiment on the system cannot be carried out, but a model of the system is available. The model can then be used to calculate or decide how the system would have reacted. This can be done analytically, that is, by mathematically solving the equations that describe the system and studying the answer. This is the way in which models typically are used, for example, in mechanics and electronics.

With effective computer power, a numerical experiment can be performed on the model. This is called *simulation* (from the Latin *simulare* which means pretend). Simulation is thus an inexpensive and safe way to experiment with the system. However, the value of the simulation results depends completely on the quality of the model of the system.

1.4 How to Build Models

The main part of this book deals with building mathematical models of real systems. There are, in principle, two sources of knowledge for system properties. One is the collected experiences of the experts and the literature in the area in question. Within this lies all the laws of nature, which have been worked out by generations of scientists. The other source is the system itself. Observations of the system and experiments on the system are the basis for all descriptions of its properties.

There are also two types of areas of knowledge for the model construction itself. One is the *domain of expertise*. This is about understanding the application and mastering all the facts that are relevant for this model. The other area is that of the *knowledge engineer*, who has to put the expert's knowledge into practice in a usable and explicit model. These terms are usually used in the construction of expert systems (or *knowledge-based systems*), but are just as important for mathematical model building. This book thus aims at describing the tools the knowledge engineer needs to construct mathematical models from the domain of expertise.

There are two basic and quite different principles for model construction.

Physical Modeling

One principle is to break down the properties of the system to subsystems whose behaviors are known. For technical systems this means that the laws of nature that describe the sub-systems are used in general. What happens when the capacitor and the resistor are connected follows Ohm's law and the relationship between charge and current for a capacitor. For nontechnical systems (economic, sociological, biological, and the like), such well-known laws of nature are usually not available, even for simple subsystems. Then hypotheses have to be introduced or generally recognized relationships have to be used.

Identification

The other basic principle is to use observations from the system in order to fit the model's properties to those of the system. This principle is often used as a complement to the first one. For technical systems the laws of nature are themselves mathematical models, which once were based on observations of small systems. Hence these models according to the first basic principle are also originally based on observations of the system. This is sound, since our models of the system ultimately have to be based on experience.

We illustrate the basic principles in Figure 1.1.

1.5 How to Verify Models

It is not difficult to build models of systems — the difficulty lies in making them good and reliable. For the model to be useful, we have to have confidence in the results and predictions that are inferred from it. Such confidence can be obtained by *verifying* or *validating* the model. In principle, model validation is done by comparing the model's behavior with the system's and evaluating the difference.

All models have a certain *domain of validity.* This may determine how exactly they are able to describe the system's behavior. Certain models are valid only for approximate, qualitative statements ("raising oil prices will lead to a slower growth of the GNP"). Most verbal models are of this kind. Other models can be valid even for more exact, quantitative predictions. The domain of validity here corresponds to the accuracy demands. It may also refer to the values of the system variables and system parameters for which the model is valid. A certain model of a pendulum might only be valid for small angular positions, while another is reliable even for large angles.

It is important to understand that *all models have a limited domain of validity.* One could say that a law of nature is a mathematical model with a large domain of validity. But it is limited. Newton's laws of motion are valid with good precision within a broad spectrum of velocities, but close to the speed of light it is unable to describe the motion of particles.

One lesson of this section is that there is a fundamental limitation in models and simulations: It is hazardous to use a model outside the area

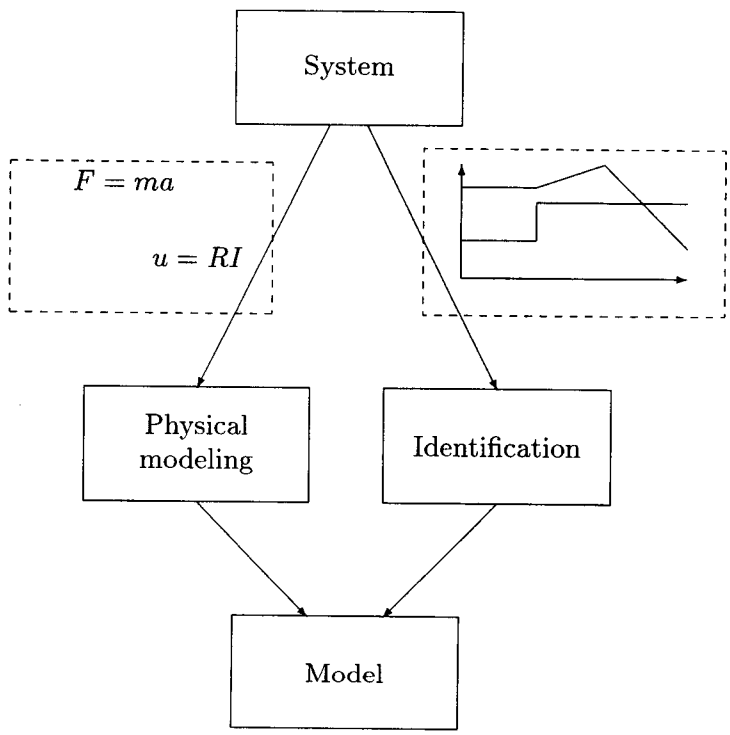


Figure 1.1: Model construction.

it has been validated for. Models and simulations can never replace observations and experiments — but they constitute an important and useful complement.

1.6 Different Types of Mathematical Models

Mathematical models that have been developed for different systems can have different characteristics depending on the properties of the system and on the tools used. A number of adjectives are used to describe the different types.

Deterministic – Stochastic

We call a model *deterministic* if it works with an exact relationship between measurable and derived variables and expresses itself without uncertainty. (It is another thing if our confidence in the expression is limited.) A model is *stochastic* if it also works with uncertainty or probability concepts. A stochastic, mathematical model contains quantities that are described using stochastic variables or stochastic processes.

Dynamic – Static

A system is usually characterized by a number of variables that change with time (position of bodies, current, voltage, unemployment numbers, blood sugar values, and the like). If there are direct, instantaneous links between these variables, the system is termed *static*. A resistor is an example of a static system, since the current through it and the voltage across it are directly related (Ohm's law). The current through it depends only on the present voltage and not on earlier values.

For other systems the variables can change also without direct outside influence, and their values will thereby also depend on earlier applied signals. Such systems are called *dynamic* (from the Greek *dynamis* for strength or power).

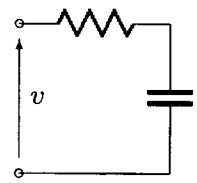


Figure 1.2: Resistor and capacitor.

Example 1.4
A capacitor connected to a resistor with an external voltage v, is a dynamic system. See Figure 1.2. The voltage over the capacitor depends on the charge and thereby on earlier values of the current and voltage v. □

Example 1.5
A national economy is a dynamic system, since the present economic situation depends on several earlier years' socioeconomical interventions. □

The list of examples of dynamic systems can be long. In this book we will use a slightly narrower definition: A dynamic system is a system that is described by differential and/or difference equations.

Continuous Time – Discrete Time

A mathematical model that describes the relationship between time continuous signals is called *time continuous*. Differential equations are often used to describe such a relationship. In practice, the signals of interest are most often obtained in *sampled* form, that is , as a result of discrete time measurements. A model that directly expresses the relationships between the values of the signals at the sampling instants is called a *discrete time* or *sampled* model. Such a model is typically

described by difference equations.

Lumped — Distributed

Many physical phenomena are described mathematically by *partial differential equations.* The events in the system are so to speak dispersed over the space variables. This description is called a *distributed parameter model.* If the events are described by a finite number of changing variables, we talk of *lumped* models. Such models are usually expressed by *ordinary differential equations.*

Change Oriented — Discrete Event Driven

The physical world is usually described in terms of continuous changes in the signals and variables we are interested in. Most laws of nature are of this character. (If we work with discrete time, the changes will of course not be continuous, but the basic idea is the same.) We call such models *change oriented* and say that they correspond to the Newtonian paradigm in the model world.

For systems constructed by humans the course of events is different in many cases. The underlying changes take place in terms of discrete events, which occur more or less randomly. Think, for example, of queuing systems or production systems where the arrival of customers drives the system. Other random events that influence the system may also occur. It can be a question of a machine breaking down, a buffer stock being emptied, and so on. Such systems and models are called *discrete event systems.*

1.7 The Book in Summary

After this survey of concepts that apply to systems and models we can define the book's goals and contents: *To describe how to build mathematical models of dynamic systems and how to use them for simulation.* We will treat both discrete time and continuous time representations and deterministic as well as stochastic models. We will, however, not discuss distributed models. They will have to be treated approximately as lumped (see Example 2). We will also not discuss the area of discrete event models. Such models demand their

own techniques both when they are constructed and simulated and thereby fall outside the framework of this book.

The book is divided into four parts:

Part I, which comprises Chapters 1–3, deals with typical ways of describing signals and dynamic systems and their basic properties.

Part II, which consists of Chapters 4–7, describes tools for physical model building. This includes straightforward methods based on common sense as well as more formalized procedures like bond graphs and computer algebraic tools.

Part III, which comprises Chapters 8–10, deals with system identification, that is, methods to arrive at mathematical models by starting with observations of the system.

Part IV, consisting of Chapters 11–12, describes how models are simulated and how they are used in simulators. The concept of model quality is also discussed.

Chapter 2

Examples of Models

2.1 Introduction

In this chapter we will study a number of simple examples of model building from different areas. The main purpose is to illustrate how to think when constructing the models and which types of mathematical models can be obtained. This will serve as a background when we discuss the formal aspects of the models in Chapter 3. There we will use the examples from this chapter as illustrations.

In these examples we will present the models in a way that seems natural within the respective applications. In Chapters 4–6 we will introduce principles for systemizing model building.

2.2 An Ecological System

As a first example we will study an idealized ecological system consisting of two animal species that either compete for the same food (case 1) or are in a predator–prey situation (case 2). We are interested in variations in the number of individuals of these species. Let $N_1(t)$ and $N_2(t)$ be the number of individuals of each species at time t. The birth rate for the species is assumed to be constants λ_1, and λ_2, respectively. There are thus $\lambda_i N_i$ offsprings of the species i born per unit time.

The mortality rate for the species is μ_i, and it depends on the availability of food as well as the risk of being eaten. In general, we can write $\mu_i = \mu_i(N_1, N_2)$. There are thus $\mu_i(N_1, N_2) \cdot N_i$ individuals

of species i dying per unit time. The net effect is now described by the differential equations

$$\frac{d}{dt}N_1(t) = (\lambda_1 - \mu_1(N_1, N_2))N_1(t) \tag{2.1a}$$

$$\frac{d}{dt}N_2(t) = (\lambda_2 - \mu_2(N_1, N_2))N_2(t) \tag{2.1b}$$

We will examine the two cases.

Case 1: The Species Compete for the Same Food

If the species subsist on the same food, the total number of specimens (possibly weighted) will determine the supply of food and thereby the mortality rate. We assume as a simple model that the mortality is proportional to this total number and assign

$$\mu_i(N_1, N_2) = \gamma_i + \delta_i(N_1 + N_2); \quad \delta_i > 0 \quad i = 1, 2$$

and get the model

$$\frac{d}{dt}N_1(t) = (\lambda_1 - \gamma_1)N_1(t) - \delta_1(N_1(t) + N_2(t))N_1(t) \tag{2.2a}$$

$$\frac{d}{dt}N_2(t) = (\lambda_2 - \gamma_2)N_2(t) - \delta_2(N_1(t) + N_2(t))N_2(t) \tag{2.2b}$$

In Figure 2.1, it is shown how N_1 and N_2 vary when

$$\lambda_1 = 3, \quad \lambda_2 = 2, \quad \gamma_1 = \gamma_2 = \delta_1 = \delta_2 = 1$$

It can be shown that if $(\lambda_1 - \gamma_1)/\delta_1 > (\lambda_2 - \gamma_2)/\delta_2$ the second species will die out and the first will approach the number $(\lambda_1 - \gamma_1)/\delta_1$, independently of the initial number of individuals.

Case 2: Predator and Prey

Assume now that the first species preys on the second. The supply of food for species 1 is thus proportional to N_2, and their mortality rate is thereby diminished when N_2 increases. We assign the simple relationship

$$\mu_1(N_1, N_2) = \gamma_1 - \alpha_1 N_2, \qquad \alpha_1 > 0$$

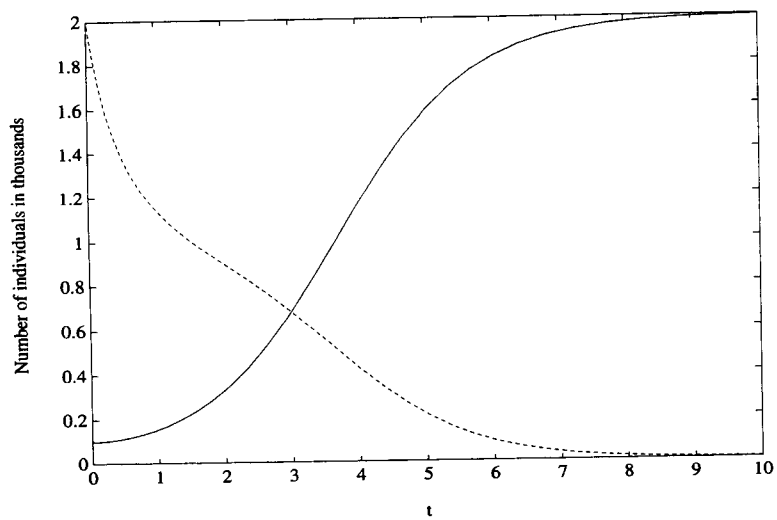


Figure 2.1: Number of individuals when the species compete for the same food. N_1, solid line, N_2, dashed line.

The mortality rate for species 2 increases in the same way when N_1 increases:

$$\mu_2(N_1, N_2) = \gamma_2 + \alpha_2 N_1, \qquad \alpha_2 > 0$$

We then get the model

$$\frac{d}{dt}N_1(t) = (\lambda_1 - \gamma_1)N_1(t) + \alpha_1 N_1(t)N_2(t) \tag{2.3a}$$

$$\frac{d}{dt}N_2(t) = (\lambda_2 - \gamma_2)N_2(t) - \alpha_2 N_1(t)N_2(t) \tag{2.3b}$$

Here it is natural to assume that the predators would die out if there were no prey, that is, $\lambda_1 - \gamma_1 < 0$, and that the prey would multiply if there were no predators, that is, $\lambda_2 - \gamma_2 > 0$. In Figure 2.2, we show how N_1 and N_2 vary in a typical case ($\lambda_1 = 1$, $\gamma_1 = 2$, $\lambda_2 = 2$, $\gamma_2 = 1$, $\alpha_1 = \alpha_2 = 1$). We see that the numbers oscillate around a certain value. This agrees well with the observations of such systems in nature. In Figure 2.3, we show how many hare pelts (snowshoe hare) and lynx pelts that were bought from hunters by the Hudson Bay Company in Canada during the years 1846–1936. It can be assumed

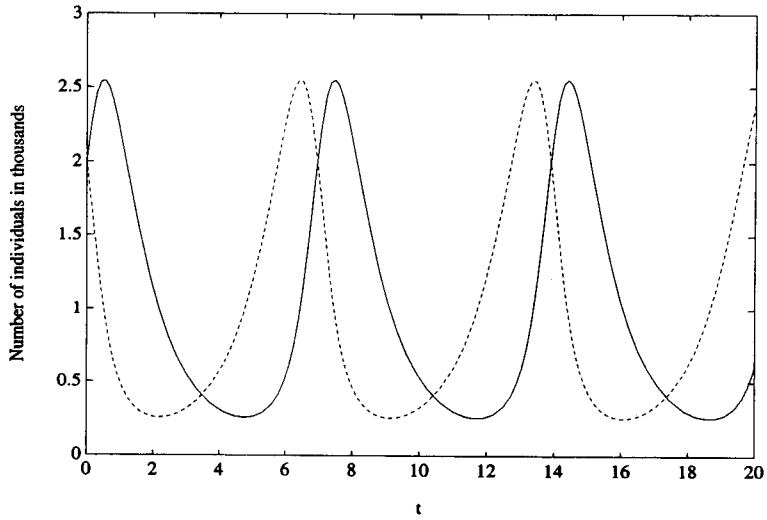


Figure 2.2: Number of predator (N_1) (solid line) and prey (N_2) (dashed line) as a function of time.

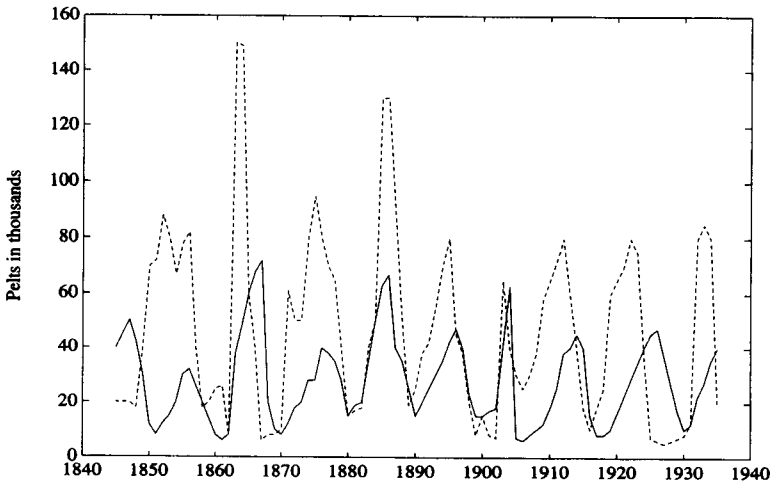


Figure 2.3: Number of lynx pelts (solid line) and hare pelts (dashed line) sold in Canada. (From D. MacLulich, University of Toronto Studies, *Biological Sciences,* No. 43, 1937, pp. 1-136.)

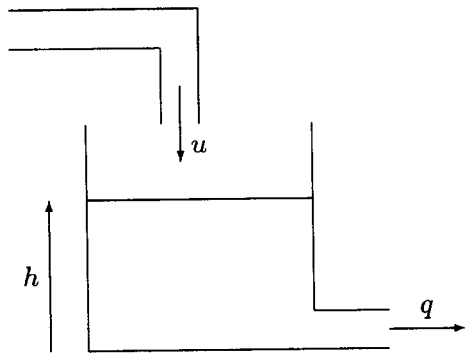


Figure 2.4: Tank with free outflow.

that these numbers are approximately proportional to the occurrence of the respective species.

Remark: This example is based on the classic article by V. Volterra, "Variations and Fluctuations of the Number of Individuals in Animal Species Living Together," *J. du Conseil*, Vol III, 1928. This article was the source of a completely new discipline within biology: population dynamics. The models can of course be made more sophisticated, but already the simple assumptions we have made have some nontrivial consequences. The predator–prey situation, for example, guarantees the survival of both species, while the case where there is competition for food eventually will lead to the extinction of one species.

2.3 A Flow System

Consider a tank with free outflow as in Figure 2.4. The tank has a crossection A (m^2) and the outflow hole has an area a (m^2). The level of the liquid in the tank is h (m), the inflow is u (m^3/s), and the outflow is q (m^3/s). We want to construct a model of how the outflow depends on the inflow.

Bernoulli's law describes the relationship between the outflow speed v (m/s) and the liquid level in the tank:

$$v(t) = \sqrt{2gh(t)} \tag{2.4}$$

Here g is the acceleration of gravity. The relation between the outflow q and the outflow speed v is by definition

$$q(t) = av(t) \tag{2.5}$$

The volume of the liquid in the tank at time t is obviously $A \cdot h(t)$ (m^3), and it changes according to the difference in inflow and outflow (this is called *mass balance given that the density is constant*):

$$\frac{d}{dt} A \cdot h(t) = u(t) - q(t). \tag{2.6}$$

The equations (2.4)–(2.6) now constitute a model for the tank system in Figure 2.4. By substituting (2.5) and (2.4) into (2.6) we get an explicit nonlinear differential equation for the liquid level:

$$\frac{d}{dt} h(t) = -\frac{a\sqrt{2g}}{A} \cdot \sqrt{h(t)} + \frac{1}{A} u(t) \tag{2.7}$$

With the help of (2.7) we can determine the level $h(t)$ when the inflow $u(t)$ is known. After that, the outflow $q(t)$ is determined as

$$q(t) = a\sqrt{2g} \cdot \sqrt{h(t)} \tag{2.8}$$

In Figure 2.5 we show how $h(t)$ varies when $u(t) = 1$, $t \geq 0$, for $h(0) = 0$ and for $h(0) = 2$ ($A = 1$, $a\sqrt{2g} = 1$).

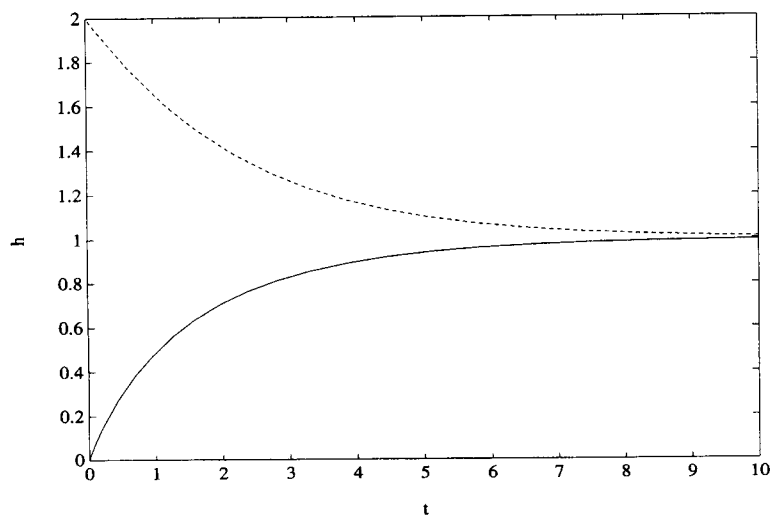


Figure 2.5: Liquid level in the tank when $u(t) = 1$ and $h(0) = 0$ or $h(0) = 2$.

2.4 An Economic System

Simple national economic models are based on the following fundamental variables:

$y(t)$: the gross national product (GNP), year t

$c(t)$: the total consumption, year t

$i(t)$: the total investments, year t

$g(t)$: the expenses of the government, year t

By definition,

$$y(t) = c(t) + i(t) + g(t) \tag{2.9}$$

There are obviously other relationships among these four variables. In reality these are certainly complicated, and some "exact" relationships of the laws of nature type are missing.

Different economic schools have assigned different simplified relationships. In this example we will study a simple Keynesian model, the multiplier–accelerator model according to P. Samuelsson. The following assumptions are then made regarding economic mechanisms:

1. The consumption for the current year is assumed to be proportional to the GNP of the previous year:

$$c(t) = a\, y(t-1); \qquad a > 0 \tag{2.10}$$

2. The investments are assumed to be proportional to the increase in consumption:

$$i(t) = b\,(c(t) - c(t-1)\,)\,; \qquad b > 0 \tag{2.11}$$

Both assumptions seem reasonable and should describe the principal character in the real relationships.

Equations (2.9) – (2.11) now constitute a simple model for the national economic system. Our aim with this model could be to investigate how the government can influence the economy by different interventions. It is then natural to view the GNP $y(t)$ as a goal variable. The government can control it in several ways, for example by influencing $c(t)$ via taxes (increases in sales taxes are used to lower the consumption) or by influencing $i(t)$ via the bank rate (a lower bank rate makes it easier to borrow money and thereby investments increase).

But in this example we will consider the expenditures $g(t)$ as the government's main instrument to influence the economy. To investigate how $g(t)$ influences $y(t)$, it is suitable to reorganize the model given by equations (2.9)–(2.11).

The variables $c(t)$ and $i(t)$ in (2.9) can be eliminated with the help of (2.10) and (2.11), which gives

$$y(t) = ay(t-1) + b\,(ay(t-1) - ay(t-2)) + g(t) \tag{2.12}$$

The relationship between $g(t)$ and $y(t)$ is thus given by the difference equation

$$y(t) - (a + ab)y(t-1) + aby(t-2) = g(t) \tag{2.13}$$

We could also organize the model by expressing how the variables $y(t)$ and $c(t)$ are changed from year to year. According to (2.10) we have

$$c(t+1) = ay(t)$$

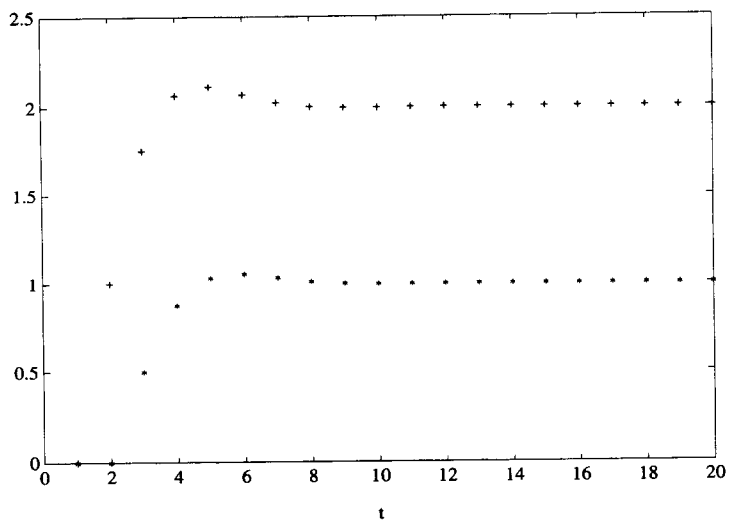


Figure 2.6: Consequences in the GNP, $y(t)$ (crosses), and consumption $c(t)$ (asterisks), of an increase in the state expenditures, $g(t)$.

Furthermore

$$\begin{aligned} y(t+1) &= c(t+1) + i(t+1) + g(t+1) \\ &= c(t+1) + b(c(t+1) - c(t)) + g(t+1) \\ &= (1+b)ay(t) - bc(t) + g(t+1) \end{aligned}$$

Here we have used (2.9) in the first equation, (2.11) in the second, and (2.10) in the third. Using vector and matrix notations, we can write the result as

$$\begin{pmatrix} c(t+1) \\ y(t+1) \end{pmatrix} = \begin{pmatrix} 0 & a \\ -b & (1+b)a \end{pmatrix} \begin{pmatrix} c(t) \\ y(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} g(t+1) \qquad (2.14)$$

The descriptions (2.13) and (2.14) are, of course, equivalent, but the matrix expression (2.14) has certain advantages in simulation, which we will see later. In Figure 2.6 it is shown how the GNP is changed according to model (2.13) or (2.14) when the expenditures of the state increase by one unit at year $t = 2$ ($a = b = 0.5$). We see that an increase in the state expenditures, according to this model, has two consequences for the GNP. One immediate rise follows directly from (2.9). This in turn results in an increase in the consumption, according

to (2.10), and thereby, according to (2.11), an increase in investments. The governments actions thus have a multiplier effect on the economy.

The model (2.14) is of course much simplified. It can be made more detailed and more exact by increasing the number of variables, dividing the investments into different areas, and so on. Such models are very useful today for studying and predicting economic variables.

2.5 Conclusions

The examples in this chapter have been gathered from different problem areas with different characteristics. However, the resulting models have shown a number of common features. The model has consisted of a differential equation or a system of differential equations, describing how some of the system variables relate to each other. It was then possible to determine other variables from the solution to the differential equation. When we worked with discrete time (Section 2.4), we obtained a difference equation instead of a differential equation. But, apart from that, the model had the same structure.

The differential equations have arisen naturally in the model building process. We have started from problems of this type: How does this variable (number of animals, liquid level, GNP) change with time? The dynamic character of the systems (see Section 1.6) therefore naturally leads to differential and difference equations.

We also note that model building contains different degrees of idealization. Equations for certain electric circuits are rather exact. Equations for motion of bodies already contain idealizations of the type of point masses, neglected air resistance, and so on. Bernoulli's law is only valid under idealized conditions, as if no turbulence exists at the outflow, and so on. Even here the approximation is rather good.

For the biological and economic example, however, it is obvious that the assumptions of the mortality rate and the economic links cannot be even approximately valid. They are, on the other hand, qualitatively of the same character as the conditions that, in all likelihood, should apply in reality. We thereby expect that the behavior of the models reveals a number of essential properties of the system. The modeling is valuable even in this case, since the model's statements are not trivial consequences of the assumptions.

Chapter 3

Models for Systems and Signals

3.1 Types of Models

Mathematical Models and Signal Models

By a mathematical model we mean a description of the system where relationships between the model's variables and signals are expressed in mathematical terms. In the examples in Chapter 2 we saw that model building naturally leads to differential and difference equations. Our mathematical models of the dynamic system will thus, in principle, consist of a collection of differential and/or difference equations. In this chapter we will discuss the formal, mathematical aspects of such equations. Several of the external signals that influence the system also have to be modeled in order to understand and simulate their effects on the system. In this chapter we will therefore also discuss typical ways of describing properties of signals.

Block Diagram Models

A block diagram of a system is a logical decomposition of the functions of the system and shows how the different parts (blocks) influence each other. This interaction is illustrated by arrows between the blocks. A given system can usually be represented by several different block diagram models, depending on how detailed we want to make them.

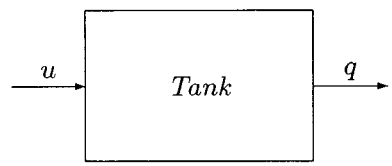


Figure 3.1: Block diagram for the tank in Section 2.3.

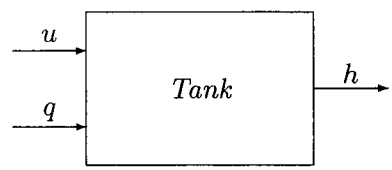


Figure 3.2: Block diagram describing how the level in the tank depends on the inflow and the outflow.

Example 3.1 Block Diagram for the Water Tank
Consider the tank in Figure 2.4. The outflow q depends on the inflow u, which we can illustrate by a simple block diagram, according to Figure 3.1. We can also make a more detailed description, which in addition contains the level h in the tank. The level h depends on the inflow u and the outflow q (Figure 3.2). The outflow q in turn depends on the level h (and the outflow area a), according to Bernoulli's law (see Figure 3.3). The left picture in Figure 3.3 is preferable if the outflow area is fixed and cannot be influenced. If the outflow area can be varied, for example, by placing a valve in the outflow, the right picture is however more natural. From the subsystems in Figures 3.2 and 3.3, we now have a block diagram for the tank according to Figure 3.4.

□

Observe the difference between the schematic picture in Figure 2.4 and the block diagram in Figure 3.4. Schematic pictures are often used

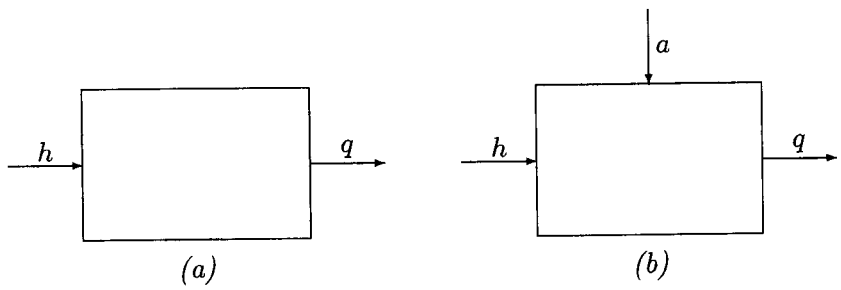


Figure 3.3: (a) Outflow as a function of the level, and (b) as a function of the level and the outflow area.

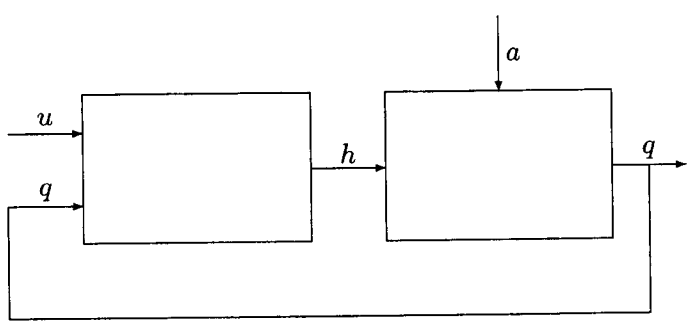


Figure 3.4: Block diagram for the tank system.

for simple illustration of the function of a system. These are, however, based on the *physical* construction of the system, whereas the block diagram is based on the *logical* description. The flows in Figure 3.4 are *information flows* and not water as in Figure 2.4.

Block diagrams are very useful when structuring a system, especially for larger and more complex ones. As models for the system they can be compared to the verbal models we discussed in the introductory chapter. They also constitute a very important starting point for the mathematical model building. When, for example, we built the model (2.7), (2.8) for the tank system, we started from the basic equations (2.6) and (2.4)-(2.5), which correspond to the two blocks in Figure 3.4. In Chapter 4 we will discuss in more detail the use of block diagrams in model building.

Block diagrams are also used as models in sciences where quantitative models usually cannot be constructed, as in, for example, ecology, sociology (sociograms), and so on.

Simulation Models

Models can be constructed for different purposes. As we noted in the introduction, simulation is often a main goal. For large, complex models it is common that the equations have not been explicitly expressed in closed form. The model might then only exist as a computer program that is used for simulation. Such models can be called *simulation models.* In Chapter 11 we will discuss them more closely in connection with program languages that have been specifically developed for simulation of dynamic systems.

3.2 Input, Output, and Disturbance Signals

A mathematical model of a dynamic system contains a number of quantities of different types. In this section we are going to discuss the characteristics of these different quantities and assign some terms for them. Certain quantities in the model do not vary in time. We will call them *constants.* Quantities that vary in time we will call *variables* or *signals.*

When modeling and simulation studies are made for design purposes, it is practical to separate them into two types of constants in

models. *System parameters* are constants that are considered given by the system and cannot be chosen by the designer. *Design parameters* are constants that can be chosen in order to give the system/model desired properties. The purpose of the simulation study is often to decide suitable values for the design parameters.

Example 3.2
If we simulate the tank model in Section 2.3 in order to test how the outflow depends on the outflow area a in an otherwise fixed tank, the area a is a design parameter, while g and A are system parameters. □

A model and a dynamic system always contain a number of variables, or signals, whose behavior is our primary interest. We will call such signals *outputs* and denote them by $y_1(t), y_2(t), \ldots, y_p(t)$.

Example 3.3
In Section 2.2 the output was

$$\begin{aligned} y_1(t) &= N_1(t) \quad \textit{(specimens of species 1)} \\ y_2(t) &= N_2(t) \quad \textit{(specimens of species 2)} \end{aligned}$$

□

Note that the output is not defined by the system itself. It is instead the model builder's interests that decide what is going to be condidered as output. Another model builder might have chosen $y_2(t) = N_2(t)$ as the model's output in Example 3.3.

We will write all outputs as a column vector:

$$y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_p(t) \end{pmatrix} \tag{3.1}$$

In systems/models, there are usually signals and variables that influence other variables in the system, but are not themselves influenced by the behavior of the system. The inflow u in the tank system (Section 2.3) is such a signal. It influences both the level of the tank and the outflow, but does not itself depend on these variables. We will call such a signal an *external signal*. In a block diagram it is easy to

recognize external signals as free arrows pointing into one or several blocks. See for example Figure 3.4, where u and a are external signals.

An external signal can be one of two types. If we have the external signal at our disposal, to influence the system's behavior, we talk of an *input* or *control signal.* We will denote such signals by

$$u_1(t),\ u_2(t), \ldots, u_m(t)$$

or by vector formalism

$$u(t) = \begin{pmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{pmatrix} \tag{3.2}$$

An external signal that we, in the application in question, cannot influence or choose, we will call a *disturbance signal.* We will use the notation

$$w_1(t),\ w_2(t), \ldots, w_r(t)$$

or

$$w(t) = \begin{pmatrix} w_1(t) \\ w_2(t) \\ \vdots \\ w_r(t) \end{pmatrix} \tag{3.3}$$

for disturbance signals.

Example 3.4 Signals for the Water Tank
If the outflow area a in the tank system in Section 2.3 can be varied, this system will have two external signals, $u(t)$ and $a(t)$. Whether they are disturbance signals or inputs depends on the application. The flow $u(t)$ could be a variable that we cannot influence, while $a(t)$ could be regulated to achieve certain goals. Think, for example, of the tank as a water reservoir, $u(t)$ as rain, and $a(t)$ as a floodgate. Then $u(t)$ is a disturbance signal and $a(t)$ an input. In other applications we can control the flow $u(t)$ and then this also will be an input. □

The example shows that the presence of external signals and the division of input and disturbance signals is not unambiguously determined by the system as such. It is instead decided by our opinion of what can vary or be varied and whether we can control the conditions.

For modeling and simulations, it is not necessary to decide on whether a certain signal is an input or a disturbance signal. The signal enters the model and the simulation program in the same way regardless of the interpretation. The distinction will only become important when discussing which properties can be obtained from the system and how to achieve them. Therefore, for simplicity we will often use the notation u for both input and disturbance signals and talk about "inputs" when we should say "inputs and/or disturbance signals."

We have now defined outputs and external signals in the models. We will call other model variables *internal variables.*

The notation we have introduced in this section can be summarized as follows:

- *Constant*: A quantity in the model that does not vary with time.
- *System parameter:* A constant that is given by the system.
- *Design parameter:* A constant that we can vary in order to give the system different properties.
- *Variable* or *signal:* A quantity in the model that varies with time.
- *Output:* A variable whose behavior is our primary interest. Denoted by y.
- *External signal:* A variable that affects the system without being affected by the system's other variables.
- *Input:* An external signal in the system whose time variations we can choose. Denoted by u.
- *Disturbance signal:* An external signal in the system that we cannot influence. Denoted by w.
- *Internal variable*: A variable in the system that is neither an output nor an external signal.

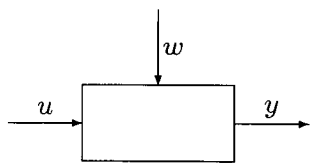


Figure 3.5: Basic block diagram for a system.

With the notation u, w, and y we can picture the system as a simple block diagram according to Figure 3.5.

By these concepts we can also more clearly define the difference between a *static* and a *dynamic* system, which we talked about in Section 1.6. The variations in the output from a static system are directly coupled to the momentary value of the input. For dynamic systems, on the other hand, the present output value depends, in principle, on all earlier input values. See Figure 3.6.

3.3 Differential Equations

In the mathematical modeling in Chapter 2, we found that the relationships between the model variables were described with the help of differential equations (in discrete time, difference equations).

There are two different ways of describing these differential equations. One is to directly relate inputs u to outputs y in one differential equation. In principle it looks like this:

$$g(y^{(n)}(t), y^{(n-1)}(t), \ldots, y(t), u^{(m)}(t), u^{(m-1)}(t), \ldots, u(t)) = 0 \quad (3.4)$$

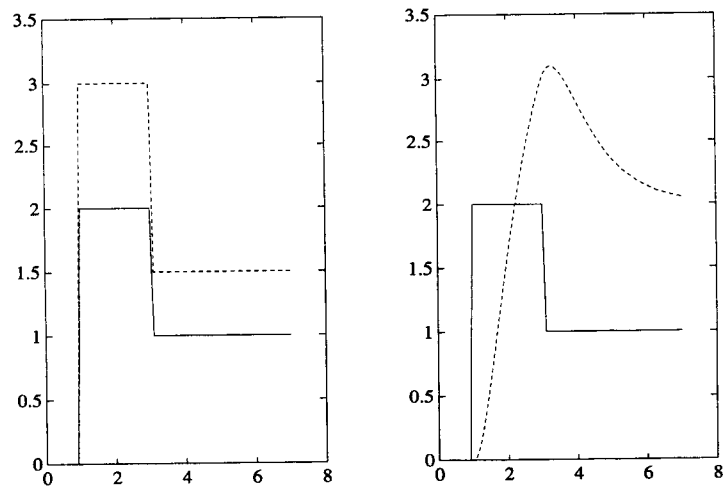
where

$$y^{(k)}(t) = \frac{d^k}{dt^k} y(t)$$

and $g(\cdot, \cdot, \ldots, \cdot)$ is an arbitrary, vector-valued, nonlinear function.

The other way is to write the differential equation as a system of first-order differential equations by introducing a number of internal variables. If we denote these internal variables by

$$x_1(t), \ldots, x_n(t)$$



(a) (b)

Figure 3.6: Example of input-output relationship for (a) a static and (b) a dynamic system. The input is the solid line, and the output the dashed line.

and introduce the vector notation

$$x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} \tag{3.5}$$

we can, in principle, write a system of first-order differential equations as

$$\dot{x}(t) = f(x(t), u(t)) \tag{3.6}$$

The dot over x denotes differentiation with respect to time t. In (3.6), $f(x, u)$ is a vector function with n components:

$$f(x, u) = \begin{pmatrix} f_1(x, u) \\ \vdots \\ f_n(x, u) \end{pmatrix} \tag{3.7}$$

The functions $f_i(x, u)$ are in turn functions of $n + m$ variables, the components of the x and u vectors. Without vector notation, (3.6)

becomes

$$\begin{aligned}
\dot{x}_1(t) &= f_1(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t)) \\
\dot{x}_2(t) &= f_2(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t)) \\
&\vdots \\
\dot{x}_n(t) &= f_n(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t))
\end{aligned} \tag{3.8}$$

The outputs of the model can then be calculated from the internal variables $x_i(t)$ and the inputs $u_i(t)$:

$$y(t) = h(x(t), u(t)) \tag{3.9}$$

which written in longhand means

$$\begin{aligned}
y_1(t) &= h_1(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t)) \\
y_2(t) &= h_2(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t)) \\
&\vdots \\
y_p(t) &= h_p(x_1(t), \ldots, x_n(t), u_1(t), \ldots, u_m(t))
\end{aligned} \tag{3.10}$$

All the models in Chapter 2 can be written as (3.6)-(3.9) or as their corresponding discrete time equations

$$x(t+1) = f(x(t), u(t)) \tag{3.11a}$$

$$y(t) = h(x(t), u(t)) \tag{3.11b}$$

Example 3.5 Internal Description of the Water Tank
The model (2.7), (2.8) for the tank in Section 2.3 is of the form (3.6), (3.9) with

$$\begin{aligned}
&x(t) = h(t),\ u(t) = u(t) \\
&y(t) = q(t),\ n = 1,\ m = 1,\ p = 1
\end{aligned}$$

$$f(x, u) = -\frac{a\sqrt{2g}}{A} \cdot \sqrt{x} + \frac{1}{A}u$$

$$h(x, u) = a\sqrt{2g}\sqrt{x}$$

□

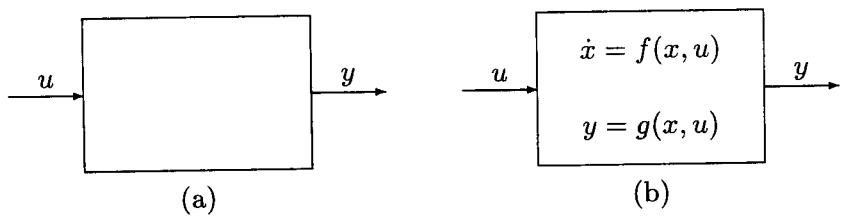


Figure 3.7: (a) External model, and (b) internal model.

A model description of the type (3.4) is sometimes said to be an external description, since it directly relates the external variables to the output. The description (3.6), (3.9) is then said to be internal, since it also describes the behavior of the internal variables, the x's. See Figure 3.7.

In this book we will use the internal description most of the time. The reason for this is that the vector $x(t)$ in (3.6) has an important interpretation as a state vector, which we will discuss in the next section.

3.4 The Concept of State and State-space Models

In the end of Section 3.2 we remarked that for a dynamic system the output depends on all earlier input values. This leads to the fact that it is not enough to know $u(t)$ for $t \geq t_0$ in order to be able to calculate the output $y(t)$ for $t \geq t_0$. We need information about the system. By *the state of the system at time* t_0 we mean an amount of information such that with this state and the knowledge of $u(t)$, $t \geq t_0$, we can calculate $y(t)$, $t \geq t_0$. This definition is well in line with the everyday meaning of the word "state."

It is also obvious from the definition of state that this concept will play a major role in the simulation of the model. The state is exactly the information that has to be stored and updated at the simulation

in order to be able to calculate the output.

Consider a general system of first-order differential equations (3.7) with the output given by (3.9):

$$\dot{x}(t) = f(x(t), u(t)) \tag{3.12a}$$

$$y(t) = h(x(t), u(t)) \tag{3.12b}$$

For this system the vector $x(t_0)$ is a state at time t_0. This follows from a general result on differential equations:

Provided that $f(x, u)$ is well behaved (it is enough, for example, that f is continuously differentiable and u is piecewise continuous), the differential equation (3.12a) with $x(t_0) = x_0$ has a unique solution for $t \geq t_0$.

Intuitively we can think as follows: Assume that we know $x(t)$ and $u(t)$ at time t_0. We can then according to (3.12a) calculate $\dot{x}(t)$. We can then also compute $x(t_0 + \delta t)$ for infinitesimally small δt according to

$$x(t_0 + \delta t) = x(t_0) + \delta t \cdot f(x(t_0), u(t_0)) \tag{3.13}$$

From this value we can continue and calculate $x(t)$ for $t > t_0$. The output $y(t)$, $t \geq t_0$, can then also be computed according to (3.12b). In fact, the equation (3.13) is Euler's method for a numerical solution of (3.12a) if δt is a small, finite number.

We have thus established that the variables $x_1(t), \ldots, x_n(t)$ or, in other words, the vector

$$x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}$$

in the internal model description (3.12) is a state for the model. Herein lies the importance of this model description for simulation. The model (3.12) is therefore called a *state-space model*, the vector $x(t)$ the *state vector*, and its components $x_i(t)$ *state variables*. The dimension of $x(t)$, that is, n, is called the *model order*.

For the discrete time model (3.11a) it is obvious that $x(t_0)$ is a state at time t_0. If we know $x(t_0)$ and $u(t)$ for $t \geq t_0$ we can clearly calculate

$x(t)$ and thereby $y(t)$ for $t = t_0 + 1,\ t_0 + 2,\ t_0 + 3, \ldots$. Equation (3.11a) is its own solution algorithm.

State-space models will be our standard model for dynamic systems. In conclusion, we have the following model:

State space models (continuous time)

$$\dot{x}(t) = f(x(t),\ u(t)) \tag{3.14a}$$

$$y(t) = h(x(t),\ u(t)) \tag{3.14b}$$

$u(t)$: input, an m-dimensional column vector
$y(t)$: output, a p-dimensional column vector
$x(t)$: state, an n-dimensional column vector
The model is said to be nth order. If the function $f(x, u)$ is continuously differentiable and if $u(t)$ is a piecewise continuous function, then a unique solution to (3.14) for $x(t_0) = x_0$ exists.

For discrete time systems we have the corresponding model:

State space models (discrete time)

$$x(t_{k+1}) = f(x(t_k), u(t_k)) \qquad k = 0, 1, 2, \ldots \tag{3.15a}$$

$$y(t_k) = h(x(t_k), u(t_k)) \tag{3.15b}$$

$u(t_k)$: input at time t_k, an m-dimensional column vector
$y(t_k)$: output at time t_k, a p-dimensional column vector
$x(t_k)$: state at time t_k, an n-dimensional column vector
The model is said to be nth order. For a given initial value $x(t_0) = x_0$, (3.15) always has a unique solution.

Linear Models

The model (3.14) or (3.15) is said to be *linear* if $f(x, u)$ and $h(x, u)$ are linear functions of x and u:

$$f(x, u) = Ax + Bu \tag{3.16a}$$

$$h(x, u) = Cx + Du \tag{3.16b}$$

Here the matrices have the following dimensions

$$A : n \times n \quad B : n \times m$$
$$C : p \times n \quad D : p \times m$$

If these matrixes are independent of time the model (3.16) is said to be *linear and time invariant.* Some facts for linear, time invariant models are summarized in Appendix A.

3.5 Stationary Solutions, Static Relationships, and Linearization

For a given initial state x_0 at time t_0,

$$x(t_0) = x_0$$

and for a given time function $u(t)$, a unique solution $x(t)$ exists to (3.14). This solution can, more completely, be denoted by

$$x(t; x_0, t_0, u(\cdot)) \tag{3.17}$$

in order to clearly show that it is tied to the initial value $x(t_0) = x_0$ and to the input $u(\cdot)$. A solution $x(t; x_0, t_0, u(\cdot))$ is also called a *trajectory* for the differential equation (3.14) [or difference equation (3.15)]. The output that corresponds to the input signal $\{u(t)\}$ and to the initial state $x(t_0) = x_0$ is of course

$$y(t) = h(x(t; x_0, t_0, u(\cdot)), u(t))$$

Stationary Solutions

In many cases in practice the input is constant over long periods of time. In this section we will discuss how the state model (3.14) behaves in such a case. We thus assume that the input is constant in time:

$$u(t) \equiv u_0 \tag{3.18}$$

For a given value of u_0, let the state vector x_0 be a solution to the equation

$$f(x_0, u_0) = 0 \tag{3.19}$$

There can be several x_0s that solve (3.19), or such a solution can be lacking, depending on u_0 and the function $f(x, u)$.

Now solve the differential equation

$$\dot{x}(t) = f(x(t),\ u_0)$$

with the initial condition

$$x(t_0) = x_0$$

where x_0 is subject to (3.19). Then $\dot{x}(t_0) = 0$ and the solution $x(t; x_0, t_0, u_0)$ will be constant and equal to x_0. Such a solution is called a *stationary solution.* $\{x_0, u_0\}$ is said to be a *stationary point* to the differential equation (3.14). Sometimes we use the notation *singular point* or *equilibrium.* On the other hand, all time constant solutions $x(t) \equiv x^*$ for the input (3.18) must fulfill

$$0 = \dot{x}(t) = f(x(t), u_0) = f(x^*, u_0)$$

That is, they have to be such that x^* is a solution to (3.19).

All stationary solutions to (3.14) for the input $u(t) = u_0$ are obtained by solving (3.19) with respect to x_0. If $x(t) = x_0$ is a stationary solution, then the corresponding output will also be time invariant:

$$y(t) \equiv y_0 = h(x_0, u_0) \quad \text{for all } t \tag{3.20}$$

The stationary solutions play an important role in the analysis of a system. They are the possible equilibrium points of the system. Such a point often corresponds to a desired behavior of the system.

Example 3.6 Stationary Points for the Population Model
What stationary points does the population model (2.3) have? We find them by setting the right side [the function $f(x)$] equal to zero:

$$\begin{aligned}(\lambda_1 - \gamma_1)N_1 + \alpha_1 N_1 N_2 &= 0\\ (\lambda_2 - \gamma_2)N_2 - \alpha_2 N_1 N_2 &= 0\end{aligned} \tag{3.21}$$

This system of equations has the solution

$$N_1 = 0, \quad N_2 = 0$$

or

$$N_1^* = \frac{\lambda_2 - \gamma_2}{\alpha_2}, \qquad N_2^* = \frac{\gamma_1 - \lambda_1}{\alpha_1} \tag{3.22}$$

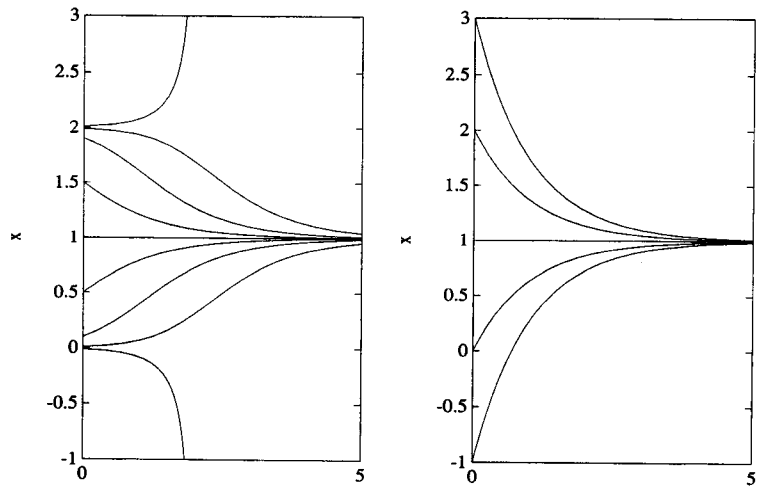


Figure 3.8: $x = 1$ is an asymptotically stable solution. (a) Solutions that start close to 1 converge to this value. (b) All solutions converge, that is, $x = 1$ is globally asymptotically stable.

There are thus two stationary points for this system. This means, for example, that if the number of specimens of each kind should be given by (3.22) at any time, the number will be constant at these values. All other combinations of numbers (except the trivial $N_1 = N_2 = 0$) will necessitate changes. □

Stability

Suppose the initial value $x(t_0) = x_0$ gives a stationary solution. What happens at another initial value $x(t_0) = x_1$? Depending on the properties of the function $f(x, u_0)$, the solution for the initial value x_1 can behave in different ways. The stationary solution x_0 is *asymptotically stable* if any solution $x(t)$ that starts close enough to x_0 converges to x_0 as t tends to infinity. The solution is *globally asymptotically* stable if *all* solutions $x(t)$ to (3.14) with $u(t) = u_0$ converges toward x_0 as t tends to infinity. See Figure 3.8.

Static Relationships

For an input u_0 and a corresponding asymptotically stable stationary solution x_0, the output of the system

$$y(t) = h(x(t), u_0)$$

will converge toward the stationary value

$$y_0 = h(x_0, u_0)$$

Let us discuss the relationship between input level u_0 and the corresponding output y_0. The stationary point x_0, which is determined by (3.19), is an implicit function of u_0. We can emphasize this by writing

$$x_0 = x_0(u_0) \tag{3.23}$$

The stationary output y_0 will thereby be a function of u_0:

$$y_0 = h(x_0(u_0), u_0) = g(u_0) \tag{3.24}$$

The expression (3.24) describes the *static relationship* that exists between a constant input and the corresponding stationary output. If the stationary solution is asymptotically stable and the input is changed from one level to another, the stationary output will in time assume the value in (3.24). How fast this will happen depends on properties of the differential equation (3.15). The term *time constant* is used to tell in which time scale the output is approaching the stationary value y_0. If we say that "the system (for the input u_0) has the time constant 5 seconds," this means that 5 seconds after a change in the input level the output will have come close to its new stationary value. A formal and more exact definition of the concept of time constant can only be given for linear systems. (Compare Appendix A.)

If the time constants for the system (3.14) should be considerably shorter than the time frame we are interested in, the dynamical differential equation model (3.14) can be replaced by the static model (3.24). See also the discussion on separation of time constants in Section 4.6.

If we make a small change of the input level u_0 from u_0 to $u_1 = u_0 + \delta u_0$, the stationary output will according to (3.24) be given by

$$y_1 = g(u_1) = g(u_0 + \delta u_0) \approx g(u_0) + g'(u_0)\delta u_0 = y_0 + g'(u_0)\delta u_0 \tag{3.25}$$

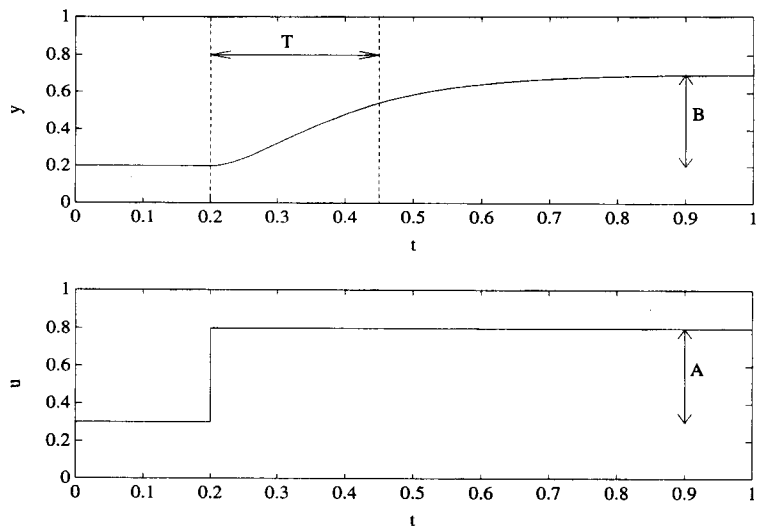


Figure 3.9: Time constant T and static gain B/A.

Here the derivative $g'(u_0)$ is a $p \times m$ matrix that describes how the stationary output varies locally with the input. It is called the *static gain* (for the input u_0). The terms time constant and static gain are illustrated in Figure 3.9.

Example 3.7 Stationary Points for the Water Tank
Consider the tank model in Section 2.3. If the inflow $u(t)$ is constant and equal to u_0, we can determine the corresponding stationary solution for the system by solving [see (2.7)]

$$-\frac{a\sqrt{2g}}{A}\sqrt{h_0} + \frac{1}{A}u_0 = 0$$

with respect to h_0. This gives

$$h_0 = \frac{1}{a^2 \cdot 2g} \cdot u_0^2$$

This relation between the state vector's stationary value h_0 and the input u_0 corresponds to (3.23). Thus the corresponding stationary outflow will be

$$q(t) \equiv q_0 = a \cdot \sqrt{2g} \cdot \sqrt{h_0} = u_0$$

This corresponds to the static relation (3.24). The static gain is thus equal to 1, *independent of the value of* u_0. *(This result is trivial, since if the level in the tank is constant the outflow has to be equal to the inflow.)* □

In this section we have discussed continuous time systems. For a discrete time system (3.15), a stationary solution corresponding to the constant input

$$u(t_k) = u_0$$

is given by x_0, the solution of

$$x_0 = f(x_0, u_0)$$

This equation corresponds to (3.19) in the continuous time case. Other than that, the results are analogous to the ones discussed previously.

Linearization

If a stationary solution is of interest, it can be meaningful to ask how the solution behaves in its neighborhood.

Consider a nonlinear system (3.14) with a stationary solution (x_0, u_0). Let $y_0 = h(x_0, u_0)$. This system can be linearized around the solution x_0, u_0 by considering small deviations $\Delta x(t) = x(t) - x_0$, $\Delta u(t) = u(t) - u_0$, and $\Delta y(t) = y(t) - y_0$. Approximately, we then have

$$\dot{\Delta x} = A\Delta x + B\Delta u \tag{3.26a}$$

$$\Delta y = C\Delta x + D\Delta u \tag{3.26b}$$

Here A, B, C, and D are partial derivative matrices (Jacobians) of $f(x, u)$ and $h(x, u)$, respectively, evaluated at (x_0, u_0). See Appendix B for more details.

Example 3.8 Linearization of the Population Model
According to Example 3.6, the population model in Section 2.2 has the stationary points $N_1 = N_2 = 0$ *and*

$$N_1^* = \frac{\lambda_2 - \gamma_2}{\alpha_2}, \qquad N_2^* = \frac{\gamma_1 - \lambda_1}{\alpha_1}$$

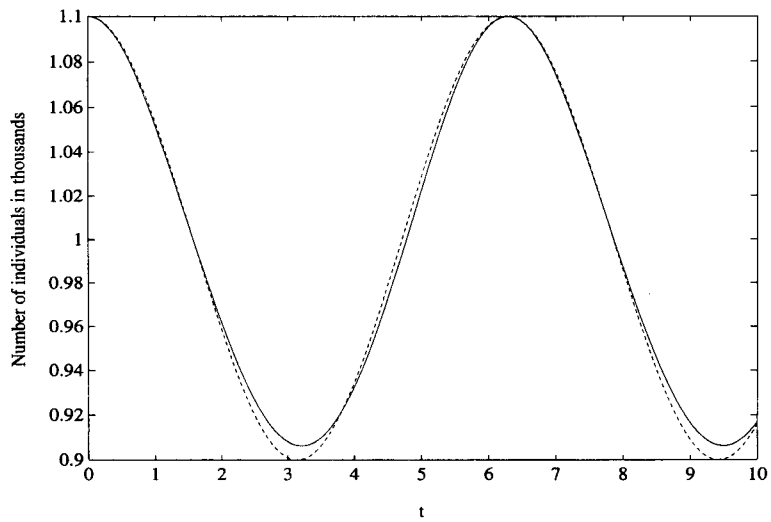


Figure 3.10: The solution to the population model. Exact (solid line) and linearized (dashed line) for $\Delta N_1(0) = 0.1$, $\Delta N_2(0) = 0$.

Let us linearize the equation (2.3) around the latter stationary point. We have

$$\begin{aligned}
\tfrac{d}{dN_1}\left((\lambda_1 - \gamma_1)N_1 + \alpha_1 N_1 N_2\right) &= \lambda_1 - \gamma_1 + \alpha_1 N_2 \\
\tfrac{d}{dN_2}\left((\lambda_1 - \gamma_1)N_1 + \alpha_1 N_1 N_2\right) &= \alpha_1 N_1 \\
\tfrac{d}{dN_1}\left((\lambda_2 - \gamma_2)N_2 - \alpha_2 N_1 N_2\right) &= -\alpha_2 N_2 \\
\tfrac{d}{dN_2}\left((\lambda_2 - \gamma_2)N_2 - \alpha_2 N_1 N_2\right) &= (\lambda_2 - \gamma_2) - \alpha_2 N_1
\end{aligned}$$

Evaluating these derivatives at the stationary point N_1^*, N_2^*, we have the linearized system:

$$\frac{d}{dt}\begin{pmatrix}\Delta N_1(t)\\ \Delta N_2(t)\end{pmatrix} = \begin{pmatrix} 0 & \frac{\alpha_1}{\alpha_2}(\lambda_2 - \gamma_2) \\ -\frac{\alpha_2}{\alpha_1}(\gamma_1 - \lambda_1) & 0 \end{pmatrix}\begin{pmatrix}\Delta N_1(t)\\ \Delta N_2(t)\end{pmatrix} \tag{3.27}$$

Figure 3.10 shows the solutions to the linearized equation (3.27) and to the original one (2.3) for the initial value $\Delta N_1(0) = 0.1$, $\Delta N_2(0) = 0$. Note the insignificant deviation from the equilibrium. □

Linearization is an important and useful tool. However, we have to point out the following important limitations:

- Linearization can only be used to study local properties in the vicinity of a stationary solution. This can be interesting in many cases, in particular if we want the system to remain around one of these stationary solutions.

- It is often difficult to quantitatively estimate how good the approximation of the linearized solution is. Conclusions based on the linearized system should be treated with a considerable amount of caution and preferably be complemented by simulations of the original nonlinear system.

Finally, the system (3.14) can be linearized by a similar procedure around any solution not necessarily stationary. The resulting linear system will then, in general, be time varying. [The matrices A, B, C, and D in (3.26) will be evaluated at $\bar{x}(t)$, $\bar{u}(t)$, where $\bar{x}(t)$, $\bar{u}(t)$ is the nominal solution to (3.14) around which we linearize.]

Remark: A time discrete system (3.15) can also be linearized around a stationary solution u_0, x_0. We then have

$$x(t_{k+1}) - x_0 = A(x(t_k) - x_0) + B(u(t_k) - u_0) \tag{3.28a}$$

$$y(t_k) - y_0 = C(x(t_k) - x_0) + D(u(t_k) - u_0) \tag{3.28b}$$

where the matrices A, B, C, and D are given by (B.2) and (B.5) in Appendix B.

3.6 Disturbances in Dynamic Models

In Section 3.2 we defined a *disturbance signal* as an external signal that we cannot choose or influence ourselves. In many systems the disturbance signal has a very obvious influence on the behavior of the system. It is then important to also have a picture of the typical properties of the disturbance signals.

The construction of a model for the disturbance signals depends to some extent on whether it is separately measurable and of known origin. In this section we will discuss the importance of these distinctions.

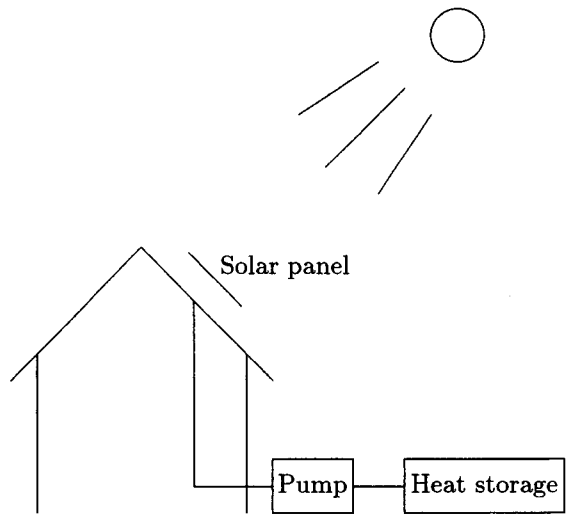


Figure 3.11: Outline of the solar house

Known Disturbance Sources — Measurable Disturbance Signals

The disturbance signal is often a well-known physical quantity, which can be measured separately. Consider the following example.

Example 3.9 A Solar-heated House
Basically a solar-heated house looks like the one depicted in Figure 3.11. The function is that the sun heats the air in the solar panel. This air is then pumped by fans into the heat storage. We use the following signals to describe the system:

$I(t)$: *Solar intensity at time t*

$y(t)$: *Temperature at the inlet to the heat storage*

$u(t)$: *Pump velocity*

We want to build a model for how the storage temperature $y(t)$ is influenced by the pump speed $u(t)$ and the intensity $I(t)$. Measurements of the solar intensity $I(t)$ are shown in Figure 3.12. (See also Example

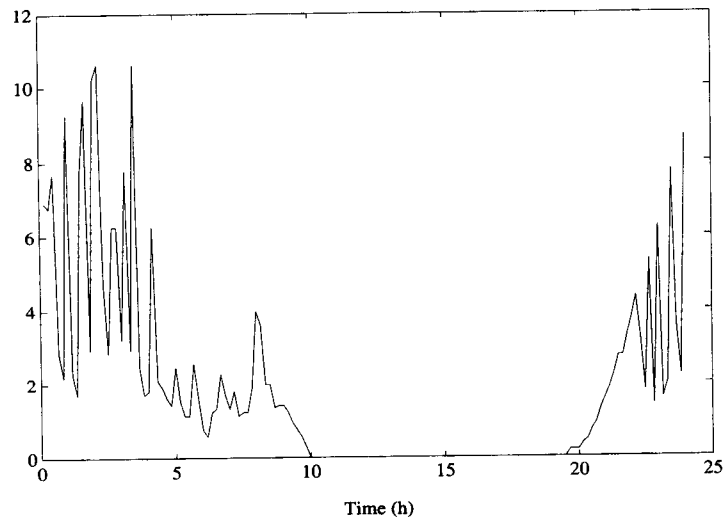


Figure 3.12: The solar intensity $I(t)$ measured during a 24-hour period with a sampling interval of 10 minutes.

10.4.) (Actually, it is more common to use water instead of air for the heat transport, but the data shown in the figure are collected from an experimental setup with air.) □

According to the classification in Section 3.2, the pump speed $u(t)$ is an *input*, while the intensity $I(t)$ is a *disturbance signal*, since we cannot influence its value. It is, however, measurable and we know its origin. We have therefore no particular difficulties to incorporate this disturbance signal in the model construction. There is no essential difference between the handling of the disturbance signal $I(t)$ and the input $u(t)$.

In such cases we are led to a model of the general character

$$\dot{x}(t) = f(x(t), u(t), w(t)) \tag{3.29a}$$

$$y(t) = h(x(t), u(t), w(t)) \tag{3.29b}$$

or, in discrete time,

$$x(t_{k+1}) = f(x(t_k), u(t_k), w(t_k)) \tag{3.30a}$$

$$y(t_k) = h(x(t_k), u(t_k), w(t_k)) \tag{3.30b}$$

[Compare with (3.14), (3.15).] Here we have distinguished between the input u and the disturbance signal w, but as pointed out, this distinction is not essential for the model construction.

A new problem arises when the model is going to be used, however. To get an idea of how the output $y(t)$ really behaves in practice, it is necessary to have a description of the typical properties of the disturbance signal $w(t)$. Consider the curve in Figure 3.12, which depicts the measured solar intensity. Obviously, it is not an easy problem to characterize this signal. We are thus led to the following problem:

$$\text{Describe } w(t) \text{ in a suitable way} \tag{3.31}$$

We will discuss this problem in Sections 3.7-3.8.

However, the situation with a known and measurable disturbance signal is favorable in the following sense: Separately from the construction of the dynamic model, a model can be built for the disturbance signal's properties based on direct measurements. We will return to how this can be done in Chapters 8-10.

Known Disturbance Sources — Nonmeasurable Disturbance Signals

Consider for example an airplane. Its motion is completely determined by the force of the engine, gravity, and the forces the air exerts. The latter depends partly on rudder and flap movements and partly on wind variations. Rudder and flap angles will be inputs. The airplane's speed and orientation will then be state variables (internal variables). The wind variations, finally, are disturbance signals.

When a mathematical model of an airplane is constructed, both the input and the disturbance signals have to be considered. In this case the physical origin of the disturbance signals is known, and their effect on the motion of the airplane is also known (follows Newton's law of motion). We are thus led to a description of the type (3.29) or (3.30) in our model building, where $w(t)$ represents wind forces. When studying airplane models, $w(t)$ has to be given typical and realistic values. The problem (3.31) becomes very significant.

There is, however, an important difference concerning our possibilities in building a disturbance model. Even if we know the origin of

the disturbance signals, they are in this case hardly separately measurable. They will mainly be noticed by their influence on other measurable variables. We have thus only indirect information about the characteristics of the disturbance signal. In principle, we could calculate $w(t)$ "backward" from $y(t)$ and $u(t)$, but we will in any case realize that the modeling of $w(t)$ is integrated into the construction of the mathematical model.

Unknown Disturbance Sources

A common situation is that the interesting signals from the system are influenced by a number of disturbance sources and that we have no possibility, energy, or wish to sort out their physical origins. We then have to lump their effect into a disturbance contribution to the output, which in such cases typically is additive:

$$y(t) = z(t) + w(t) \tag{3.32}$$

$z(t)$ is here the undisturbed output, for example

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \\ z(t) &= h(x(t), u(t)) \end{aligned} \tag{3.33}$$

The problem (3.31) of describing $w(t)$ does however remain unchanged. In this case, obviously, $w(t)$ is not directly measurable. We then must proceed from indirect observations of $w(t)$ inferred from measurements of input and output signals.

Example 3.10 Electric Welding
In electric welding the heat power in the weld depends on the current. This heat in turn influences how fast the welding steel melts and thereby how wide the weld will be. Figure 3.13 shows the related values of current and weld width. The latter is obviously also influenced by a number of other signals, and it is not so easy to get a clear picture of the physical origins of all of them. □

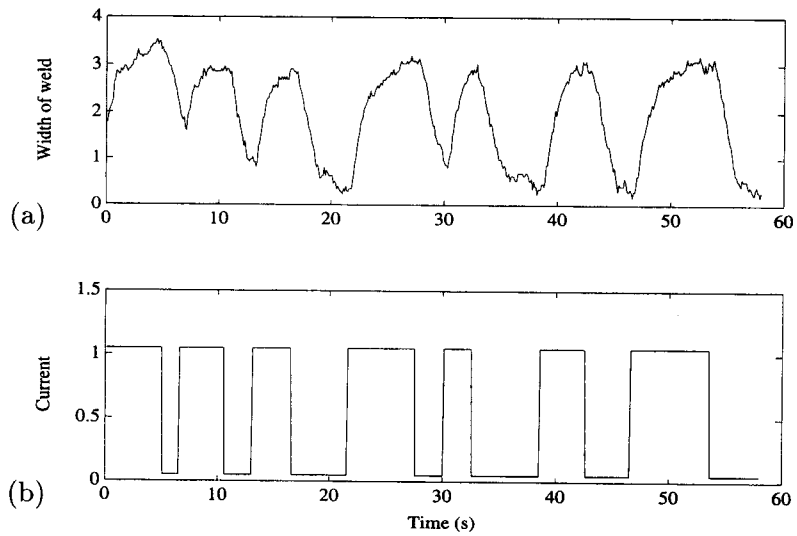


Figure 3.13: (a) The width of the weld. (b) Current in the welding machine.

3.7 Description of Signals in the Time Domain

Deterministic Models

A natural way to approach problem (3.31) is to describe the disturbance signal $w(t)$ as the output from a dynamic system with a simple and known input $u_w(t)$:

$$\begin{aligned} \dot{x}_w(t) &= f_w(x_w(t), u_w(t)) \\ w(t) &= h_w(x_w(t), u_w(t)) \end{aligned} \tag{3.34}$$

and similarly in discrete time. A typical choice of input is an impulse

$$u_w(t) = \delta(t) \tag{3.35}$$

$\delta(t)$ is Dirac's delta function. In discrete time we interpret $\delta(t)$ as a pulse:

$$\delta(t) = \begin{cases} 1 \; t = 0 \\ 0 \; t \neq 0 \end{cases} \tag{3.36}$$

$w(t)$ is then the impulse response to the system in question. If $u_w(t)$ is chosen as a pulse train, i.e regularly or irregularly recurring pulses (impulses), repetitive patterns can be described.

We sometimes have some insight into the generation of the disturbances. The modeling techniques according to Chapters 4-7 can then be used to construct (3.34). Such detailed knowledge is, however, often missing, and we have to resort to obtaining models by identification methods. See Chapters 8-10.

It is common that the model (3.34) be linear. It can then be written as

$$\begin{aligned} \dot{x}_w(t) &= A_w x_w(t) + B_w u_w(t) \\ w(t) &= C_w x_w(t) + D_w u_w(t) \end{aligned} \tag{3.37}$$

and similarly in discrete time.

According to Appendix A [see (A.4)] we can then also represent the disturbance signal by the transfer function

$$W(s) = G_w(s)U_w(s) \tag{3.38}$$

In terms of the time functions (rather than their Laplace transforms) we let the transfer function have the argument $p = d/dt$, that is, the differentiation operator:

$$w(t) = G_w(p)u_w(t) \tag{3.39}$$

In the same way, in discrete time, we then have, according to (A.13), for the signals' z-transforms

$$W(z) = G_w(z)U_w(z) \tag{3.40}$$

which according to Appenix A corresponds to the difference equation.

$$\begin{aligned} &w(t_{k+n}) + d_1 w(t_{k+n-1}) + \ldots + d_n w(t_k) \\ &= c_0 u_w(t_{k+n}) + c_1 u_w(t_{k+n-1}) + \ldots + c_n u_w(t_k) \end{aligned} \tag{3.41}$$

where

$$G_w(z) = \frac{c_0 z^n + c_1 z^{n-1} + \ldots + c_n}{z^n + d_1 z_{n-1} + \ldots + d_n} = \frac{C(z)}{D(z)} \tag{3.42}$$

Similarly to (3.39) in the time domain, we can write

$$w(t_k) = G_w(q)u_w(t_k) \tag{3.43}$$

where q is the shift operator $qy(t_k) = y(t_{k+1})$. For signals that have been sampled uniformly such that $t_k = kT$, we call T the *sampling interval*, the frequency $2\pi/T$ the *sampling frequency* (radians per second), and the frequency π/T the *Nyquist frequency*.

Example 3.11 Wave Models

Waves constitute disturbances for systems like oil drilling platforms and ships. Much work has been done in hydromechanics both in developing models of waves and in trying to find out how they influence ships.

A simple model in the time domain will lead to the following model of the wave profile:

$$w(\ell, t) = \frac{h}{2}\cos(k\ell - \omega t) \tag{3.44}$$

$w(\ell, t)$ is the water level at time t at the (length) coordinate ℓ, h is the wave height, k is the wave number, and ω the wave frequency. The following relationship holds:

$$k = \frac{\omega^2}{g} \tag{3.45}$$

where g is the acceleration of gravity.

The dependence of h and ω on the wind speed v has also been determined experimentally:

$$h = 0.015v^2 + 1.5 \quad (h \text{ in meters, } v \text{ in } m/s)$$

$$\omega = 2\pi/(-0.0014v^3 + 0.042v^2 + 5.6) \quad (\omega \text{ in } rad/s)$$

(Reference: W. G. Price and R. E. D. Bishop: Probabilistic Theory of Ship Dynamics, Chapman-Hall, London, 1974.)

We can write (3.44) in the form (3.37), for example, as

$$\dot{x}_w(t) = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} x_w(t) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \delta(t) \tag{3.46}$$

$$w(\ell, t) = \frac{h}{2}(\cos(k\ell)\ \sin(k\ell))x_w(t)$$

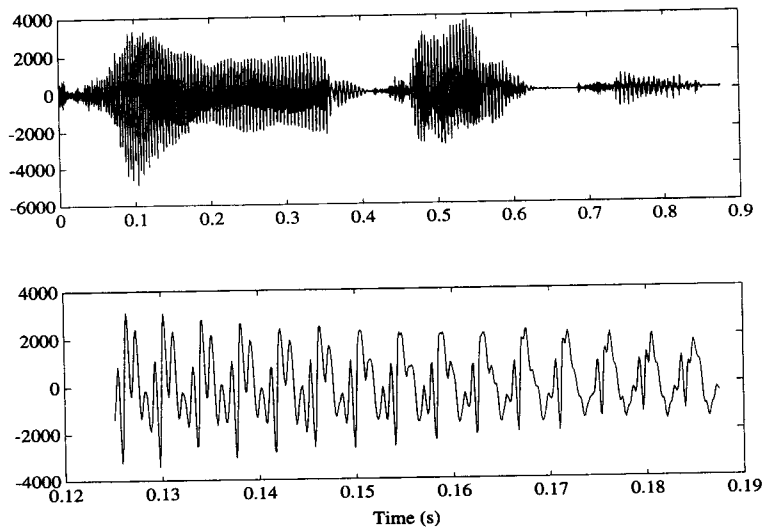


Figure 3.14: Air pressure variation as a result of human speech.

This follows from the fact that the state-space equation has

$$x_w(t) = \begin{pmatrix} \cos(\omega t) \\ \sin(\omega t) \end{pmatrix}$$

as its solution. □

Example 3.12 Models of Human Speech

To be able to store, recognize, transfer, and artificially generate human speech, we need models for the speech signal. Figure 3.14 shows some typical speech signals (air pressure variations as a function of time).

An attempt that has been successfully tested is to describe the speech signal $w(t)$ as

$$w(t+8T) + d_1 w(t+7T) + \cdots + d_8 w(t) = u_w(t) \tag{3.47}$$

where $u_w(t)$ is a pulse train with a frequency equal to the basic pitch of the speech. The sampling interval is typically $T = \frac{1}{8} \cdot 10^{-3} s$ (sampling frequency 8 kHz).

The model (3.47) has a certain physical background. The speech is formed when the vocal cords generate a train of pressure pulses. These

are then filtered in the oral cavity and by the lips. Pressure variations are described by a linear, partial differential equation, for which the form of the oral cavity constitutes boundary values. If the form is described approximately as four consecutive cylinders with different diameters, the solution to the partial differential equation is of the form (3.47), where $u_w(t)$ is the input from the vocal cords.

A certain sound (phoneme) can now be described with the help of the numbers $d_1, d_2, \ldots, d_8$ plus the basic frequency (pitch) in the pulse train. This can be stored or transferred instead of the speech signal itself. □

Stochastic Models

A characteristic feature in a disturbance signal is that its future cannot be predicted exactly. Consider, for example, the solar intensity signal in Figure 3.12. It is not reasonable to consider a model of the type (3.41), which exactly describes the solar intensity. On the other hand, qualified guesses of the expected future behavior can be made. It is therefore natural to introduce stochastic elements in the signal description. This can be done in several ways. It is most easily accomplished in the linear time discrete description (3.41) by choosing $u_w(t_k) = e(t_k)$ as a series of independent stochastic variables, *white noise.* (In the sequel we will deal with uniformly sampled signals: $t_k = k \cdot T$.) This gives

$$\begin{aligned} &w(t) + d_1 w(t-T) + \cdots + d_n w(t-nT) \\ &= c_0 e(t) + c_1 e(t-T) + \cdots + c_n e(t-nT), \\ &\quad e(t) \text{ and } e(s) \text{ independent if } t \neq s \end{aligned} \tag{3.48}$$

The probability distribution of $e(t)$ plays a major role for the typical appearance of $w(t)$. The most common model is that $e(t)$ are independent, normally distributed variables:

$$\begin{aligned} &e(t) \in N(0, \lambda) \text{ (normally distributed} \\ &\text{with mean value 0 and variance } \lambda) \end{aligned} \tag{3.49}$$

This gives "noisy" variables $w(t)$. If, on the other hand, $e(t)$ is zero most of the time, but is a pulse now and then, then $w(t)$ has a different

character. Such a behavior in $e(t)$ can be modeled with the distribution

$$\begin{cases} e(t) = 0 \quad \text{with probability } 1-\mu \\ e(t) \in N(0, \lambda/\mu) \quad \text{with probability } \mu \end{cases} \tag{3.50}$$

Also, in this case $e(t)$ will have zero mean and variance λ.

From (3.48), $w(t)$ will then have properties that depend on c_i and d_i, and on the probability distribution of $e(t)$. See Figure 3.15 for an illustration. This shows realizations of (3.48) when $e(t)$ is a series of independent normally distributed, stochastic variables with zero mean and variance one. The following numbers have been used:

(a) $n = 1$, $d_1 = -0.9$, $c_0 = 1$, $c_1 = 0$

(b) $n = 1$, $d_1 = 0.9$, $c_0 = 1$, $c_1 = 0$

(c) $n = 2$, $d_1 = -0.5$, $d_2 = 0.7$, $c_0 = 1$, $c_1 = 0.5$, $c_2 = 0$

(d) same system as in case (c), but $e(t)$ not normally distributed, instead with the distribution $P(e(t) = 0) = 0.98$, $P(e(t) = \sqrt{50}) = 0.01$, and $P(e(t) = -\sqrt{50}) = 0.01$ (thus still independent with average 0 and variance 1)

The signal $\{w(t)\}$ is defined through (3.48) as a *stochastic process*, that is, a series of stochastic variables with a certain simultaneous distribution. The values that $\{w(t)\}$ assume for a certain turnout of the random variables $\{e(t)\}$ are called a *realization* of the process.

A complete characterization of a stochastic process means that all simultaneous distribution functions for $w(t_1), w(t_2), \ldots, w(t_N)$ are given. In our cases we will let these distribution functions be determined indirectly with the help of the numbers c_i and d_i and $e(t)$'s probability distribution. We will consequently limit ourselves to stochastic processes that are obtained as *linearly filtered white noise.* Such a process is also called an *ARMA process*(AutoRegressiveMovingAverage). The numerator polynomial $C(z)$ is the MApart and the denominator polynomial $D(z)$ is the ARpart. [Compare (3.42).] If $c_i = 0, \quad i \neq 0$, we talk of an *AR process*, and if $d_i = 0$, $i \neq 0$, we have an *MA process.*

If the distribution of $e(t)$ does not depend on t, the properties of $w(t)$ will not depend on absolute time either (after any possible transients have died out). We then have a *stationary process.*

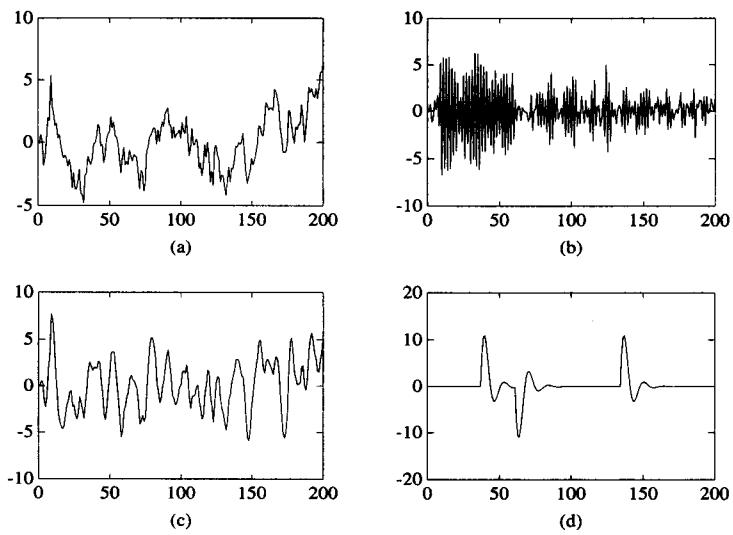


Figure 3.15: Realizations of different stochastic processes. See the text for details.

The time function

$$m_w(t) = Ew(t) \tag{3.51}$$

is called the *mean value function* and the covariance

$$R_w(t,s) = E(w(t) - m_w(t))(w(s) - m_w(s)) \tag{3.52}$$

gives *the covariance function.* Here and elsewhere E denotes *mathematical expectation.* For a stationary process, $R_w(t,s)$ depends only on the time difference $t - s$, and we then write

$$R_w(\tau) = R_w(t+\tau, s) \tag{3.53}$$

The covariance functions for the processes in Figure 3.15 are shown in Figure 3.16.

3.8 Description of Signals in the Frequency Domain

To describe a signal's properties in terms of its frequency contents is appealing both intuitively and from an engineering point of view. This

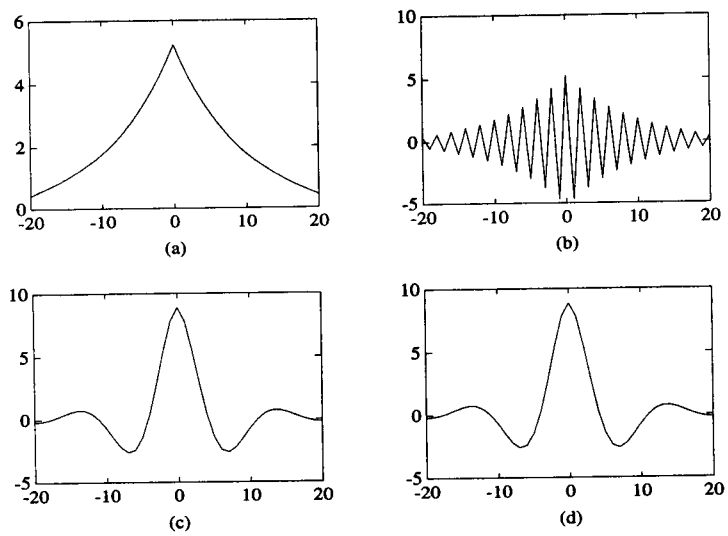


Figure 3.16: The covariance functions for the processes in Figure 3.15.

is probably connected to the fact that our senses are well suited to a frequency description of the observed signals. A new spectral peak in the motor sound of a car — a discord — is revealed quickly by the ear. Added to this is the fact that the mathematical tool for frequency description — the Fourier transform — is powerful.

Spectra

The frequency content of a signal is described by its *spectrum*. We will use the notation $\Phi_w(\omega)$ for $w(t)$'s spectrum. A more precise term is really *spectral density*, since

$$\int_{\omega_1}^{\omega_2} \Phi_w(\omega) d\omega \tag{3.54}$$

is a measure of the signal's energy (power) between the frequencies ω_1 and ω_2. Φ_w thus has the "dimension" of energy (power) per frequency. There are several variations of spectrum definitions, depending on whether the signal is time continuous or time discrete, deterministic or stochastic, whether it has finite or infinite energy, and whether we deal with power or amplitude. The definitions are, however, very closely related and are all meant to describe the (mean) frequency content of

the signal in question. The exact definitions are given in Appendix C. All that is necessary to know is that a signal's spectrum is the square of the absolute value of its Fourier transform, possibly normalized and possibly formed by mathematical expectation. In summary, we have the following:

1. For signals with finite energy, we define (energy) spectrum as the absolute square of the signal's Fourier transform. This is valid for both time continuous and time discrete signals.

2. For signals with infinite energy, we calculate energy spectrum for a truncated signal, normalize with the time interval's length, and then let this interval tend to infinity. The *power spectrum* is thereby defined. This is valid in both continuous and discrete time.

3. For signals that are perceived as realizations of stationary stochastic processes, we define spectrum as the expected value of the realization's power spectrum (more exactly expressed as the limit value of the expected value of the normalized energy spectra for the truncated realizations).

For a given signal, only one of these definitions is relevant. We will therefore not distinguish between the different variants in the notations, but use

$$\Phi_w(\omega) \tag{3.55}$$

for the *spectrum of the signal* $w(t)$. The context will decide which definition applies. In Figure 3.17 the spectra for the disturbance signals in Figure 3.15 are shown. Note especially that for a time discrete signal that has been sampled with sampling interval T the spectrum is defined only up to the Nyquist frequency π/T. (More exactly, the spectrum is symmetric with respect to frequency and periodic with period $2\pi/T$.) It is therefore enough to consider the spectrum between the frequency zero and the Nyquist frequency. [See (C.12).]

Remark: We deal in this book only with energy and power spectra. In some applications it is customary to work with the square root of these spectra, which are called *amplitude spectra.*

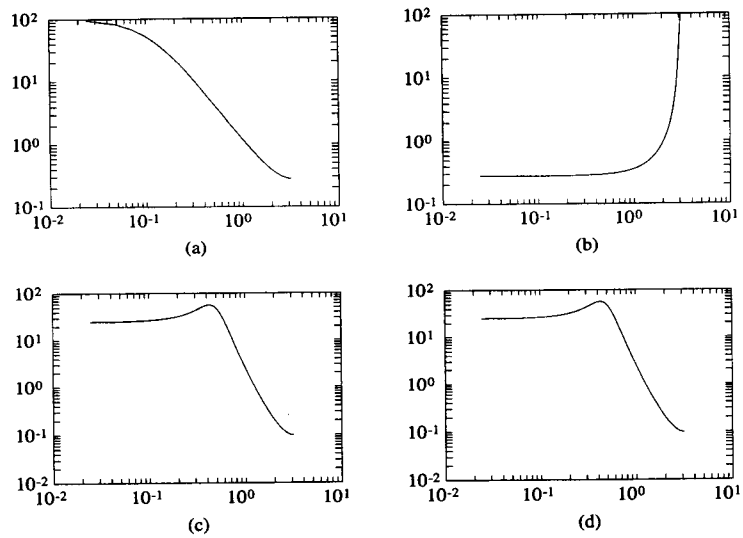


Figure 3.17: Spectra for the processes in Figure 3.15. The x-axis has a logarithmic frequency range, and the y-axis has a logarithmic amplitude range.

Cross Spectra

Consider two signals $u(t)$ and $y(t)$. It is obviously interesting to find out how they "vary together," that is, what relationship exists between the two signals. Analogously to spectra, the *cross spectrum* between u and y,

$$\Phi_{yu}(\omega) \tag{3.56}$$

is defined as the product between the Fourier transform of y and the conjugate of the Fourier transform of u. The result is normalized and the expected value is applied exactly as for spectra. Exact definitions are given in Appendix C. $\Phi_{uu}(\omega) = \Phi_u(\omega)$; that is, the cross spectrum between a signal and the signal itself is what we defined as spectrum earlier. We say that two signals are *uncorrelated* if their cross spectrum is identically zero.

$$u \text{ and } y \text{ uncorrelated} \iff \Phi_{yu}(\omega) \equiv 0 \tag{3.57}$$

Cross spectra are mainly used for signals that are described as realizations of stochastic processes [that is, with the definition (C.22)].

$\Phi_{yu}(\omega)$ is then a complex number equal to the covariance between $Y(\omega)$ and $\overline{U(\omega)}$, that is, the respective signal's Fourier transform at the frequency ω. Intuitively we can think as follows: If $u(t)$ has a "typical" signal component $\cos\omega t$, $y(t)$ will "on the average" have this component $|\Phi_{yu}(\omega)|$ times larger and $\arg\Phi_{yu}(\omega)$ radians phase delayed.

Links with the Time Domain Description

Assume that three signals y, u and w are related by

$$y(t) = G(p)u(t) + w(t) \tag{3.58}$$

[here p is the differentiation operator, see (3.39)], where $u(t)$ and $w(t)$ are uncorrelated. Their Fourier transforms then [compare (A.4)] obey

$$Y(\omega) = G(i\omega)U(\omega) + W(\omega) \tag{3.59}$$

By taking the absolute square of both terms and possibly normalizing by the length of the time interval and possibly using expected value, we have

$$\Phi_y(\omega) = |G(i\omega)|^2\Phi_u(\omega) + \Phi_w(\omega) \tag{3.60}$$

regardless of which spectral definition we apply. Multiplying (3.59) by $\overline{U(\omega)}$ (and possibly normalizing and possibly using the expected value), yields

$$\Phi_{yu}(\omega) = G(i\omega)\Phi_u(\omega) \tag{3.61}$$

If the relationship between y, u and w instead is given as a time discrete expression

$$y(t) = G_T(q)u(t) + w(t) \tag{3.62}$$

[here q is the shift operator, see (3.43)], then

$$\Phi_y(\omega) = |G_T(e^{i\omega T})|^2\Phi_u(\omega) + \Phi_w(\omega) \tag{3.63}$$

holds, and

$$\Phi_{yu}(\omega) = G_T(e^{i\omega T})\Phi_u(\omega) \tag{3.64}$$

[compare (C.25)].

For the special case $w(t) \equiv 0$, $\Phi_u(\omega) \equiv 1$, $G(p) = \frac{C(p)}{D(p)}$ [which corresponds to $u(t)$ being an (im)pulse or white noise and that G is finite dimensional], we have in continuous time

$$\Phi_y(\omega) = |G(i\omega)|^2 = \frac{|C(i\omega)|^2}{|D(i\omega)|^2} \tag{3.65}$$

This gives a simple and obvious link between the signal spectrum and the linear system, $G(p)$, which represents the signal in the time domain with a white noise or (im)pulse input. If a given spectrum $\Phi_y(\omega)$ is a rational function of ω^2, it is always possible to find a stable and inversely stable system $G(p)$ such that (3.65) is valid. (That is all poles and zeros are in the left half-plane.) This is called *spectral factorization.*

In discrete time we analogously have

$$\Phi_y(\omega) = |G_T(e^{i\omega T})|^2 = \frac{|C(e^{i\omega T})|^2}{|D(e^{i\omega T})|^2} \tag{3.66}$$

It is always possible to find a stable and inversely stable $G_T(q)$ if $\Phi_y(\omega)$ is a rational function of $\cos \omega T$.

Use of Spectral Description

To characterize the properties of a signal by plotting or verbally describing its spectrum is, as we said, both practical and intuitively appealing. "We have low frequency disturbances, almost all under 5 Hz," "The disturbances are concentrated to the region around 50 Hz," and so on. The link to a more exact mathematical description is obtained from the spectral factorization (3.65). The function $\Phi_y(\omega)$ is then formed, and its ω dependence is reflected in the verbal description. The polynomial C and D can then be determined in (3.65) or its time discrete counterpart (3.66), and a time domain description is thereby obtained.

Example 3.13 Wind Models

Models of wind strength and wind gusts play, as we pointed out earlier, an important role in many applications. Much work has been done to obtain such models. One of the most used models is the von Karman

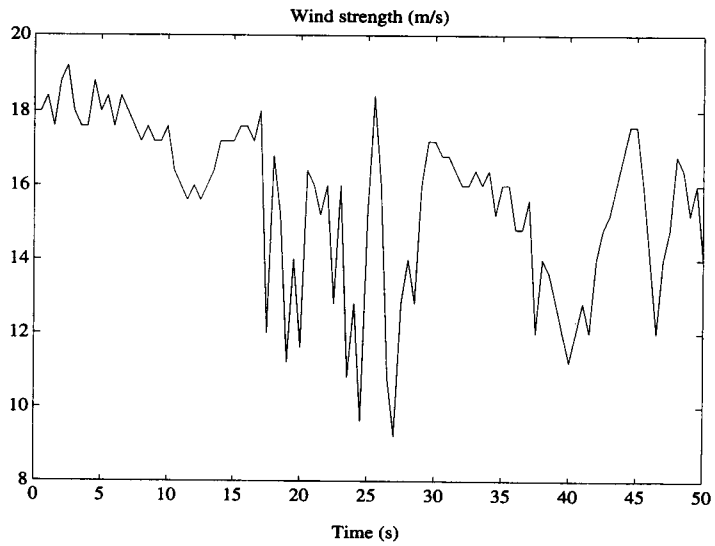


Figure 3.18: Wind strength as a function of time.

spectrum

$$\Phi(\omega) = \frac{K}{(1 + (\omega T)^2)^{5/6}} \tag{3.67}$$

The model is a frequency description. Figure 3.18 shows a typical wind strength signal and Figure 3.19 shows the von Karman spectrum. Most other wind models are of the general form

$$\Phi(\omega) = \frac{K|\omega|^\gamma}{(1 + (\omega T)^2)^{5/6}} \tag{3.68}$$

of which (3.67) is a special case. □

Finally it has to be said that in practice we seldom can *calculate* the spectrum for an observed signal. We have only a finite sequence of numbers $\{w(1), w(2), \ldots, w(N)\}$ available, and all spectral definitions contain limits when $N \to \infty$ and/or expected values. We instead have to be satisfied by *estimating* spectra from the measured values. How this is done will be discussed in Chapter 8.

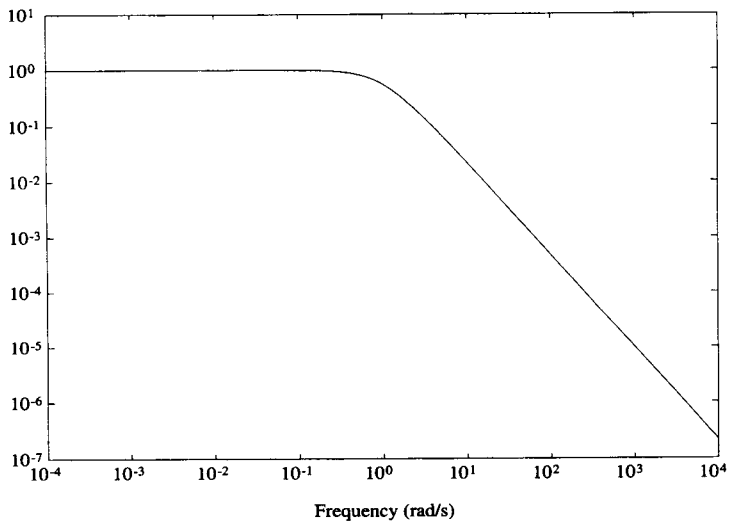


Figure 3.19: Von Karman's spectrum for wind strength.

3.9 Links between Continuous Time and Discrete Time Models

So far we have treated time continuous models of systems and signals, like (3.14), parallel to the time discrete counterparts, like (3.15). Many of the formal properties, especially for linear systems are also analogous. Which representation to choose, time continuous or time discrete, depends partly on how it is natural to model and partly on the purpose of the model.

If the system is mechanical–physical–technical and the modeling is done based on laws of nature, it is often easier to build a time continuous model. This naturally is due to the fact that most laws of nature are formulated as time continuous differential equations. If the model building is done based on measured data, we are often led to time discrete models, simply because data are collected at time discrete moments. We say that the signals are *sampled.* Economic models are also often built in discrete time due to the fact that it is natural to think of economic variables as sampled data: "GNP this year depends on GNP last year and ... ."

In this section we will discuss the links between a time discrete

and a time continuous model of the same system. When is it interesting to use such relationships? A typical situation is when a time discrete model has been built based on measured data and we want to compare it to a time continuous one, which is based on physical modeling. Another situation is when we have a time continuous model and want to describe how states and outputs vary between different sampling instants while the input, for example, is constant between the sampling instants. This of course can always be dealt with by simulation, but sometimes it is important to have an analytical expression for the variations, that is, a time discrete model. A typical application of this is when we want to use the time continuous model to determine computer-based regulators (mechanisms for computing a suitable input).

Time Continuous and Time Discrete Models

We are now going to discuss the formal links between a time continuous model

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)) \end{aligned} \tag{3.69}$$

and a time discrete model

$$\begin{aligned} x(t_{k+1}) &= \tilde{f}(x(t_k), u(t_k)) \\ y(t_k) &= \tilde{h}(x(t_k), u(t_k)) \end{aligned} \tag{3.70}$$

These links can be divided into *approximate* relationships, which are based on difference approximations of $\dot{x}$ or the Taylor expansion of $x(t)$, and *exact* relations, based on the analytical solution to (3.69) over the time interval $[t_k, t_{k+1}]$.

Let us first consider the approximate relations. The simplest approximation (the Euler approximation) of $\dot{x}(t)$ is

$$\dot{x}(t) \approx \frac{x(t_{k+1}) - x(t_k)}{t_{k+1} - t_k} \tag{3.71}$$

This gives

$$x(t_{k+1}) \approx x(t_k) + (t_{k+1} - t_k) f(x(t_k), u(t_k)) \tag{3.72}$$

The link between (3.69) and (3.70) is thus

$$\tilde{f}(t_k, x, u) \approx x + (t_{k+1} - t_k) f(x, u) \tag{3.73}$$

Note that according to the mean value theorem the expression (3.72) will be exactly valid, in the scalar case, if in $f_i(x(t_k), u(t_k))$ we replace $x(t_k)$ and $u(t_k)$ with values between $x(t_k)$ and $x(t_{k+1})$ and $u(t_k)$ and $u(t_{k+1})$, respectively. The quality in the approximation (3.72) thus depends on how much the variables $x(t)$ and $u(t)$ change over the time interval (t_k, t_{k+1}). In other words, the approximation (3.73) is good if the time interval $t_{k+1} - t_k$ is small compared to the model's time constants.

The approximation can of course be improved in this regard by replacing the difference approximation (3.71) with higher-order approximations or by using more expressions in the Taylor expansion (3.73).

Let us now discuss the exact links that can be established between (3.69) and (3.70). We are only going to study linear equations.

Linear Models

More precise expressions can be obtained for linear models. Consider a system with the transfer function $G(s)$ and the state representation

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{3.74}$$

If the values of the input at each time instant can be reconstructed from its values at the sampling instants, we should be able to translate (3.74) exactly to discrete time. The simplest case is when the input is piecewise constant:

$$u(t) = u(t_k) \qquad \text{for} \quad t_k \le t < t_{k+1} \tag{3.75}$$

This is the case for instance in computer-controlled systems. We then have $x(t_k)$ and $y(t_k)$ given exactly by

$$x(t_{k+1}) = \tilde{A}_k x(t_k) + \tilde{B}_k u(t_k) \tag{3.76a}$$

$$y(t_k) = Cx(t_k) + Du(t_k) \tag{3.76b}$$

where

$$\tilde{A}_k = e^{A(t_{k+1}-t_k)} \tag{3.77a}$$

$$\tilde{B}_k = \int_0^{t_{k+1}-t_k} e^{A\tau} B d\tau \tag{3.77b}$$

See Appendix A. Observe that (3.77) also can be used to compute A and B in (3.74) from given $\tilde{A}_k$ and $\tilde{B}_k$.

If the input is piecewise linear,

$$u(t) = u(t_k) + (t - t_k)\frac{u(t_{k+1}) - u(t_k)}{t_{k+1} - t_k} \quad t_k \le t \le t_{k+1} \tag{3.78}$$

we can also obtain an exact relationship. The simplest way to achieve this link is to reason as follows: If $u(t)$ is piecewise linear, then $\dot{u}(t)$ is piecewise constant. Therefore, form a state representation for the system

$$G(s) \cdot \frac{1}{s}$$

and run it with $\dot{u}(t)$. Sample the state representation for $G(s)/s$ according to (3.76)-(3.77). We then have a model of the form

$$\begin{aligned} x(t_{k+1}) &= \tilde{A}_k x(t_k) + \tilde{B}_k \tfrac{u(t_{k+1})-u(t_k)}{t_{k+1}-t_k} \\ y(t_k) &= Cx(t_k) + D\tfrac{u(t_{k+1})-u(t_k)}{t_{k+1}-t_k} \end{aligned} \tag{3.79}$$

If the real input is neither piecewise constant nor piecewise linear, the expressions are not exactly valid. If the signal is twice continuously differentiable, it can be shown that the approximation (3.75) will give rise to an error in $y(t_k)$ bounded by $C \cdot \sup_{t\in[t_k,t_{k+1}]} |\dot{u}(t)| \cdot T$, while the approximation (3.78) gives an error of no more than

$$C \cdot \sup_{t\in[t_k,t_{k+1}]} |\ddot{u}(t)| \cdot T^2$$

Here is $T = \max |t_{k+1} - t_k|$ and C is a constant, independent of u and T.

Transfer Functions

Consider the time continuous model (3.74) with the frequency function $G(i\omega)$. If $t_k = kT$ and the input is piecewise constant, the sampled model (3.76) gives a sampled frequency function [see (A.14)]

$$G_T(e^{i\omega T}) = C(e^{i\omega T} \cdot I - \tilde{A})^{-1}\tilde{B} + D \tag{3.80}$$

Even if $G(i\omega)$ and $G_T(e^{i\omega T})$ describe the same system and give exactly the same values of $y(t_k)$ [when (3.75) holds], they themselves will not

be identical functions of ω. It can be shown (see the appendix to this chapter) that

$$|G(i\omega) - G_T(e^{i\omega T})| \leq \omega \cdot T \cdot \int_0^\infty |g(\tau)| d\tau \tag{3.81}$$

where $g(\tau)$ is the impulse response of $G(s)$. The two frequency functions are thus in good agreement for low frequencies. The rule of thumb is that this good agreement usually extends to up to one-tenth of the sampling frequency $[\omega < 2\pi/(10T)]$. For a system that is sampled fast compared to the interesting frequencies, it follows that the sampled frequency function gives a good picture also of $G(i\omega)$.

Example 3.14 Sampling of a System

Consider a time continuous system with the transfer function

$$G(s) = \frac{1}{s^2 + s + 1} \tag{3.82}$$

Its impulse response and frequency function are shown in Figures 3.20 and 3.21, respectively. If we represent (3.82) in a state-space form and sample according to (3.76)–(3.77) for some different sampling intervals T and then compute the time discrete transfer function, we have

$$\begin{aligned}
T = 0.1s: &\quad G_T(q) = \tfrac{0.0048q^{-1}+0.0047q^{-2}}{1-1.8953q^{-1}+0.9048q^{-2}} \\
T = 0.5s: &\quad G_T(q) = \tfrac{0.1044q^{-1}+0.0883q^{-2}}{1-1.4138q^{-1}+0.6065q^{-2}} \\
T = 1s: &\quad G_T(q) = \tfrac{0.3403q^{-1}+0.2417q^{-2}}{1-0.7859q^{-1}+0.3679q^{-2}} \\
T = 2s: &\quad G_T(q) = \tfrac{0.8494q^{-1}+0.40409q^{-2}}{1+0.1181q^{-1}+0.1353q^{-2}}
\end{aligned} \tag{3.83}$$

The impulse response and the frequency function $G_T(e^{i\omega T})$ for these time discrete systems are also shown in Figures 3.20 and 3.21. We see, as expected, that the shorter the sampling interval the better the correspondence will be between the sampled and time continuous systems. □

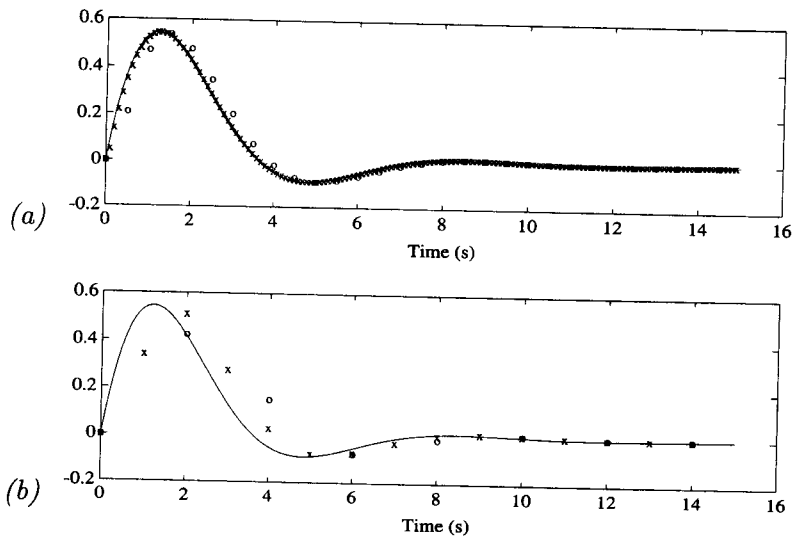


Figure 3.20: The continuous system's impulse response (solid line) together with the sampled system's impulse response. (a) $x : T = 0.1$ s, $o : T = 0.5$ s. (b) $x : T = 1$ s, $o : T = 2$ s. The levels of the impulses have been fitted so that they all give the same energy.

Signal Spectra

Consider a time continuous signal

$$w(t) \quad t > 0 \tag{3.84}$$

that has been sampled with the sampling frequency $\omega_s = 2\pi/T$, which gives the time discrete signal

$$w[k] = w(kT), \quad k = 1, 2, ... \tag{3.85}$$

The *Nyquist frequency* is then given by $\omega_N = \omega_s/2$.

Let the Fourier transform of (3.84) be $W(\omega)$ [see (C.2)] and that of (3.85) be $W^{(T)}(\omega)$ [see (C.8)]. *Poisson's summation formula* then says that

$$W^{(T)}(\omega) = \sum_{r=-\infty}^{\infty} W(\omega + r\omega_s) \tag{3.86}$$

The proof is given in Appendix C. From Poisson's summation formula a number of facts can be obtained. We see that $W^{(T)}(\omega)$ is periodic

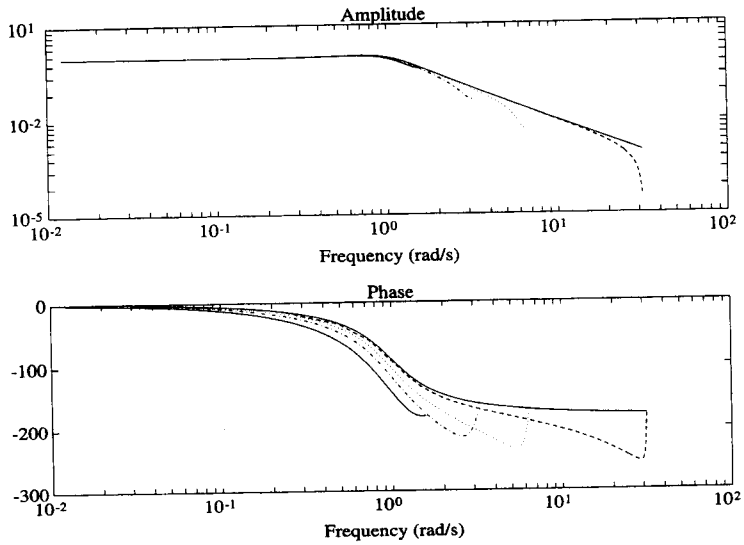


Figure 3.21: The continuous system's Bode diagram together with the sampled systems'. The curves are, from top to bottom, the continuous system, $T = 0.1$ s, $T = 0.5$ s, $T = 1$ s, $T = 2$ s. Note that they end at their respective Nyquist frequencies.

with period ω_s. It is thus enough to consider it over the interval

$$I_T : \quad -\omega_N \leq \omega \leq \omega_N \tag{3.87}$$

The frequencies in the time continuous signal that lie outside this interval are "folded into" I_T and are interpreted as slower frequencies in the sampled signal. The *sampling theorem* follows from this: If the time continuous signal's spectrum is zero outside I_T, the sampled signal will have exactly the same spectrum. In other words, no information is lost at the sampling.

The practical consequence is that, if (3.84) has insignificant energy above the Nyquist frequency (for example eliminated by a presampling filter), then the spectrum of (3.85) can be used as a good approximation of the spectrum of (3.84).

3.10 Appendix

Proof of (3.81)

Let

$$G(i\omega) = \int_0^\infty g(\tau)e^{-i\tau\omega}d\tau$$

We have

$$G_T(e^{i\omega T}) = \sum_{k=1}^{\infty} \tilde{g}_k e^{-ik\omega T}$$

where $\tilde{g}_k$ is the sampled system's impulse response. From

$$y(\ell T) = \int_0^\infty g(\tau)u(\ell T - \tau)d\tau = \sum_{k=1}^{\infty} \int_{(k-1)T}^{kT} g(\tau)d\tau u(\ell T - kT)$$

which is valid if (3.75) is true ($t_k = kT$), we realize that

$$\tilde{g}_k = \int_{(k-1)T}^{kT} g(\tau)d\tau \tag{3.88}$$

Thus

$$\begin{aligned}
|G(i\omega) - G_T(e^{i\omega T})| &= \left|\sum_{k=1}^{\infty} \int_{(k-1)T}^{kT} \left(g(\tau)e^{-i\tau\omega} - g(\tau)e^{-ik\omega T}\right)d\tau\right| \\
&\leq \sum_{k=1}^{\infty} \max_{(k-1)T\leq\tau\leq kT} |e^{-i\omega\tau} - e^{-ik\omega T}| \int_{(k-1)T}^{kT} |g(\tau)|d\tau \\
&\leq \omega T \cdot \sum_{k=1}^{\infty} \int_{(k-1)T}^{kT} |g(\tau)|d\tau = \omega T \int_0^\infty |g(\tau)|d\tau
\end{aligned}$$

which gives (3.81).

Part II

Physical Modeling

Deriving Equations from Physical Knowledge

This part of the book discusses how to construct a mathematical model from knowledge of the basic mechanisms of a system. We will use the term *physical modeling*, since it is usually a knowledge of physics that is relevant in the examples we consider. The principles that we discuss, mainly in Chapter 4, are useful for most types of models, however. The knowledge of physics will then have to be replaced by other relevant knowledge of chemistry, biology, economics, or whatever is required.

It is unrealistic to expect a general methodology for modeling all the types of systems that are possible. There is, however, a way of structuring the problem in different phases that is fairly general. This is discussed in Chapter 4. For many engineering applications it turns out that the same type of equations appear despite the diversity of the physical systems. These analogies are discussed in Chapter 5. Starting from these analogies it is possible to do systematic modeling for a broad class of systems. This is done using *bond graphs*, which are presented in Chapter 6. Finally, we discuss in Chapter 7 how modern computer facilities can help modeling.

Deriving Equations from Physical Knowledge

This part of the book discusses how to construct a mathematical model from knowledge of the basic mechanisms of a system. We will use the term *physical modeling*, since it is usually a knowledge of physics that is relevant in the examples we consider. The principles that we discuss, mainly in Chapter 4, are useful for most types of models, however. The knowledge of physics will then have to be replaced by other relevant knowledge of chemistry, biology, economics, or whatever is required.

It is unrealistic to expect a general methodology for modeling all the types of systems that are possible. There is, however, a way of structuring the problem in different phases that is fairly general. This is discussed in Chapter 4. For many engineering applications it turns out that the same type of equations appear despite the diversity of the physical systems. These analogies are discussed in Chapter 5. Starting from these analogies it is possible to do systematic modeling for a broad class of systems. This is done using *bond graphs*, which are presented in Chapter 6. Finally, we discuss in Chapter 7 how modern computer facilities can help modeling.

Chapter 4

Principles of Physical Modeling

4.1 The Phases of Modeling

In this chapter we will discuss how to build mathematical models of dynamical systems. We will thus consider the problem of how to arrive at a model of the form

$$\begin{aligned} \frac{d}{dt}x(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)) \end{aligned} \tag{4.1}$$

starting from a physical, engineering, biological, economic, or other system.

Modeling is, in common with other scientific and engineering activities, as much an art as a science. Successful modeling is based as much on a good feeling for the problem and common sense as on the formal aspects that can be taught. In this chapter we will take a pragmatic commonsense attitude toward modeling. Later, in Chapter 6, we will demonstrate a systematic approach that solves many but not all modeling problems.

We can distinguish the following three phases in the work of arriving at a mathematical model:

1. The problem is structured.
2. The basic equations are formulated.
3. The state-space model is formed.

Phase 1 consists of an attempt to divide the system into subsystems, an effort to determine cause and effect, what variables are important and how they interact. When doing this work, it is important to know the intended use of the model. The result of phase 1 is a block diagram or some similar description. This phase puts the greatest demands on the modeler in terms of the understanding of and intuition for the physical system. It is also in this phase that the level of complexity and degree of approximation are determined.

Phase 2 involves the examination of the subsystems, the blocks, that the structuring of phase 1 produced. The relationships between variables and constants in the subsystems are formed. In doing that, we use those laws of nature and basic physical equations that are assumed to hold. This often means that we introduce certain approximations and idealizations (point mass, ideal gas, and the like) in order to avoid too complicated expressions. For nontechnical systems, for which generally accepted basic equations are often lacking, this phase gives the opportunity for new hypotheses and innovative thinking.

Phase 3, in contrast to the other phases, is a more formal step aiming at the suitable organization of the equations and relationships left by phase 2. Even if the model in some sense is defined already after phase 2, this step is usually necessary to give a model suitable for analysis and simulation. During phase 3 a computer algebra program can be helpful (see Chapter 7), and it is not always necessary to carry the work all the way to an explicit form (4.1). For the purpose of simulation it might be enough to arrive at state-space models for the subsystems together with instructions for the interconnections.

These three phases will be discussed in more detail in the following sections of this chapter, where we will also illustrate their use in a physical example. An important problem in modeling is to find a model that is not too complicated. This means that we must always make simplifications, idealizations, and approximations. The simplification of models is discussed in Section 4.6

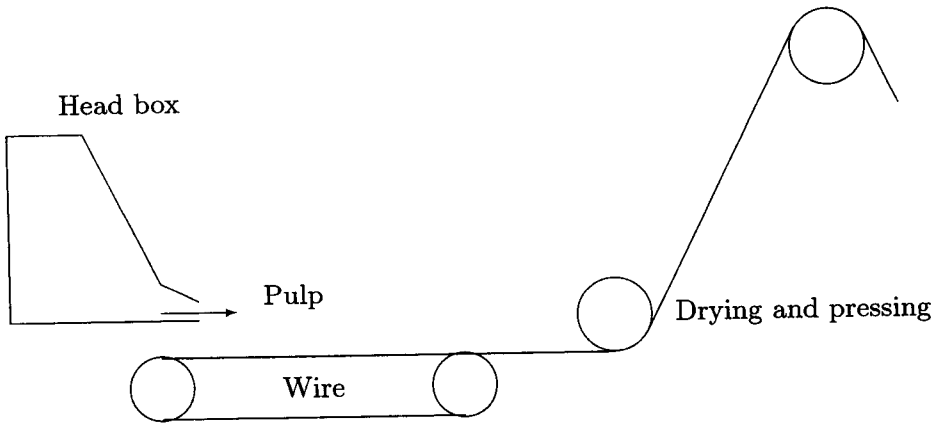


Figure 4.1: A paper machine.

4.2 An Example: Modeling the Head Box of a Paper Machine

To give a concrete dimension to our discussion of the phases of modeling, we will illustrate their application to a physical example, the head box of a paper machine.

The manufacturing of paper from paper pulp is basically done in the following way. The pulp is a dilute solution of fibers in water. It is poured over the wire, which is a continuously moving fine mesh, 10–20 m long. On the wire most of the water drains away, producing a sheet of paper, which then is dried, pressed, and rolled. See the schematic diagram of Figure 4.1. The pulp is delivered to the wire by a *head box*. It is important that this be done in a well-controlled manner in order to get uniform paper quality. This is achieved by forcing the pulp through a narrow slit. Modern head boxes use compressed air to achieve an even flow and a sufficiently high velocity ($\approx$ 10m/s). A paper machine head box is shown in Figure 4.2. We will illustrate the principles of this chapter by modeling the head box.

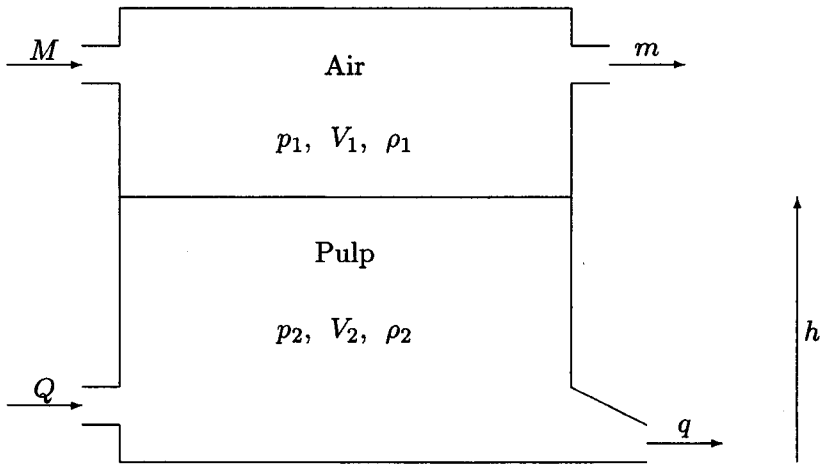


Figure 4.2: Diagram of a head box. The variables are defined in Section 4.3.

4.3 Phase 1: Structuring the Problem

When addressing a modeling task, the main difficulty is often understanding the general structure:

- What signals are of interest (that is, to be considered as outputs)?
- Which quantities are important to describe what happens in the system?
- Of these quantities, which are time varying and which should be regarded as internal variables?
- What quantities are approximately time invariant and should be regarded as constants?
- What variables affect certain other variables?
- Which relationships are static and which are dynamic? (Compare Figure 3.6.)

Answering these questions can demand considerable insight into the system and much work, but it is always necessary as a first step in modeling. When modeling an existing system, we often use simple experiments to assist these preliminary steps, for example, to determine time constants and the influences of signals (see Section 8.1).

Note that the intended use of the model must be known when we answer the preceeding questions. A model that will only be used for simple, order of magnitude calculations (what are the dominating time constants and the approximate static gain?) can allow crude approximations and can neglect many effects and variables. A model to be used for important design decisions demands greater care when answering the questions. The intended use of the model thus determines its complexity.

In later modeling work, maybe after the beginning of simulations, we get a deeper insight into the properties of the system. Then it is not uncommon to go back to this initial stage to revise some decisions.

In some cases it is natural to work with several models in parallel. These models can have very different complexity and can be used to answer different questions. When modeling jet engines, we sometimes uses models with hundreds of states to study thermal problems and models with a handful of states to design regulators.

When we have decided what variables in the system are of interest, the next step is to make clear how they interact. At this stage it is mainly the question of cause and effect that is considered. The quantitative expression, formulas, and equations are introduced in phase 2 (see Section 4.1). Often the result of this analysis is presented as a block diagram (see Section 3.1), which can be seen as the result and goal of phase 1.

Summarizing phase 1 of the modeling, we have to accomplish the following:

- Determine outputs and external signals for the model. Decide what internal variables are important to describe the system.
- Illustrate the interactions between external signals, internal variables, and outputs in a block diagram.

Sometimes, in particular for more complex systems, it is convenient to first divide the system into subsystems and then divide the subsystems

further into blocks. In that case we iterate between these two steps.

An Example of Phase 1: The Head Box

Consider the head box of Section 4.2 and Figure 4.2. We are given the following:

A Inputs to the system:
M: air flow rate (mass flow) (kg/s)
Q: pulp flow rate (volume flow) (m^3/s)

B Outputs from the system:
Here we can choose what interests us. The output pulp flow rate q should definitely be of interest, since it determines what is delivered onto the wire. It is also suitable to regard the pulp level h as an output, since there will be practical restrictions on it. We also choose the excess pressure p_e of the air pad as an output:
q: output pulp flow rate (volume flow) (m^3/s)
h: pulp level (m)
p_e: excess pressure of air pad (N/m^2)

C Division into subsystems:
It is natural to regard the paper pulp and the air pad as separate subsystems. The variables affecting the air subsystem are M and V_1 (the available volume), while the pulp subsystem has the inputs Q and p_e. We have the following relationships between these variables:
M: is an input to the overall system
V_1: depends on the pulp level h
Q: is an input to the overall system
p_e: is an output of the air subsystem

This gives us the block diagram of Figure 4.3.

To get a more detailed block diagram, we introduce the variables and constants of the subsystem in question.

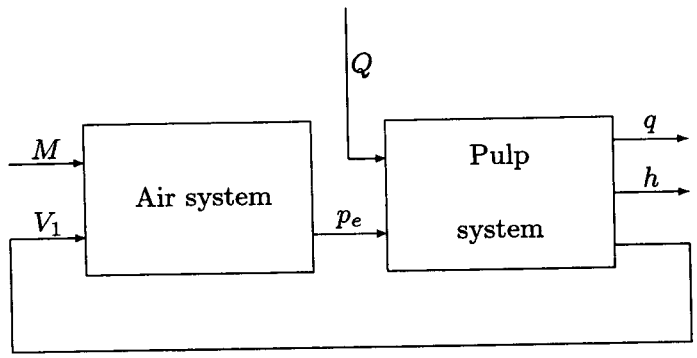


Figure 4.3: Block diagram for the head box.

Air Subsystem

Inputs:
M: inflow of air (kg/s)
V_1: volume of air (m^3)

Output:
p_e: excess pressure of the air (N/m^2)

Internal variables:
ρ_1: density of air (kg/m^3)
m: mass outflow of air (kg/s)
p_1: pressure in air pad (N/m^2)
N: mass of air in air pad (kg)

Constants:
T: absolute temperature of air (K)
(We regard the physical processes in the air pad as isothermal.)
a_1: cross-sectional area of air outflow (m^2)
R: gas constant for air ($m^2/K/s^2$)
p_0: atmospheric pressure (N/m^2)

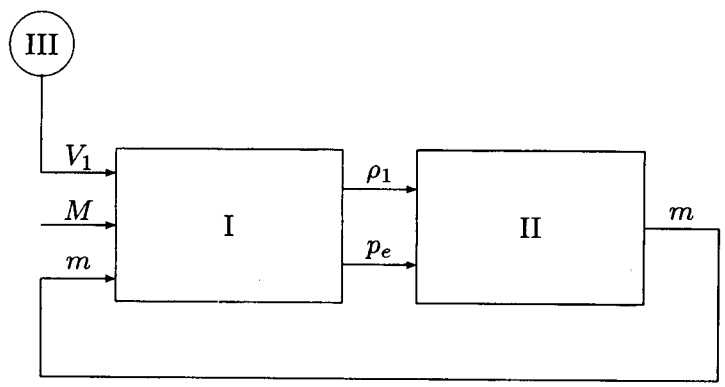


Figure 4.4: Air subsystem. III marks a signal from block III of Figure 4.5.

Pulp Subsystem

Inputs:
Q: input flow rate (m^3/s)
p_e: excess pressure in air pad (N/m^2)
Outputs:
q: output flow rate (m^3/s)
h: pulp level (m)
Internal variables:
h_{eff}: the effective pulp level (m) (see Section 4.4)
V_2: pulp volume (m^3)
Constants:
A: cross sectional area of head box (m^2)
a_2 : cross sectional area of slit (m^2)
C: coefficient of slit area (see Section 4.4)
V: total volume of head box (m^3)
ρ_0: density of pulp (kg/m^3) (assumed to be incompressible)
g: gravitational acceleration (m/s^2)

Some simple considerations now lead to the block diagrams of Figure 4.4 and 4.5.

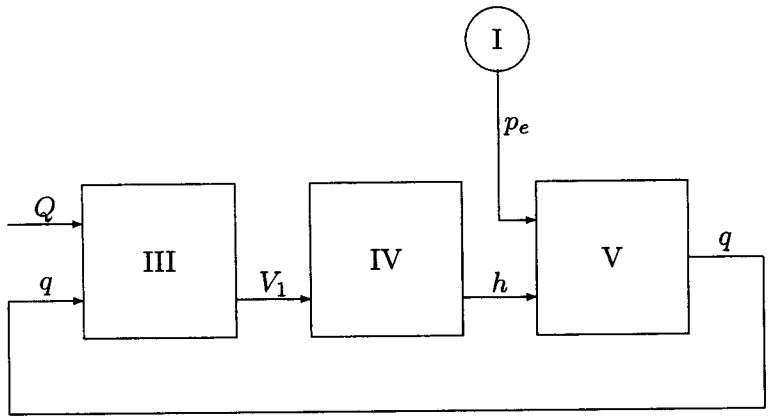


Figure 4.5: Pulp subsystem.

4.4 Phase 2: Setting up the Basic Equations

We now have to transform the block diagram model derived in phase 1 into a mathematical model. We do this by formulating quantitative relationships between the inputs and outputs of the different blocks. In this phase, we use knowledge of mechanics, physics, economics, and the like.

The relationships between the system variables can be of different kinds. Sometimes they go back to some reliable and well-established law of nature, like Ohm's law describing the relationship between current and voltage for a resistor. In other cases the relationship might be given by an experimental curve giving, for example, the pressure over a valve as a function of the flow. A third situation occurs when we use simple formulas that should describe the general character of the relationship but are obviously crude approximations. That was the situation for the ecological model and the economic model in Sections 2.2 and 2.4, respectively.

The work in phase 2 is therefore very problem dependent. However, for systems in physics and engineering we can give certain guidelines for the organization of the work. To that end we note that relationships between physical quantities usually can be divided into two groups: conservation laws and constitutive relationships.

1. *Conservation laws* relate quantities of the same kind. Common examples are the following:

   - Power in – power out = stored energy per unit of time
   - Input flow rate – output flow rate = stored volume per unit of time
   - Input mass flow rate – output mass flow rate = stored mass per unit of time

   Kirchhoff's laws (the sum of currents at a junction is zero, the sum of all voltages around a circuit is zero) are also examples of conservation laws.

   The conservation laws express the conservation of some basic physical quantity: conservation of energy, mass, electrons (Kirchhoff's law), and so on

2. *Constitutive relationships* relate quantities of different kinds. Examples are the relationships between voltage and current for a resistor, a capacitor, or an inductor, the relationship between level and output flow of a tank and the relationship between pressure drop and flow for a valve.

   Constitutive relations often describe the properties of a certain material or a certain component or block in the system. This is illustrated in Figure 4.6

   Note that the relationships are *static* in the chosen variables. From the user's point of view there is no fundamental difference between the law of nature of Figure 4.6a and the experimental curve (or table) of Figure 4.6b, which is specific for a certain valve. In practice it is of course easier to use simple, linear relationships.

   Constitutive relations, like those of Figure 4.6, are always approximate. The inaccuracy depends on the type of system and on the care that was used to determine the curve.

Remark: For practical reasons, we will also count relationships that follow from definitions as constitutive relationships. The curve in Figure 4.6c, for example, can be derived from conservation of energy.

1. *Conservation laws* relate quantities of the same kind. Common examples are the following:

   - Power in – power out = stored energy per unit of time
   - Input flow rate – output flow rate = stored volume per unit of time
   - Input mass flow rate – output mass flow rate = stored mass per unit of time

   Kirchhoff's laws (the sum of currents at a junction is zero, the sum of all voltages around a circuit is zero) are also examples of conservation laws.

   The conservation laws express the conservation of some basic physical quantity: conservation of energy, mass, electrons (Kirchhoff's law), and so on

2. *Constitutive relationships* relate quantities of different kinds. Examples are the relationships between voltage and current for a resistor, a capacitor, or an inductor, the relationship between level and output flow of a tank and the relationship between pressure drop and flow for a valve.

   Constitutive relations often describe the properties of a certain material or a certain component or block in the system. This is illustrated in Figure 4.6

   Note that the relationships are *static* in the chosen variables. From the user's point of view there is no fundamental difference between the law of nature of Figure 4.6a and the experimental curve (or table) of Figure 4.6b, which is specific for a certain valve. In practice it is of course easier to use simple, linear relationships.

   Constitutive relations, like those of Figure 4.6, are always approximate. The inaccuracy depends on the type of system and on the care that was used to determine the curve.

Remark: For practical reasons, we will also count relationships that follow from definitions as constitutive relationships. The curve in Figure 4.6c, for example, can be derived from conservation of energy.

1. *Conservation laws* relate quantities of the same kind. Common examples are the following:

   - Power in – power out = stored energy per unit of time
   - Input flow rate – output flow rate = stored volume per unit of time
   - Input mass flow rate – output mass flow rate = stored mass per unit of time

   Kirchhoff's laws (the sum of currents at a junction is zero, the sum of all voltages around a circuit is zero) are also examples of conservation laws.

   The conservation laws express the conservation of some basic physical quantity: conservation of energy, mass, electrons (Kirchhoff's law), and so on

2. *Constitutive relationships* relate quantities of different kinds. Examples are the relationships between voltage and current for a resistor, a capacitor, or an inductor, the relationship between level and output flow of a tank and the relationship between pressure drop and flow for a valve.

   Constitutive relations often describe the properties of a certain material or a certain component or block in the system. This is illustrated in Figure 4.6

   Note that the relationships are *static* in the chosen variables. From the user's point of view there is no fundamental difference between the law of nature of Figure 4.6a and the experimental curve (or table) of Figure 4.6b, which is specific for a certain valve. In practice it is of course easier to use simple, linear relationships.

   Constitutive relations, like those of Figure 4.6, are always approximate. The inaccuracy depends on the type of system and on the care that was used to determine the curve.

Remark: For practical reasons, we will also count relationships that follow from definitions as constitutive relationships. The curve in Figure 4.6c, for example, can be derived from conservation of energy.

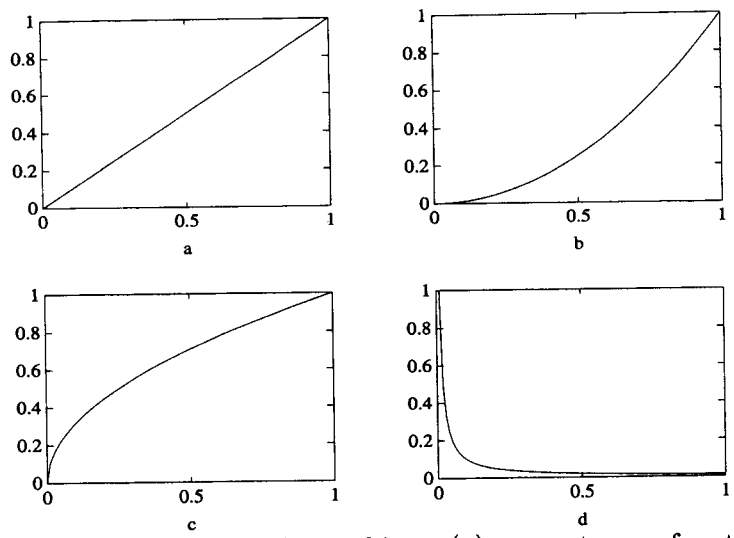


Figure 4.6: Constitutive relationships. (a) current as a function of voltage for a resistor. (b) flow as a function of area for a valve. (c) output flow as a function of level for a tank. (d) pressure as a function of volume for an ideal gas at constant temperature. (All variables are normalized to the interval 0-1.)

A good way of formulating the basic equations of a block is the following:

- Write down the conservation law(s) that is relevant for the block/subsystem.

- Use suitable constitutive relationships to express the conservation laws in the model variables. Calculate the dimensions of the different quantities as a check.

We illustrate the procedure using the head box example.

An Example of Phase 2: The Head Box

We continue the example from Figures 4.4 and 4.5. We get the following results for the different subsystems.

Air System I + II

Conservation law (conservation of mass)

$$\dot{N} = M - m \tag{4.2}$$

Constitutive relationships

$$N = \rho_1 \cdot V_1 \qquad (\text{ mass} = \text{density} \cdot \text{volume}) \tag{4.3}$$

$$p_1 = R \cdot T \cdot \rho_1 \qquad (\text{pressure}) \tag{4.4}$$

The mass flow m is determined by Bernoulli's law for gases:

$$m = a_1 \sqrt{2 p_e \rho_1} \tag{4.5}$$

The total pressure is the sum of the atmospheric and excess pressures:

$$p_1 = p_e + p_0 \tag{4.6}$$

Pulp Subsystem II + IV + V

Conservation law (conservation of volume)

$$\dot{V}_2 = Q - q \tag{4.7}$$

Constitutive relationships

$$V_2 = Ah \qquad (\text{ volume} = \text{area} \cdot \text{height}) \tag{4.8}$$

$$V_1 = V - V_2 \tag{4.9}$$

The flow q is determined by Bernoulli's law, as in Section 2.3. A complication is the excess pressure above the pulp. Converting this pressure into an effective pulp level, we get

$$h_{eff} = h + \frac{p_e}{\rho_2 g} \tag{4.10}$$

The flow out of the head box now becomes

$$q = a_2 \cdot C \cdot \sqrt{2 h_{eff} g} \tag{4.11}$$

The coefficient C compensates for the fact that Bernoulli's law is valid only for holes with small cross-sectional area and for flow without energy losses. We define an effective cross-sectional area $Ca_2(m^2)$, where a_2 is the geometric area.

4.5 Phase 3: Forming the State-space Model

After phase 2 we have, strictly speaking, a mathematical model for the system. The equations are often unstructured, however, and it is not easy to go from phase 2 to a simulation. In the example we treated in the last section, the result of phase 2 was 10 equations. Obviously these could be better organized. In this section we shall demonstrate how to go from phase 2 to a state space-model of the form (4.1).

The general recipe is obvious:

1. Choose a set of state variables.
2. Express the time derivative of each state variable with the help of state variables and inputs.
3. Express the outputs as functions of the state and input variables.

If steps 2 and 3 are successful, we have a state-space description as in (4.1). The difficulties in the procedure obviously lie in step 1. How do you choose the state variables? As a guidance it is useful to remember the interpretation of the state (see Section 3.4). The state variables represent the memory of what has happened earlier. All internal variables corresponding to storage of different quantities are therefore candidates for the role of state variable.

Example 4.1 Stored quantities
The following are examples of stored quantities:

- *Position of point mass (stored potential energy)*
- *Velocity of point mass (stored kinetic energy)*
- *Charge of capacitor (stored electrical field energy)*
- *Current through inductor (stored magnetic field energy)*
- *Temperature (stored thermal energy)*
- *Tank level (stored volume)*

□

Another rule of thumb is that internal variables, whose time derivatives occur in the equations of phase 2, are suitable state variables.

When forming the state-space model (4.1), it is not necessary to collect all formulas and equations together into one (vector-valued) equation. For large systems it is often advantageous to make separate state-space models for the subsystems and then connect them according to the block diagrams. A modularized model of this type makes it possible to check (in a simulation) what happens when a certain block is replaced by another.

An Example of Phase 3: The Head Box

Guided by the discussion in the previous section, we make the following choice of state variables:

$$\begin{aligned} x_1 &= V_2 \qquad \text{(volume of pulp)} \\ x_2 &= N \qquad \text{(mass of air)} \end{aligned}$$

We now have to express the time derivatives $\dot{x}_1$ and $\dot{x}_2$ using only x_1, x_2, M, Q, and the constants.

$$\begin{aligned} \dot{x}_1 = \dot{V}_2 = Q - q &= Q - a_2 C\sqrt{2hg + \frac{2p_e}{\rho_2}} \\ &= Q - a_2 C\left[\frac{2gV_2}{A} + \frac{2(RT\rho_1 - p_0)}{\rho_2}\right]^{1/2} \\ &= Q - a_2 C\left[\frac{2g}{A}x_1 + \frac{2}{\rho_2}\left(\frac{RTx_2}{V - x_1} - p_0\right)\right]^{1/2} \end{aligned} \tag{4.12}$$

Here we have used at the successive steps (4.7), (4.10), (4.11), (4.8) and (4.4), (4.6), (4.3) and (4.9), (4.8).

$$\begin{aligned} \dot{x}_2 = \dot{N} = M - m &= M - a_1\sqrt{2(p_1 - p_0)\rho_1} = \\ &= M - a_1\left[2\left(RT\frac{N}{V_1} - p_0\right)\frac{N}{V_1}\right]^{1/2} = \\ &= M - a_1\left[2\left(RT\frac{x_2}{V - x_1} - p_0\right)\frac{x_2}{V - x_1}\right]^{1/2} \end{aligned} \tag{4.13}$$

Here the following equations were used: (4.2), (4.5), (4.6) (4.3), (4.4) and (4.8), (4.9). The outputs are calculated in a similar fashion:

$$q = a_2 C \left[\frac{2g}{A} x_1 + \frac{2}{\rho_2} \left(\frac{RTx_2}{V - x_1} - p_0 \right) \right]^{1/2} \tag{4.14}$$

$$p_e = RT \frac{x_2}{V - x_1} - p_0 \tag{4.15}$$

$$h = \frac{x_1}{A} \tag{4.16}$$

The equations (4.12), (4.13) and (4.14)-(4.16) now form a state-space model for the head box, and we have reached the goal of the modeling.

The Number of State Variables

It is easy to check whether we have chosen enough state variables. The test is that their time derivatives can be expressed using only state variables and inputs to the system, together with constant quantities. It can be more difficult to determine if there are unnecessary state variables. In our head box example, N, p_1, V_2, and h are possible candidates for state variables. We see directly from (4.8) that it is unnecessary to let both V_2 and h be state variables. Apart from that, it is not easy, however, to see which variables are redundant. For linear state-space models there are tests based on the rank of certain matrices to determine if we have a minimal number of states. A corresponding test for nonlinear systems is much more difficult to carry out, both from a mathematical and a computational point of view. However, when the model is used for simulation, the only disadvantage (in principle) of too many state variables is that unnecessary computations are done. The simulation result is the same.

4.6 Simplified Models

All models contain simplifications of the real processes. We are forced to use simplifications simply because we do not know the exact relationships. But even if we knew them we would still construct approximate, simplified models. The reason is that the model must be manageable for our purposes. A model with thousands of variables

is impossible to use for analysis and requires long execution times for simulation. In other words, we seek simple models and we consciously introduce approximations and generalizations. By a simple model we mean primarily a model whose order (the dimension of the state vector) is small. Simple could also mean that the relationships between variables are easily computable or that the model is linear rather than nonlinear.

In this section we will study the principles that can be used to simplify models. The simplification can be done under the first two phases of modeling but also in the completed model to reduce complexity.

There is of course a trade-off between the complexity of a model and the accuracy requirements in order to simulate the physical system. The intended use of the model decides how this trade-off is handled. It is important to have some balance between the approximations used in different parts of the model, since usually the overall model is no better than the crudest approximation. If you accept a general approximation in one model, it maks no sense to resort to hair-splitting in another.

We will discuss simplifications of three kinds:

1. Small effects are neglected — approximate relationships are used.
2. Separation of time constants.
3. Aggregation of state variables.

Most model simplifications can be placed under one of these headings. We treat them one at a time.

Small Effects Are Neglected — Approximate Relationships Are Used

In the block diagram phase we usually consider what relationships could be present between the different variables. It is then often clear that certain effects are more important than others. When modeling the head box, we assumed the paper pulp to be incompressible. This ought to be a good approximation here, since the compressibility is much lower than that of the air pad. There are other situations however, for example, in hydraulic servos, where the compressibility of a liquid can give rise to important phenomena, like resonances that are clearly noticeable. In all modeling we must make trade-offs. Is

an effect important enough to be included in the block diagram? In different areas of modeling, different practices have developed from experience. Often we know when to neglect friction and air drag. For example, in hydrodynamics we know what approximations are suitable and how good they are. In general, we must use our physical intuition and insight together with the developed practice to arrive at a suitable degree of approximation.

When formulating the basic equations of phase 2, we often encounter similar problems. Many relationships between variables in an engineering system are complicated and do not fit the idealized situations for which physical laws are formulated. Real gases are not ideal, real liquids are not incompressible, real flow is not laminar, and so on. When modeling a nontechnical system, the difficulties of getting a reliable description are even greater.

In practice, we have to accept working with approximate relationships. The degree of approximation that we can tolerate depends on the desired accuracy of the model. For example, is it sufficient to use a linear relationship between force and elongation for a spring, or is it necessary to use a nonlinear description for large elongations? In the latter case we probably have to make separate experiments and tabulate the results. This is only justified if the model has to be very accurate or if its intended use is precise simulation of large amplitudes. As we noted earlier, we must also have some balance between approximations in different subsystems.

Separation of Time Constants

In Section 3.5 we introduced the notion of time constant to describe the time scale of change in the variables. The time constant for the car dynamics from accelerator to velocity is on the order of a few seconds, while the time constant of the economy (say from investment stimulants to export) can be several years.

In the same system we often have time constants of different orders of magnitude. Still, our interest might be focused on a certain time scale. If we are going to construct a model of a nuclear reactor to be used for simulations of the control rods, then we are mainly interested in time constants around a few seconds. The time constants associated with the burnout of nuclear fuel, which are on the order of months,

are then uninteresting even though they affect the dynamics.

We can give the following advice:

- Concentrate the modeling on phenomena whose time constants are of interest when considering the intended use of the model.
- Subsystems whose dynamics are considerably faster are approximated with static relationships. (See Section 3.5.)
- Variables of subsystems whose dynamics are appreciably slower are approximated as constants.

With these rules, we gain two important advantages

1. By ignoring very fast and very slow dynamics, we lower the order of the model.
2. By giving the model time constants that are on the same order of magnitude (say $T_{max}/T_{min} \leq 10 - 100$)), we get simpler simulations. (Differential equations that have a great spread of time constants are called *stiff*. We will see in Chapter 11 that they pose special problems for the simulation.)

For some systems we might really be interested in time constants of quite different magnitude. An example is the model of the heat storage for a solar-heated house. On the one hand, we want to see the variations in temperature during the day (time constant a few hours). On the other hand, we are interested in the yearly variation in temperature (time constant a couple of months). In such a situation we should consider the possibility of using two different models, one for each time scale, and using the preceding procedure to simplify each one of them.

Aggregation of State Variables

In Section 3.4 we defined the state as the set of information needed to predict the future behavior of the system, provided the external signals are known. A strict application of this definition would in most cases lead to an excessive number of states. For the head box example we would have to assign each point in the air pad a value of pressure, temperature and density. We would then have infinitely many state

variables (connected via partial differential equations). However, it seems reasonable that the spatial variations of these three variables are so small that one single value for each would be sufficient.

This is an example of *aggregation* of state variables:

> To merge several similar variables into *one* state variable. Often this variable plays the role of average or total value.

Aggregation is a very common method for reducing the number of state variables in a model. For economic models we often have a hierarchy of models with different amounts of aggregation. In a very simple model the level of investment might be one single variable (compare Section 2.4). In a less aggregated model the investments are perhaps divided into private and government ones. In detailed models the investment level of each sector of the economy might be modeled. There are economic models in use with more than a thousand state variables.

Aggregated models are also common in physics. A typical example is thermodynamics. To know the state of a volume of gas, we would strictly speaking have to know the speed and position of every molecule. Instead we use pressure and temperature when dealing with a gas on a macro level. Those variables are aggregated state variables connected to the average distance between molecules and the average velocity.

A number of physical phenomena are described by partial differential equations (PDE). Typical examples are field equations, waves, flow, and heat conduction problems. In mathematical models of dynamic systems to be used in simulation, PDEs are often unsuitable. The reason is that the standard simulation programs for dynamic systems assume that models are given in the form

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x,u) \end{aligned}$$

with a finite dimensional state vector x. Often partial differential equations are reduced to ordinary differential equations via difference approximations of the spatial variables. This corresponds to aggregation, as shown in the following example.

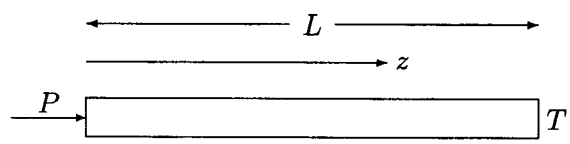


Figure 4.7: Heating of a metal rod

Example 4.2 Heat Conduction

Consider a metal rod whose left end point is heated by an external source. See Figure 4.7. The power in the heat source is denoted P and is an input. The system output is the temperature T at the other end point. The heat conduction is described by the heat equation

$$a\frac{\partial^2}{\partial z^2}x(z,t) = \frac{\partial}{\partial t}x(z,t) \tag{4.17}$$

where $x(z,t)$ is the temperature at time t at the distance z from the left end point. The number a is the heat conductivity coefficient of the metal. We have disregarded heat losses to the environment. At the end points we have

$$T(t) = x(L,t) \tag{4.18}$$

$$P(t) = \alpha\frac{\partial}{\partial z}x(z,t)|_{z=0} \tag{4.19}$$

where α is a constant depending on the heat transfer from the external source to the metal.

The description (4.17)-(4.19) requires that we know the whole function $x(z,t_1)$, $0 \leq z \leq L$, in order to determine the temperature $T(t)$ for $t \geq t_1$. The function $x(z,t_1), 0 \leq z \leq L$ is the system state at time t_1, so we have to measure and store infinitely many temperatures (one for each value of z) to know the state. Systems described by partial differential equations are therefore often called infinite-dimensional systems.

To get an approximate model that is more manageable for simulation purposes, we can use aggregation. Let us make a third-order model of the system in Figure 4.7. This means that we work with three

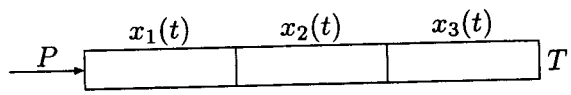


Figure 4.8: An aggregate model for heat conduction.

state variables. We divide the rod into three parts and assume the temperature to be homogeneous in each of them. See Figure 4.8. These temperatures at time t we denote $x_1(t)$, $x_2(t)$, and $x_3(t)$, respectively. In this way we have aggregated the function $x(z,t)$, $0 \leq z \leq L/3$, into the aggregate $x_1(t)$ and so on. Let the heat capacity of each part be denoted C and let the heat transfer coefficient between the parts be K. Writing the conservation of energy relationships for each part gives

$$\frac{d}{dt}(\text{heat stored in part 1}) = (\text{power in}) - (\text{power out to part 2})$$

This gives the equation

$$\frac{d}{dt}(C \cdot x_1(t)) = P - K(x_1(t) - x_2(t))$$

In an analogous manner we get

$$\frac{d}{dt}(C \cdot x_2(t)) = K(x_1(t) - x_2(t)) - K(x_2(t) - x_3(t))$$

$$\frac{d}{dt}(C \cdot x_3(t)) = K(x_2(t) - x_3(t))$$
$$T(t) = x_3(t)$$

Rearranging these equations gives the linear state-space model

$$\dot{x} = \frac{K}{C}\begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix} x + \frac{1}{C}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} P \tag{4.20}$$

$$y = (0 \quad 0 \quad 1)x$$

This is essentially the same model that we would get using a difference approximation of the partial derivative

$$\frac{\partial^2}{\partial z^2} x(z,t)$$

of (4.17). By choosing a finer division of the rod (giving more state variables), we could get a more accurate model. □

Example 4.3 Forrester's World Model
In 1979, J. W. Forrester, working for the Club of Rome, constructed a dynamic model for the development of the world. The model is of order five with the following state variables
x_1: *World population*
x_2: *Investments*
x_3: *Natural resources*
x_4: *Agricultural share of capital*
x_5: *Pollution*
This is probably the most grandiose example of aggregation. The model, its construction, and the result of simulations are described in J. W. Forrester: World Dynamics, Wright-Allen Press, Inc., 1971. □

4.7 Conclusions

We can now summarize the phases of modeling in the following scheme:

Phase 1. Structure the problem:

Decide on:

- The intended use of the model
- Outputs, inputs, variables, constants
- How the variables interact

Draw a block diagram.

Phase 2. Describe the relationships between the variables:

For each block:

- Write down conservation laws.
- Add relevant constitutive relations.

Phase 3. Form a state-space model:

- Choose a set of state variables.
- Express their time derivatives as functions of state variables and inputs.
- Express the outputs as functions of state variables and inputs.

We have shown how it is possible to produce a model in a fairly systematic fashion, starting from physical (or biological, chemical, or economical) knowledge of a system. One difficulty is that the modeling sometimes ranges over completely different types of physical systems. It is then useful to draw on the analogies that exist between different physical phenomena. We will discuss this aspect in the next two chapters. The general structuring of the modeling into three phases will also be relevant in that context.

Chapter 5

Some Basic Relationships in Physics

5.1 Introduction

When modeling physical systems, we must of course start with a knowledge of physics. In this chapter we shall summarize the most common relationships within a number of areas in physics. Together the equations presented here cover many of the situations encountered in physical system modeling. At the same time we get the advantage that more general relationships become visible. They can form the basis for the more general modeling methods presented in Chapter 6. To make comparisons easier, we will label the equations in a special way in this chapter. The label will be the section number followed by a letter, that indicates the type of equation. The labels (5.2C), (5.3C), (5.4C) etc. will all denote analogous equations, for instance.

5.2 Electrical Circuits

Consider electrical circuits consisting of resistors, capacitors, inductors, and transformers. The basic equations used to describe such circuits consist of relationships between the fundamental quantities:

Voltage u (volt) (5.2A)

Current i (ampere) (5.2B)

An ideal inductor, for instance, is described by

$$u(t) = L \cdot \frac{d}{dt} i(t)$$

where $u(t)$ and $i(t)$ are voltage and current at time t. The constant L is the *inductance* (henry), and the relationship is sometimes called the law of inductance. We can also write it as

$$i(t) = \frac{1}{L} \int_0^t u(s) ds \tag{5.2C}$$

In the same way an ideal capacitor is described by

$$i(t) = C \frac{d}{dt} u(t)$$

where C is the *capacitance* (farad). We can also write

$$u(t) = \frac{1}{C} \int_0^t i(s) ds \tag{5.2D}$$

For a linear resistor with *resistance* R (ohm) we have Ohm's law:

$$u(t) = Ri(t) \tag{5.2E}$$

We can of course also consider nonlinear resistances, with the general description

$$u(t) = h_1(i(t)) \tag{5.2F}$$

or

$$i(t) = h_2(u(t)) \tag{5.2F'}$$

for some nonlinear function h. An ideal rectifier, for instance, has

$$h_2(x) = \begin{cases} x, \, x > 0 \\ 0, \, x \leq 0 \end{cases}$$

In the resistor, energy is lost (as heat). The power is

$$P(t) = u(t) \cdot i(t) \tag{5.2G}$$

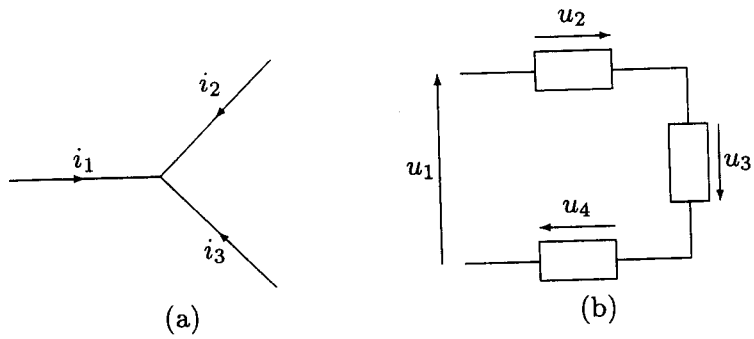


Figure 5.1: Kirchhoff's laws. (a) The sum of currents (with signs) is zero. (b) The sum of voltages (with signs) over a circuit is zero.

(P is measured in watts, 1 W = 1 J/s.) In a similar manner the inductor and the capacitor represent *energy storage* (magnetic and electric field energy, respectively). For the inductor we have

$$T(t) = \frac{1}{2}Li^2(t) \tag{5.2H}$$

(T is measured in joules), and for the capacitor

$$T(t) = \frac{1}{2}Cu^2(t) \tag{5.2I}$$

When connecting the electric circuit elements, the rule is

$$\sum_k i_k(t) \equiv 0 \tag{5.2J}$$

for current and

$$\sum_k u_k(t) \equiv 0 \tag{5.2K}$$

for voltage (Kirchhoff's laws). See Figure 5.1. An *ideal transformer* transforms voltage and current in such a way that their product is constant:

$$\begin{aligned} u_1 \cdot i_1 &= u_2 \cdot i_2 \\ u_1 &= \alpha u_2 \\ i_1 &= \tfrac{1}{\alpha} i_2 \end{aligned} \tag{5.2L}$$

See Figure 5.2.

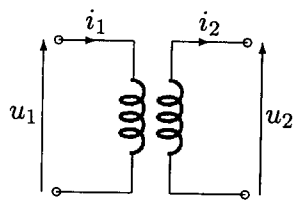


Figure 5.2: A transformer. α is the ratio of the number of turns on each side.

5.3 Mechanical Translation

Mechanical translation is governed by the laws of mechanics, which are relationships between the variables

Force F (newton) (5.3A)
Velocity v (meters per second) (5.3B)

Note that these quantities are three dimensional vectors in the general case. Most of the following relationships can be formulated as vector equations. For simplicity we will treat everything as scalars here.

Newton's force law gives

$$F(t) = m \cdot \frac{d}{dt} v(t)$$

where the constant m (kg) is the *mass* of the body. We can also write

$$v(t) = \frac{1}{m} \int_0^t F(s) ds \tag{5.3C}$$

For elastic bodies (for example, a linear spring), the force is proportional to the elongation (or compression). This in its turn is proportional to the integral of the difference in velocity between the end points:

$$F(t) = k \cdot \int_0^t v(s) ds \tag{5.3D}$$

Here k is the *spring constant* (N/m). In many cases there is a more complicated relationship between the force and the elongation (this is an important part of materials science). In general we can write

$$F(t) = \mathcal{H}(\int_0^t v(s)ds) \tag{5.3D'}$$

for some nonlinear function $\mathcal{H}$.

An important problem in mechanical systems is the description of the phenomena of *friction*. In general the description is a direct relationship between (frictional) force and velocity

$$F(t) = h(v(t)) \tag{5.3F}$$

The most common case perhaps is *dry friction*

$$h(x) = \begin{cases} +\mu \text{ if } x > 0 \\ F_0 \text{ if } x = 0 \\ -\mu \text{ if } x < 0 \end{cases}$$

(Here F_0 is the frictional force at rest, whose value depends on the details of the friction model.)

Air drag is often described by

$$h(x) = cx^2 \operatorname{sgn}(x)$$

while *viscous friction* (for example, in dampers) corresponds to

$$h(x) = \gamma x \tag{5.3E}$$

The power lost as heat through friction is

$$P(t) = F(t) \cdot v(t) \tag{5.3G}$$

In a similar way the velocity of the body and the compression of the spring represent *energy storage* (kinetic and elastic energy, respectively). For the kinetic energy we have

$$T(t) = \frac{1}{2}mv^2(t) \tag{5.3H}$$

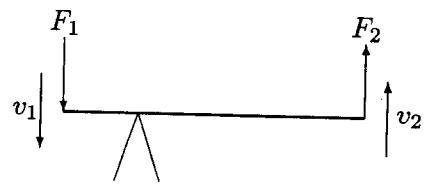


Figure 5.3: A lever. α is the ratio of the distances from the pivot.

and for the elastic energy of the linear spring

$$T(t) = \frac{1}{2k} F^2(t) \tag{5.3I}$$

When a number of forces act on a body at rest, their sum is zero:

$$\sum_k F_k(t) \equiv 0 \tag{5.3K}$$

Forces can be amplified by levers and similar mechanical devices. See figure 5.3. The relationship is then

$$\begin{aligned} F_1 v_1 &= F_2 v_2 \\ F_1 &= \alpha F_2 \\ v_1 &= \tfrac{1}{\alpha} v_2 \end{aligned} \tag{5.3L}$$

5.4 Mechanical Rotation

Mechanical systems with rotational parts like motor and gear boxes, are very common. For these systems the laws of mechanics relate the basic variables:

Torque M (newton - meter) (5.4A)

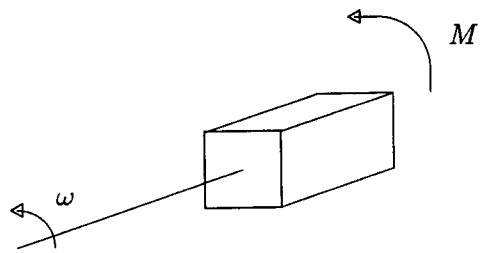


Figure 5.4: Rotational mechanics.

Angular velocity ω (radians per second) (5.4B)

See Figure 5.4. The counterpart of Newton's law of force says that the angular acceleration is proportional to the torque on the axis:

$$M(t) = J \cdot \frac{d}{dt}\omega(t)$$

where the constant of proportionality J (Nm/s^2) is the *moment of inertia.* We write this as

$$\omega(t) = \frac{1}{J}\int_0^t M(s)ds. \tag{5.4C}$$

The torsion of an axis gives rise to a torque described by

$$M(t) = k\int_0^t \omega(s)ds \tag{5.4D}$$

The integral corresponds to the angular displacement between the ends where k is the *torsional stiffness.*

The rotational friction is a function of the angular velocity

$$M(t) = h(\omega(t)) \tag{5.4F}$$

with different functions h analogous to translational friction. The power dissipation at rotation is

$$P(t) = M(t) \cdot \omega(t) \tag{5.4G}$$

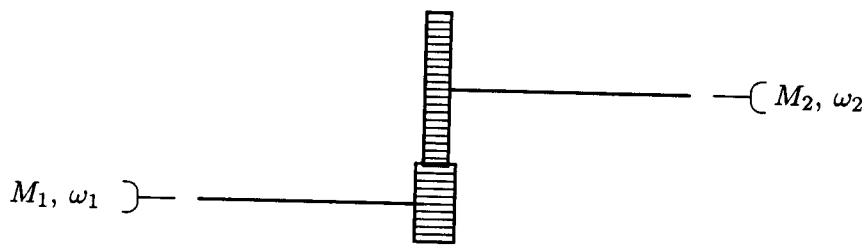


Figure 5.5: A pair of gears. α is the ratio of the circumferences.

and the stored rotational energy is

$$T(t) = \frac{1}{2} J\omega^2(t) \tag{5.4H}$$

while turning of a torsional axis to $M(t)$ corresponds to a stored torsional energy according to

$$T(t) = \frac{1}{2k} M^2(t) \tag{5.4I}$$

For a rotational mechanical system at rest, the sum of all torques must be zero:

$$\sum_k M_k(t) \equiv 0 \tag{5.4K}$$

A pair of gears transforms torque and angular velocity as follows:

$$\begin{aligned} M_1\omega_1 &= M_2\omega_2 \\ M_1 &= \alpha M_2 \\ \omega_1 &= \tfrac{1}{\alpha}\omega_2 \end{aligned} \tag{5.4L}$$

See Figure 5.5.

5.5 Flow Systems

By a flow system we mean connections of fluid flows in tubes and tanks. Typical applications are in chemical industrial systems and hydraulic systems. We will only treat *incompressible* fluids, that is,

those for which the volume is unaffected by the pressure. The treatment of compressible fluids is more complicated, partly because there are temperature changes when the volume is altered.

Flow systems are described by two basic quantities:

Pressure p (newtons per square meter) (5.5A)
Flow Q (cobic meters per second) (5.5B)

(We will work with volume flows. Multiplying by the density would give mass flows. For incompressible flows there is no essential difference.)

Consider a fluid flowing through a tube. See Figure 5.6. The pressure difference p between the end points of the tube results in a force that accelerates the fluid. If the cross-sectional area is A, the force is $p \cdot A$. The mass to be accelerated is $\rho \cdot \ell \cdot A$, where ρ is the density of the fluid and ℓ the length of the tube. Newton's force law gives

$$pA = \rho \cdot \ell \cdot A \cdot \frac{d}{dt} v(t)$$

where $v(t)$ is the velocity of the fluid. The velocity corresponds to a fluid flow $Q(t) = v(t) \cdot A$ (m^3/s). We get

$$p(t) = \frac{\rho \cdot \ell}{A} \cdot \frac{d}{dt} Q(t)$$

or

$$Q(t) = \frac{1}{L_f} \int_0^t p(s) ds \tag{5.5C}$$

where $L_f = \rho \cdot \ell / A$ is the *inertance* ($\mathrm{kg/m^4}$) of the tube.

Consider a fluid that is accumulated in a tank as shown in Figure 5.7. The volume in the tank is the integral of the flow: $V = \int Q\, dt$. The pressure at the bottom of the tank is equal to the level ($h = V/A$) multiplied by the density ρ and the gravitational acceleration g,

$$p(t) = \frac{\rho g}{A} \int_0^t Q(s) ds \tag{5.5D}$$

The number $C_f = A/\rho g$ ($\mathrm{m^4 s^2/kg}$) is called the *fluid capacitance.* If the area of the tank depends on the level, we get a nonlinear relationship

$$p(t) = \mathcal{H}(\int_0^t Q(s) ds) \tag{5.5D$'$}$$

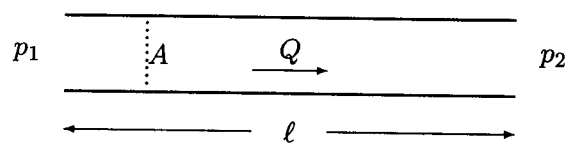


Figure 5.6: Flow through a tube. p_1 and p_2 are the pressures at the end points of the tube.

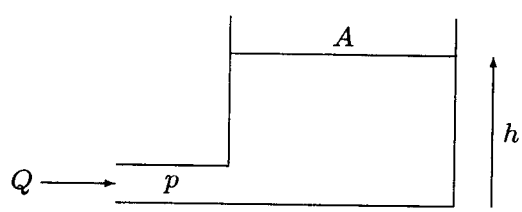


Figure 5.7: A tank as a fluid store.

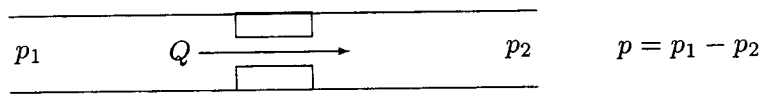


Figure 5.8: Flow through an orifice.

When liquid flows through a tube there is normally a loss of power through friction against the walls and internal friction in the fluid. This leads to a pressure drop over the tube. Conversely, we can say that a pressure drop is needed to maintain a certain flow. [Note that we disregarded these effects in (5.5C).] The pressure drop depends on the flow. In general we can write

$$p(t) = h_1(Q(t)) \tag{5.5F}$$

The properties of the function h_1 depend on the properties of the tube. If the tube is thin or filled with a porous medium, d'Arcy's law applies:

$$p(t) = R_f Q(t) \tag{5.5E}$$

where R_f is called the *flow resistance*. If the tube contains a sudden change in area (an orifice or a valve), we have the approximate relationship

$$p(t) = \mathcal{H} \cdot Q^2(t) \cdot \operatorname{sgn}(Q(t)) \tag{5.5F$'$}$$

for some constant $\mathcal{H}$. See Figure 5.8. The energy loss through these phenomena is

$$P(t) = p(t) \cdot Q(t) \tag{5.5G}$$

The frictionless flow through a tube (Figure 5.6) corresponds in the same way to an accumulation of energy,

$$T(t) = \frac{1}{2} L_f Q^2(t) \tag{5.5H}$$

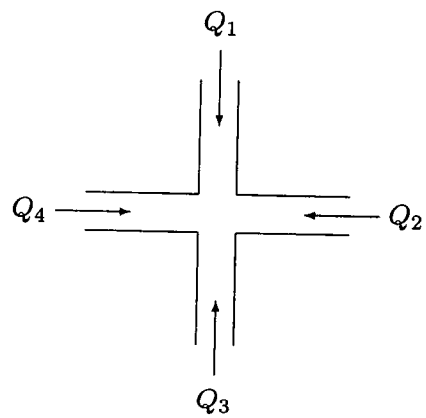


Figure 5.9: Flows at a junction.

while the tank of Figure 5.7 corresponds to potential energy:

$$T(t) = \frac{1}{2} C_f p^2(t) \tag{5.5I}$$

When flows are connected in a junction (see Figure 5.9) their sum must be zero:

$$\sum_k Q_k(t) \equiv 0 \tag{5.5J}$$

Likewise, the total pressure over a series connection as in Figure 5.10 must be the sum of the pressure drops

$$p_{r+1} = \sum_{k=1}^{r} p_k$$

or, going around in a loop and taking account of signs,

$$\sum_{k=1}^{r+1} p_k = 0 \tag{5.5K}$$

Finally, flows and pressures can be transformed as shown in Figure

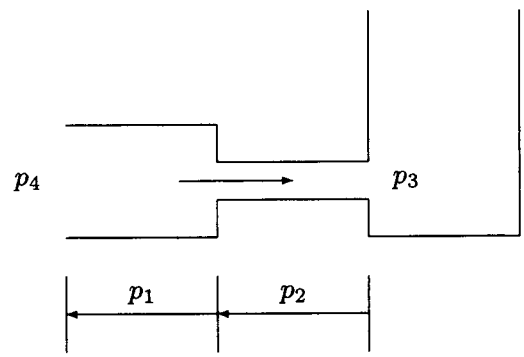


Figure 5.10: Pressures over subsystems [$r = 3$ in (5.5K)].

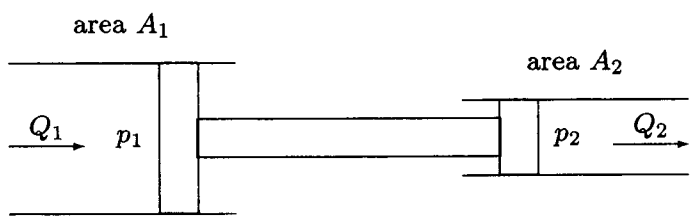


Figure 5.11: A flow transformer.

5.11. It is easily seen that we get

$$\begin{aligned} p_1 Q_1 &= p_2 Q_2 \\ p_1 &= \alpha p_2 \\ Q_1 &= \tfrac{1}{\alpha} Q_2 \end{aligned} \tag{5.5L}$$

where $\alpha = A_2/A_1$.

5.6 Thermal Systems

Thermal systems involve heating of objects and transport of thermal energy. The laws governing these phenomena are typically expressed

as relationships between the quantities:

Temperature T (kelvin) (5.6A)
Heat flow rate q (watt). (5.6B)

Heating of a body means that the temperature increases as heat flows into it.

$$q(t) = C\frac{d}{dt}T(t)$$

Here C is the *thermal capacity* [J/(K· s)]. We write this relationship as

$$T(t) = \frac{1}{C}\int_0^t q(s)\,ds \tag{5.6D}$$

If the thermal capacity depends on the temperature, then (5.6D) is replaced by a nonlinear expression. Normally, $T(t)$ represents the deviation from a reference temperature (different from absolute zero) and then (5.6D) can be a good approximation.

Heat transfer between two bodies with different temperatures is often considered to be proportional to the temperature difference $T(t)$:

$$q(t) = WT(t) \tag{5.6E}$$

The coefficient W is called the *heat transfer coefficient* [J/(K· s)].

Furthermore, the sum of all heat flow rates at one point must be zero:

$$\sum q_k(t) \equiv 0 \tag{5.6J}$$

5.7 Some Observations

There are obvious similarities among the basic equations for different physical systems. (Check all the equations ending with the same letter!) We have seen that almost all equations are relationships between two variables:

A: Effort variables e
B: Flow variables f

The relationships have the following characteristics:

C: Effort storage: $f = \alpha^{-1} \cdot \int e$

System	Effort	Flow	C	D	F
Electrical	Voltage	Current	Inductor	Capacitor	Res
Mechanical:					
Translational	Force	Velocity	Body	Spring	Fri
Rotational	Torque	Angular velocity	Axis	Torsion spring	Fri
Hydraulic	Pressure	Flow	Tube	Tank	Or
Thermal	Temperature	Heat flow rate	-	Heating	He

Table 5.1: Some Physical Analogies.

D: Flow storage: $e = \beta^{-1} \int f$
F: Static relationship: $e = h(f)$
G: Power dissipation: $P = e \cdot f$
H: Energy storage via C: $T = \frac{1}{2\alpha} f^2$
I: Energy storage via D: $T = \frac{1}{2\beta} e^2$
J: Sum of flows equal to zero: $\sum f_i = 0$
K: Sum of efforts (with signs) equal to zero: $\sum e_i = 0$
L: Transformation of variables: $e_1 f_1 = e_2 f_2$

In this way we get many analogies. In certain cases the analogies are not complete. The relationship G, for instance, is invalid for thermal systems. We would have to use entropy flow rate instead of heat flow rate to get completely parallel results. The analogies are summarized in Table 5.1.

5.8 Conclusions

We have seen that there are far reaching analogies between different types of physical systems. An important aspect is that it should be possible to create systematic, application-independent modeling methods starting from these analogies. In Chapter 6 we will use the analogies in a thoroughly systematic attack on the modeling problem.

Chapter 6

Bond Graphs

In Chapter 5 we saw that there are far reaching analogies between electrical, mechanical, hydraulic, and thermal systems. At the end it was also suggested that we can find a systematic modeling scheme based on these analogies. One way of doing this is to use bond graphs, introduced by H. Paynter. We then make the modeling systematic by following the energy flow in the system under study. Since the processes that are modeled usually involve energy exchange and energy is conserved, there is some assurance that nothing important is forgotten in the modeling.

6.1 Efforts and Flows

The basic idea of bond graphs is the observation of Section 5.7 that many models can be described in terms of effort variables e and flow variables f. This is represented graphically in Figure 6.1a for the electrical, mechanical, hydraulic, and general cases, respectively. As we noted before, the products ui, Fv, pQ, and ef have the dimension of power. The horizontal line, the *bond*, is therefore interpreted as a connection between subsystems that exchange energy. To mark in what direction the energy flows when the product ef is positive, we can put a half-arrow on the bond (Figure 6.1b). (Ordinary arrows are reserved for other purposes; we will return to them later.) The half-arrow will also be used to distinguish between the variables: it will be written on the same side as the effort variable. (Unfortunately, no

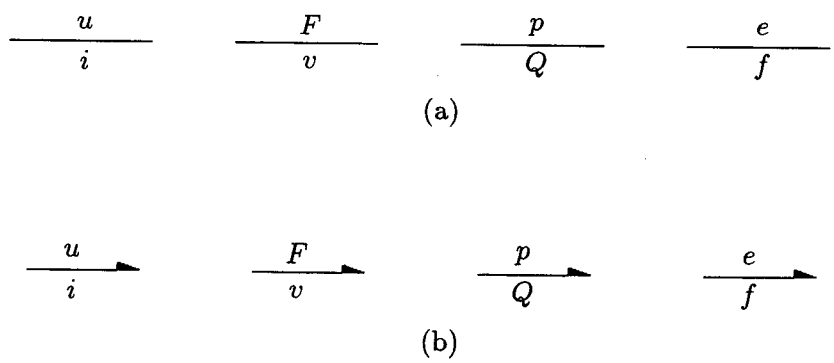


Figure 6.1: Efforts and flows.

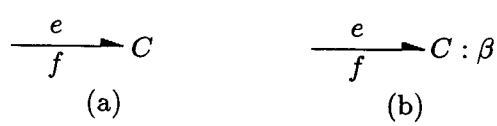


Figure 6.2: Flow storage.

convention of this type is universal in the bond graph literature.) We will now present some elements that can be connected by bonds.

Flow Storage

We noted in Chapter 5 that many physical elements can be modeled as flow storage, for example, a capacitor for electric current or a tank for water flow. With bond graphs this is represented according to Figure 6.2a. Here the C denotes the storage element (C as in capacitive). The bond symbolizes the energy flow from the rest of the system into

the element. If the relationship between e and f is linear,

$$e(t) = \frac{1}{\beta}\int^t f(\tau)d\tau \quad \text{or} \quad \frac{de}{dt} = \frac{f}{\beta}$$

then the coefficient β is often included in the graph as in Figure 6.2b.

Effort Storage

We also saw in Chapter 5 that, for example, an inductor could be regarded as an effort storage with a relationship

$$f(t) = \frac{1}{\alpha}\int^t e(\tau)d\tau \tag{6.1}$$

(in the linear case). In a bond graph this is represented as in Figure 6.3. Elements of this type are often called inductive or inertial (compare the electrical and mechanical applications in Sections 5.2 and 5.3). The symbol $I : \alpha$ shows the type of element (I) and the parameter α of equation (6.1).

Resistive Elements

We have also considered static relations between effort and flow

$$e(t) = h(f(t)), \quad f(t) = h^{-1}(e(t))$$

These are called resistive elements following electrical terminology. The graphical representation is shown in Figure 6.4a. For linear elements with the relation

$$e = \gamma \cdot f$$

we often use the notation of Figure 6.4 b.

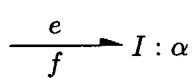


Figure 6.3: Effort storage.

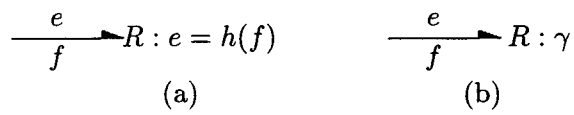


Figure 6.4: Resistive element.

Figure 6.5: (a) Effort source and (b) flow source.

Sources

In the general discussion of models (Section 3.2), we saw that there are in general inputs, that is, externally generated signals. When dealing with bond graphs, they are called *sources*. They can be of two types, depending on whether the input is an effort or a flow (see Figure 6.5).

In the effort source, e is regarded as input, while f is the input for the flow source.

6.2 Junctions

The discussion in Chapter 5 of analogies between different physical phenomena also showed that summation laws for flows and efforts are important. (Equations J and K in Section 5.7.) We will now scrutinize these relationships within the framework of bond graphs.

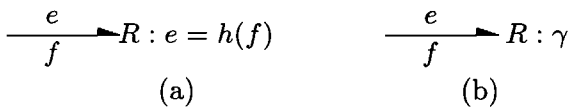


Figure 6.4: Resistive element.

Figure 6.5: (a) Effort source and (b) flow source.

Sources

In the general discussion of models (Section 3.2), we saw that there are in general inputs, that is, externally generated signals. When dealing with bond graphs, they are called *sources.* They can be of two types, depending on whether the input is an effort or a flow (see Figure 6.5).

In the effort source, e is regarded as input, while f is the input for the flow source.

6.2 Junctions

The discussion in Chapter 5 of analogies between different physical phenomena also showed that summation laws for flows and efforts are important. (Equations J and K in Section 5.7.) We will now scrutinize these relationships within the framework of bond graphs.

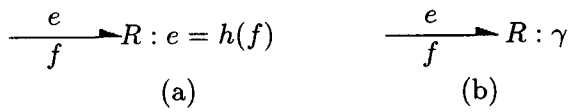


Figure 6.4: Resistive element.

Figure 6.5: (a) Effort source and (b) flow source.

Sources

In the general discussion of models (Section 3.2), we saw that there are in general inputs, that is, externally generated signals. When dealing with bond graphs, they are called *sources*. They can be of two types, depending on whether the input is an effort or a flow (see Figure 6.5).

In the effort source, e is regarded as input, while f is the input for the flow source.

6.2 Junctions

The discussion in Chapter 5 of analogies between different physical phenomena also showed that summation laws for flows and efforts are important. (Equations J and K in Section 5.7.) We will now scrutinize these relationships within the framework of bond graphs.

Series Junction

Consider the electric circuit of Figure 6.6a. Let us regard the voltage v as an input signal. We then get an effort source, (see Figure 6.6b). We see directly that the energy flowing into the system is divided between the resistor, the capacitor, and the inductor (see Figure 6.6c).

We have put an s into the junction representing the division of the energy flow to show that it represents a series connection. Note that the flow variable i is the same for all bonds at the junction. Also note that the efforts satisfy

$$v - v_1 - v_2 - v_3 = 0$$

The sum of the efforts is thus zero, if we use the convention that an outward pointing half-arrow represents a change of sign for the corresponding variable. By using what we know about capacitive, inductive, and resistive elements, the graph can now be completed (see Figure 6.6d).

We can now generalize the example and define a general *series junction*, an s junction (see Figure 6.7a). It is characterized by the following.

- The same flow: $f_1 = f_2 = ... = f_n$
- The sum of efforts equal to zero: $e_1 + e_2 + ... + e_n = 0$

As we saw in the example, it is natural to have the convention that an outward pointing arrow gives a change of sign for the effort variable. The graph of Figure 6.7b, for example, has the equation

$$e_1 + e_2 - e_3 = 0$$

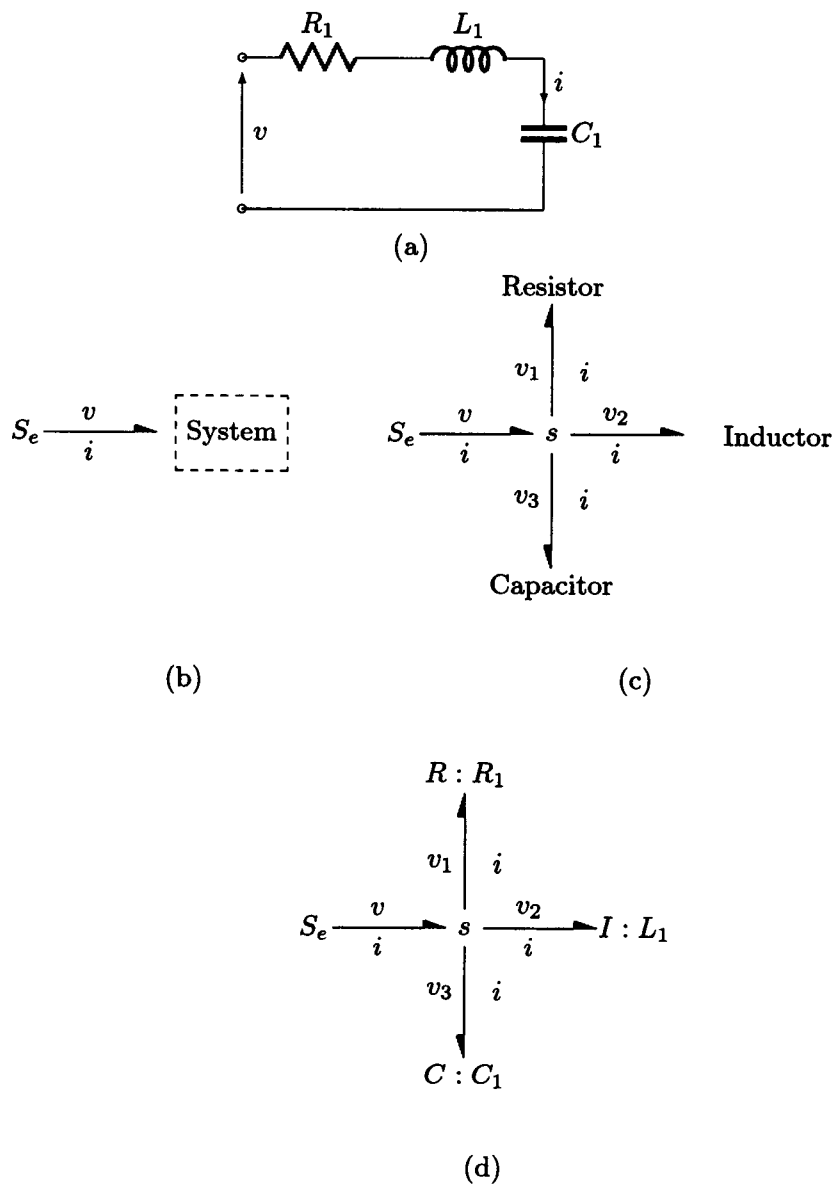


Figure 6.6: Power flow in an electric circuit.

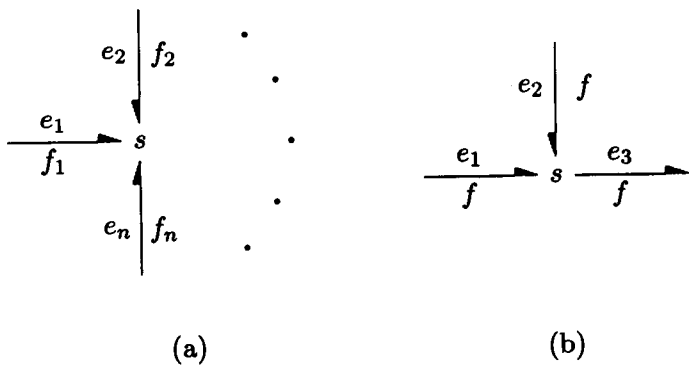


Figure 6.7: s junction.

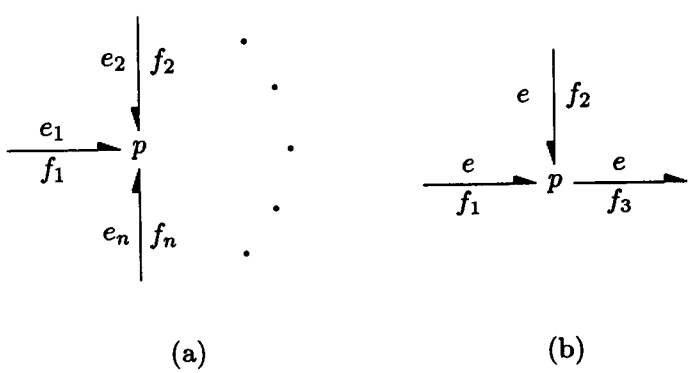
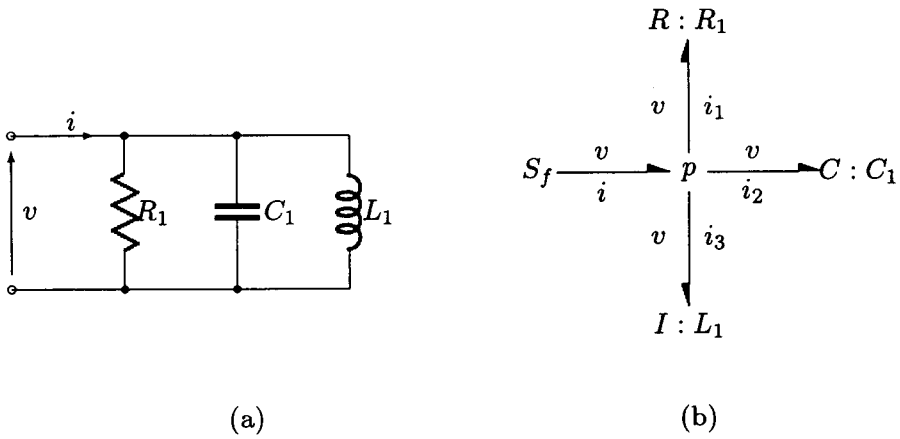


Figure 6.8: p junction.



(a) (b)

Figure 6.9: Electrical parallel circuit.

Parallel Junction

Analogously to the previous section, it is natural to define a junction where e and f have changed roles. So let us define a *parallel junction* (p junction) with the graphical representation shown in Figure 6.8a and the following rules:

- Identical efforts: $e_1 = e_2 = ... = e_n$
- Sum of flows equal zero: $f_1 + f_2 + ... + f_n = 0$

As before the sign convention should be that an outward pointing half-arrow gives a change of sign in the sum, so in Figure 6.8b the relation is $f_1 + f_2 - f_3 = 0$. Let us finally motivate the name parallel junction by applying the definition to the electrical parallel circuit shown in Figure 6.9a. We get the graph in Figure 6.9b. We will see later that the source should be a flow source (Section 6.6). We see

Figure 6.10: Removal of junctions.

that the equation

$$i - i_1 - i_2 - i_3 = 0$$

fits into what we know about electrical currents. Also there is the same voltage (effort) at all the bonds.

Simplifications in Bond Graphs

In a number of cases, bond graphs can be simplified. We give two examples.

1. Junctions that have only two bonds (with the half arrows pointing in the same direction) can be removed, as shown in Figure 6.10.

2. A direct application of the definitions shows that two adjacent junctions of the same kind can be merged. In Figure 6.11, two s junctions (a and b) and two p junctions (c and d) are merged.

6.3 Simple Bond Graphs

We now have enough tools to set up bond graphs for some simple physical systems. Consider the mechanical system of Figure 6.12a. A body, which is moving without friction with the velocity v, is pulled with the force F. It is natural to regard F as the input, that is, an effort source (see Figure 6.12b). In this case, all energy is stored as kinetic energy. From Section 5.3 we see that this is an effort storage, that is, an I element. The graph is shown in Figure 6.12c.

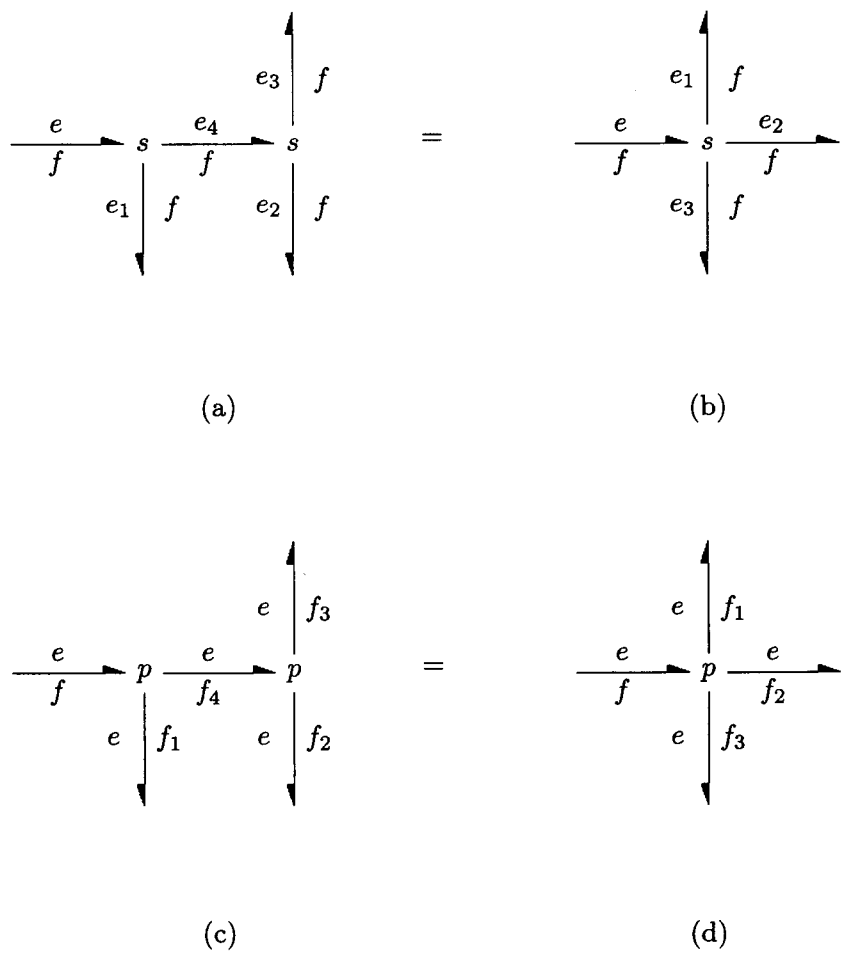


Figure 6.11: Merging of junctions.

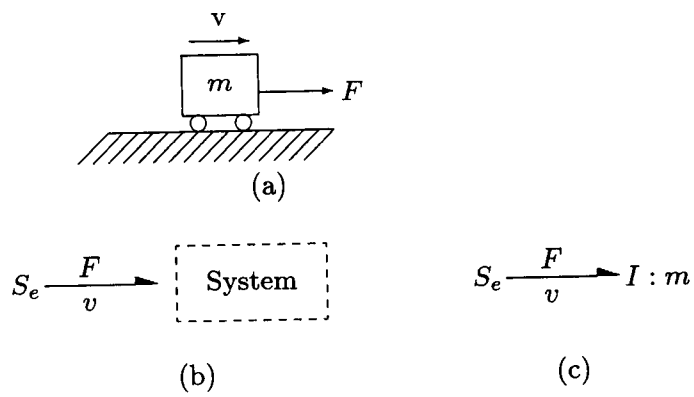


Figure 6.12: Mechanical system.

Now suppose that we add a spring with spring constant k, as shown in Figure 6.13a. F is still represented as an effort source. Now, however, the energy is divided between kinetic energy and energy stored in the spring (see Figure 6.13b). This corresponds to effort and flow storage, respectively, that is, I and C elements. Since these elements have the velocity in common, this is a junction with common flow — a series junction. This gives the graph of Figure 6.13c with the relationship $F = F_1 + F_2$.

Now consider the situation in which the mass slides along the surface with a certain friction as shown in Figure 6.14a. The friction is assumed to be a nonlinear function of the velocity, $F_f = \varphi(v)$. The energy supplied by F is now divided into three components (Figure 6.14b). We see that the friction loss is an R element. Furthermore, the velocity v is common to all bonds, that is, we still have an s junction. The graph is shown in Figure 6.14b.

We will now consider the electrical system shown in Figure 6.15a. The voltage v is regarded as input. This corresponds to an effort source (see Figure 6.15b). If we regard R_2 and C_1 as one element, we see that we have a series connection where R_1 and L_1 are R and I elements respectively (Figure 6.15c). In the subsystem consisting of R_2 and C_1

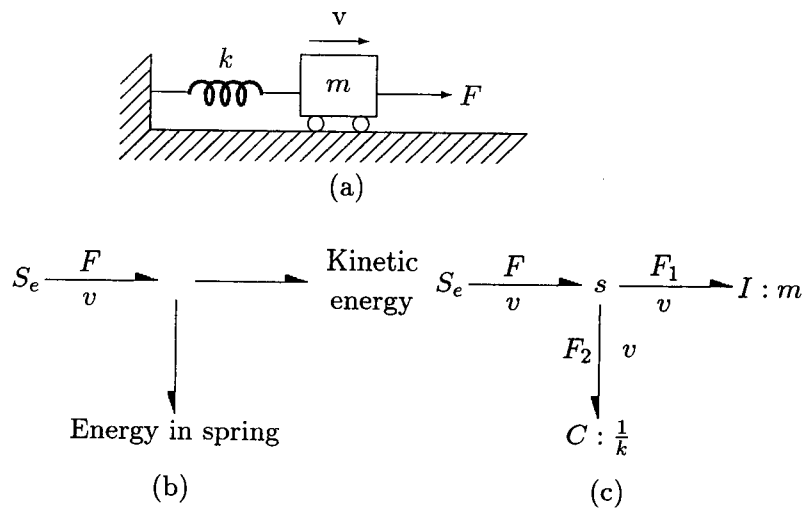


Figure 6.13: Mechanical system with spring.

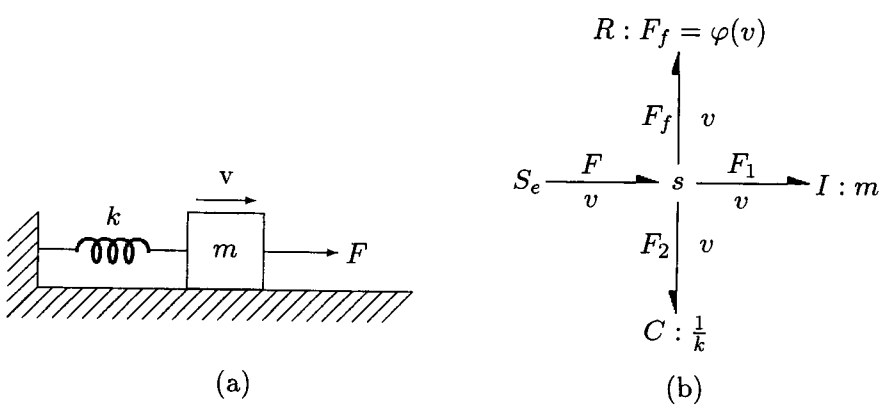


Figure 6.14: Mechanical system with spring and friction.

the energy is divided between R_2 (R element) and C_1 (C element) in a p junction (a common voltage). The full graph is displayed in Figure 6.15d.

6.4 Transformers and Gyrators

In Chapter 5 it was shown that the ideal electrical *transformer* (Figure 6.16a) with the relationships

$$u_2 = nu_1, \quad i_2 = \frac{1}{n} i_1$$

also has counterparts for mechanical and hydraulic systems. In bond graphs it is usually symbolized as shown in Figure 6.16b, with the relationships

$$e_2 = ne_1$$
$$f_2 = \tfrac{1}{n} f_1$$

An interesting feature of the transformer is that it can also be used to describe connections between different types of physical variables. An example is a hydraulic cylinder as shown in Figure 6.17a. The relationship between the mechanical variables, the force F and the velocity v, and the hydraulic variables, the pressure p and the flow Q, is

$$p = F/A$$
$$Q = vA$$

where A is the cross section of the cylinder. Since we have defined p and F to be effort variables and Q and v to be flow variables, this is a transformer (Figure 6.17b).

There is an element similar to the transformer in which efforts and flows have a "crosswise" dependence. It is called a *gyrator* and its symbolic description is shown in Figure 6.18. It is characterized by the equations

$$e_2 = rf_1$$
$$f_2 = \tfrac{1}{r} e_1$$

Note that $e_2 f_2 = e_1 f_1$, so power is conserved, as in the transformer.

The gyrator is found for instance in the conversion between electrical and rotational energy in electric motors. In principle the situation is described by Figure 6.19a. A wire carrying the current i rotates with

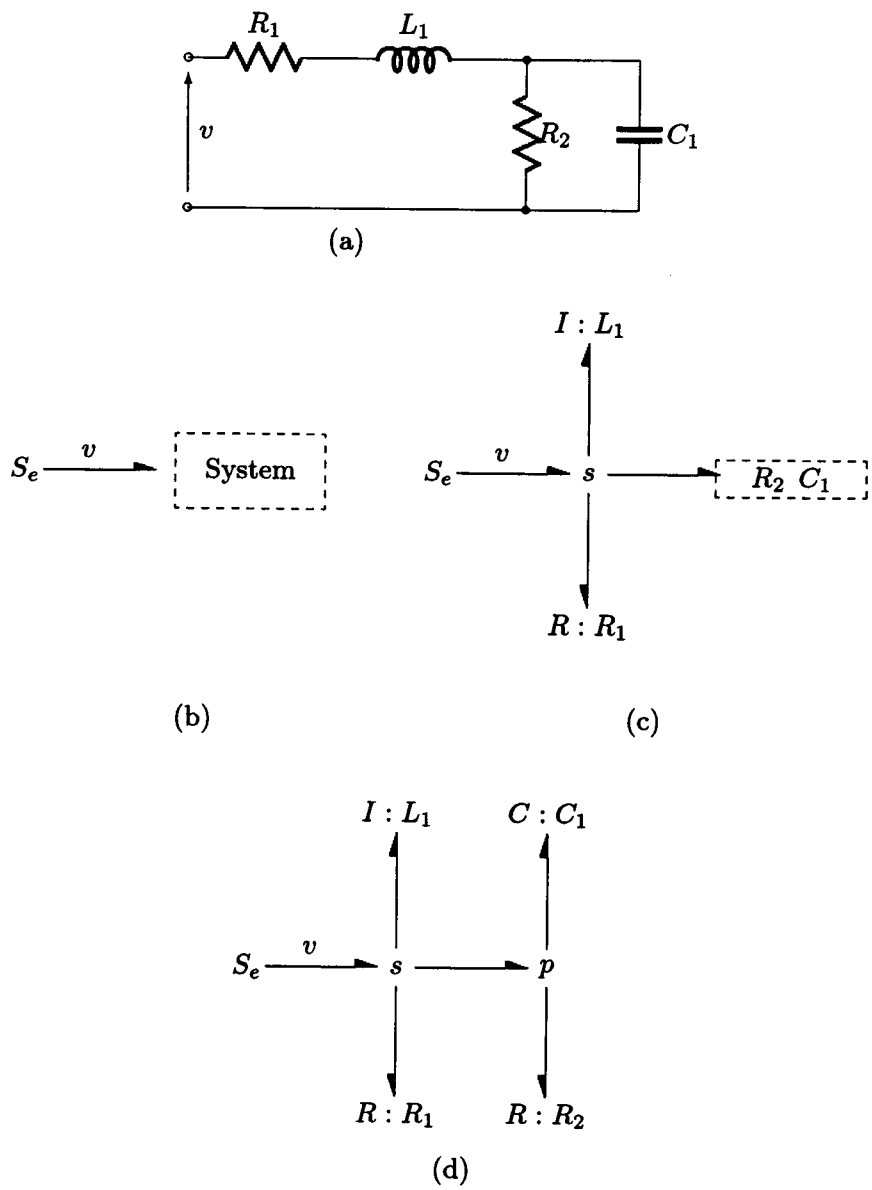


Figure 6.15: Electrical system.

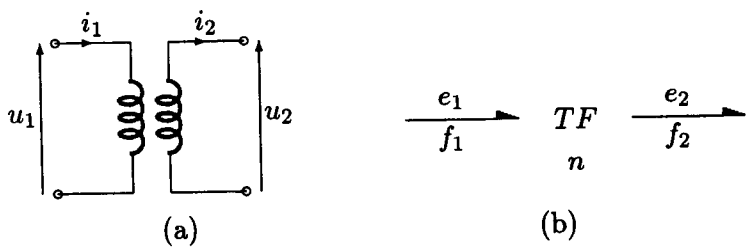


Figure 6.16: Ideal transformer.

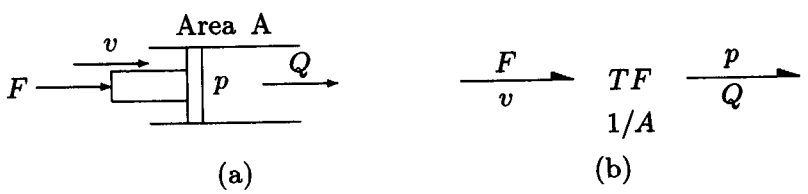


Figure 6.17: A mechanical-hydraulic conversion.

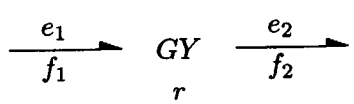


Figure 6.18: Gyrator.

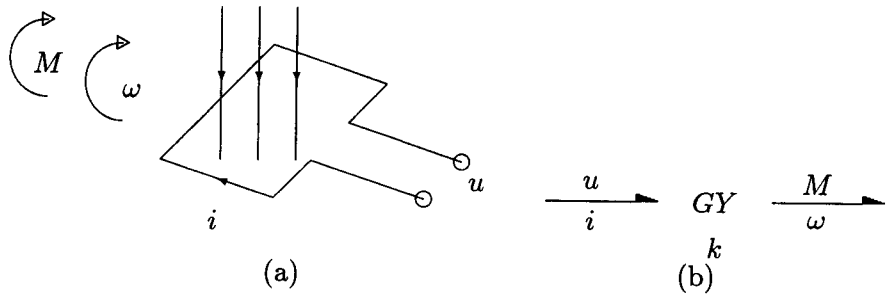


Figure 6.19: An electromechanical example of a gyrator.

the angular velocity ω in a magnetic field. It is turned by a moment M proportional to the current

$$M = ki$$

At the same time an emf proportional to the angular velocity is generated

$$u = \tilde{k}\omega$$

Since $M\omega = ui$ if there are no energy losses, we must have $\tilde{k} = k$. We can thus symbolize the system with a gyrator as shown in Figure 6.19b.

6.5 Systems with Mixed Physical Variables

We shall use tranformers and gyrators as tools to describe some mixed systems.

Example 6.1 A DC Motor
We first consider the dc motor shown in Figure 6.20a. The motor is controlled by an external voltage u, giving rise to a current i and a rotation with the angular velocity ω. We assume that the windings have a resistance R_1 and an inductance L_1. The rotating parts have a moment of inertia J. There is also some friction. The energy supplied will be partially stored in the inductor and partially lost in the resistor, while the remainder will be converted into mechanical energy. This energy in its turn will be stored as rotational energy or be lost by

friction. The graph will then in principle be the one of Figure 6.20b. As we saw earlier, the storages in the inductor and as rotational energy correspond to I elements, while the resistor and the friction are R elements. Both junctions have common flow variables, so they are s junctions. The conversion between electrical and mechanical energy is a gyrator as we saw in the last section. We then get the bond graph of Figure 6.20c, where we have assumed a general, possibly nonlinear, relationship for the friction: $M_f = \varphi(\omega)$. □

Example 6.2 Mechanics-Hydraulics
Let us now consider a mechanical-hydraulic example as shown in Figure 6.21a. The force F *acts on a piston with mass* m_1 *and cross section* A_1. *This results in a flow* Q_1, *which is partly stored in a tank with cross section* A_3 *and partly goes on as* Q_2 *to a piston with cross section* A_2. *This piston moves the mass* m_2 *with velocity* v_2 *over a surface with friction* $F_f = \varphi(v_2)$. *The pressure in the connecting tube is* p. *Starting with the left cylinder, we see that the energy is partly stored in the mass* m_1 *and partly transferred to the hydraulic subsystem. This transfer of energy is, as we have seen, represented by a transformer (Figure 6.17). We thus get Figure 6.21b.*

The energy flow to the right is partly stored in the tank (see equation (5.5D)) and partly taken to the cylinder on the right. A common pressure means a p junction [see Figure 6.21c]. At the cylinder to the right there is a reconversion to mechanical energy which is divided into kinetic energy and friction losses. We then get the full graph shown in Figure 6.21d. □

6.6 Causality: Signals between Subsystems

We have discussed input signals and represented them by sources. If an effort e is an input, then we have an effort source (see Figure 6.22). (We do not include the half-arrow in the bond, since the sign of the energy flow is of no importance here.) Our physical feeling tells us that the system will produce a certain f when we apply an e. (A certain voltage gives a certain current, a certain pressure gives a certain flow, and so on.) The flow f should then be regarded as the output of the

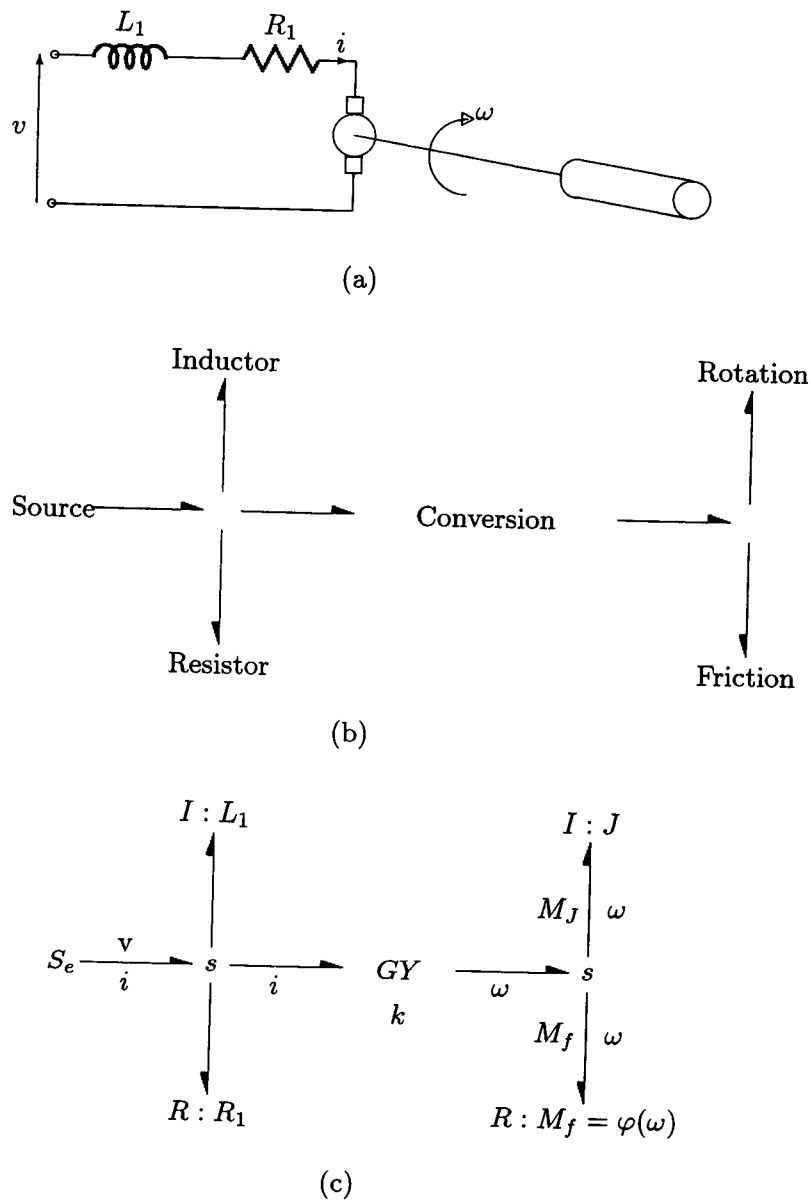


Figure 6.20: DC motor.

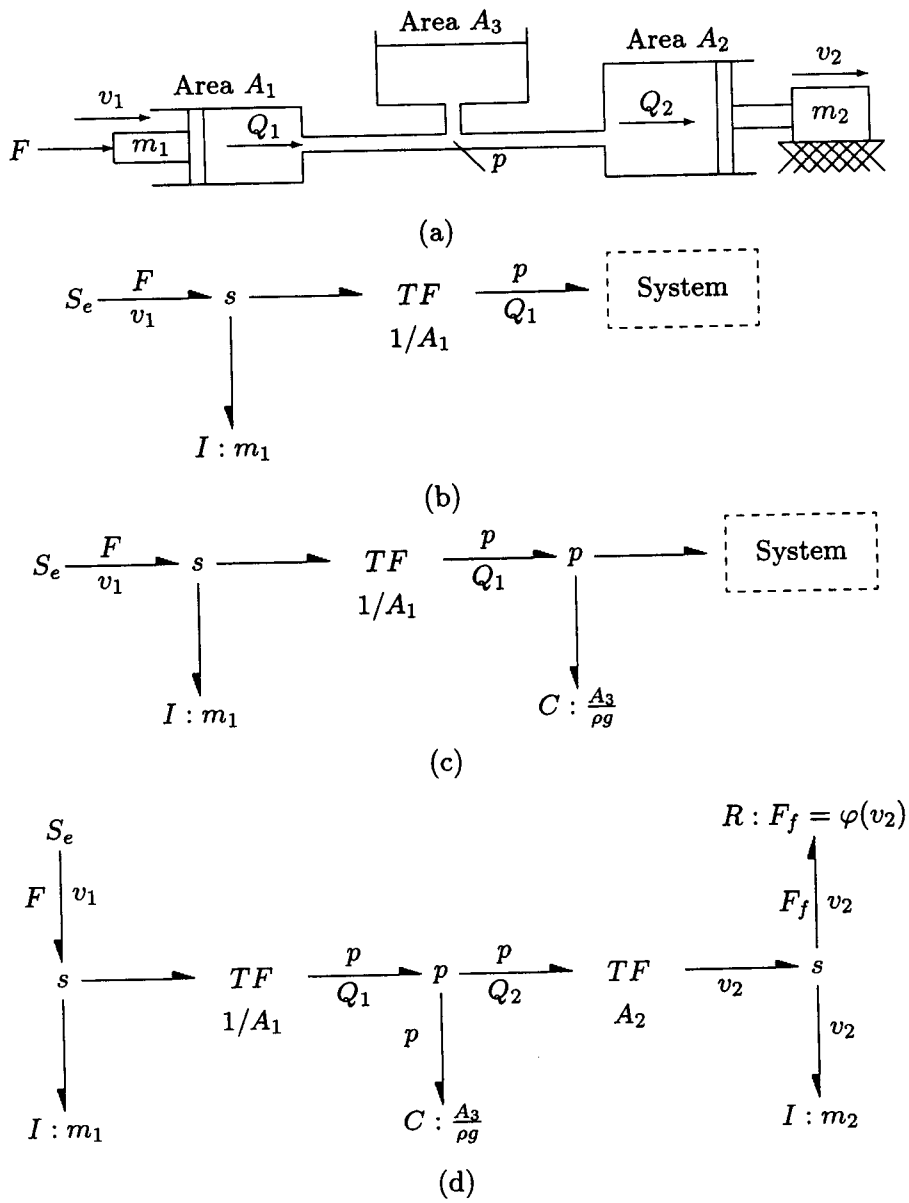


Figure 6.21: Mechanical-hydraulic system.

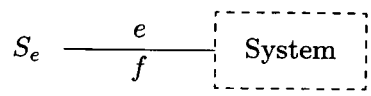


Figure 6.22: Effort as input.

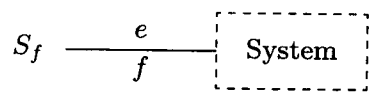


Figure 6.23: Flow as input.

system. For a flow source we have the reverse situation (see Figure 6.23). We regard f as input to the system and it is then reasonable to regard e as the output (a current gives rise to a voltage drop, a flow gives a pressure drop).

The point is now that we can use the same arguments about any bond. Consider a bond between two arbitrary subsystems (Figure 6.24). Now imagine a situation where it is natural to regard the subsystem A as an effort source. The variable e then becomes an input to the subsystem B, and from the preceding argument f should be an output of B. To make the situation symmetrical it is then logical to

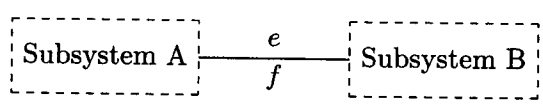


Figure 6.24: A bond between subsystems.

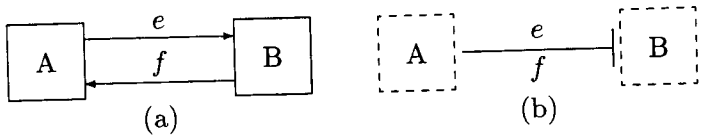


Figure 6.25: Signals between subsystems.

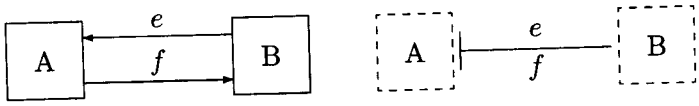


Figure 6.26: Signals between subsystems.

regard f as an input to A, so e becomes an output. Using a block diagram description (see Chapter 4), we can illustrate the situation as shown in Figure 6.25(a). We regard e as a signal from A to B and f as a signal from B to A. If we want to view the bonds in this way, we put a cross stroke on the bond adjacent to subsystem B, as shown in Figure 6.25(b), that is, adjacent to the subsystem that has effort as input. A cross stroke at the opposite side of the bond, as shown in Figure 6.26, means that e is regarded as input of A and f as input of B. This process of assigning inputs and outputs to subsystems is often called *causality* assignment – the input can be regarded as the cause of the output. The cross strokes are often called *causal strokes*.

Note that the causal stroke says that we *choose* to regard the system as having a certain causality. The causality is thus not given by physics (except for sources). We will see however that there is often a natural choice of causality that makes it simple to transfer a bond graph to state equations.

Causality and Sources

For sources we must, by definition, have the causality of Figure 6.27.

Figure 6.27: Causality of sources.

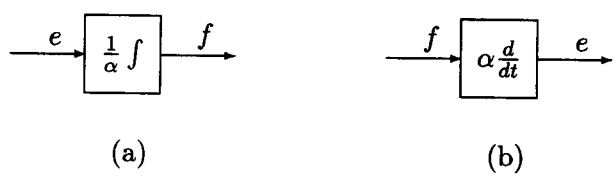


Figure 6.28: Different choices of input for an effort storage.

Causality of Energy Storage

Let us consider an effort storage described by

$$f(t) = \frac{1}{\alpha}\int^t e(\tau)d\tau$$

Regarding e as the input gives the block diagram of Figure 6.28a. Choosing f as the input gives instead the block diagram of Figure 6.28b. Since integration is a more natural operation than differentiation in signal processing, it seems natural to prefer Figure 6.28a. We could also consider a physical example. Suppose a body with velocity f is acted on by the force e. If f is regarded as the input, an instantaneous input change will give an infinitely large force. There is no corresponding problem if the force e is regarded as input. Every time

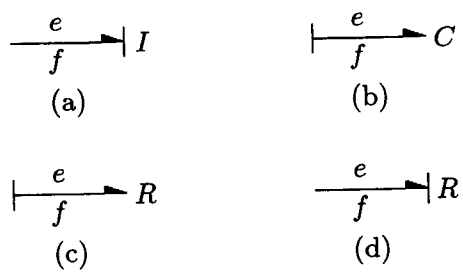


Figure 6.29: Causal strokes for I, C and R elements.

function e (with the exception of mathematical pathologies) can be integrated to a nice f. We conclude that it is natural to let e be the input and then get the causality of Figure 6.29a. Since e and f change roles for a flow storage, the natural choice of input is f, giving the causality of Figure 6.29b.

Causality of Resistive Elements

For a resistive element there is a static relationship between e and f; for example,

$$e(t) = \varphi(f(t)), \quad f(t) = \varphi^{-1}(e(t))$$

where we have assumed the function φ to be invertible. If this is the case we see that we are free to choose either e or f as input. Both (c) and (d) of Figure 6.29 are thus possible causalities. We will see later that this freedom can be used to adjust the causality to be compatible with the rest of the graph.

Causality of Junctions

For an S junction as shown in Figure 6.30a, we know that there is a common flow and that the sum of the efforts is zero:

$$f_1 = f_2 = ... = f_n, \quad e_1 + e_2 + ... + e_n = 0 \tag{6.2}$$

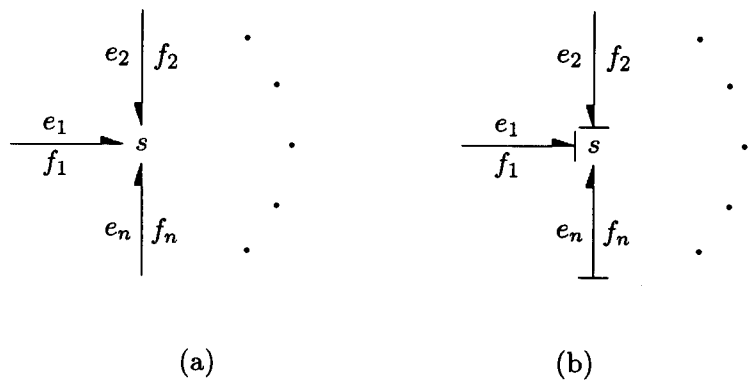


Figure 6.30: Causality of an s junction.

We see that it is possible to choose the flow in only one of the bonds; when this f_i is chosen, the others are automatically fixed to the same value. For the efforts however we can choose $n-1$ of them as inputs and let the remaining one be an output, given by (6.2). If we take e_1, e_2, ... e_{n-1} as inputs then we get the causality of Figure 6.30b.

The argument can be summarized in the following rule:

$$\begin{array}{c}\text{All bonds of an } s \text{ junction except one}\\ \text{shall have the causal stroke at the } s.\end{array} \tag{6.3}$$

For a p junction the roles of effort and flow change places and the rule therefore becomes

$$\begin{array}{c}\text{Precisely one of the bonds of a } p \text{ junction}\\ \text{shall have the causal stroke at the } p.\end{array} \tag{6.4}$$

Causality of Transformers and Gyrators

Since a transformer relates the variables as

$$e_2 = ne_1, \quad f_2 = \frac{1}{n} f_1$$

we must choose inputs and outputs in the same way in the bonds. This gives two possibilities for the causal strokes (see Figure 6.31).

A gyrator ties the effort of one side to the flow of the other side, so the causal stroke has to be moved (see Figure 6.32).

Figure 6.31: Possible causalities of a transformer.

Figure 6.32: Possible causalities for a gyrator.

Propagating Causality in a Graph

We have seen that some bonds have an automatic (sources) or a natural (I and C elements) causality. Also it is clear that the rules for p and s junctions, transformers, and gyrators will force a certain choice of causality for some other bonds. We formalize this in the following algorithm.

1. Choose a source and mark its automatic causality.
2. Some adjacent bonds have only one possible causality because of the rules for junctions, transformers, and gyrators. Mark these causalities as far into the graph as possible.
3. Repeat steps 1 and 2 for all sources.
4. Choose an I or C element and mark its natural causality.
5. Mark the causalities that are now fixed analogous to step 2.
6. Repeat steps 4 and 5 for all I and C elements.

7. Choose some R element that has no causality mark and fix an arbitrary one.

8. Do the analogy of step 2.

9. Repeat steps 7 and 8 for all remaining R elements.

The normal case is that the algorithm terminates after step 6 with causality for all bonds and all rules satisfied. If the algorithm stops because the causality rules are in conflict, then the problem might be ill-posed. We will defer a discussion of such questions to Section 6.8. Instead we will show how to go from a causality marked bond graph to a state-space description. First, we will give an example.

Example 6.3 Causality for an Electrical System
We will use the bond graph of an electrical system constructed in Section 6.3 (Figure 6.15). See Figure 6.33a. Here we have already performed step 1 in the algorithm by marking the causality at the effort source. Steps 2 and 3 of the algorithm do not give anything new. At step 4 we give causality to the bond at the I element. See Figure 6.33b. In step 5, rule (6.3) says that we must mark causalities at the left junction as shown Figure 6.33c. According to step 6, we have to repeat step 4 for the C element. This gives Figure 6.33d. Finally, step 5, using (6.4), gives causality to the last bond. See Figure 6.33e. We have now succeeded in giving causality to all bonds without conflict with the rules. □

6.7 State Equations from Bond Graphs

A great advantage of bond graphs is that the choice of state variables is natural. The memory of the system lies in the I and C elements. From the physical interpretation of the states described in Section 3.4 it then follows that it is natural to associate one state variable with each such element. Mathematically, we do it in the following way. Consider an effort storage

$$f(t) = \frac{1}{\alpha}\int^{t} e(\tau)d\tau$$

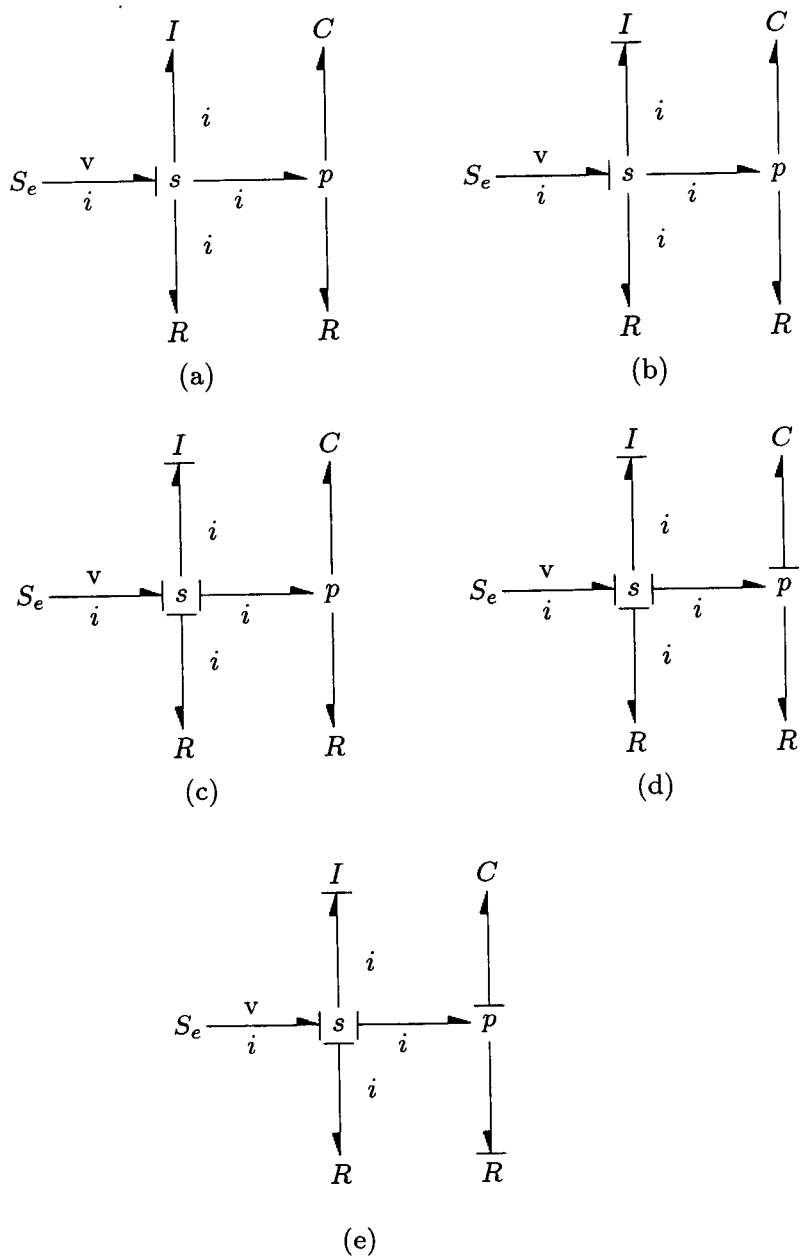


Figure 6.33: Causality in a bond graph of an electrical system.

We see that it is natural to choose the state variable x as

$$x = f \;\Rightarrow\; \dot{x} = \frac{1}{\alpha} e \tag{6.5}$$

or

$$x = \alpha f \;\Rightarrow\; \dot{x} = e \tag{6.6}$$

These choices are equivalent but one of them might be more natural from a physical point of view. In both cases the flow f is uniquely determined by x and (6.5) or (6.6) is a state equation for x, *provided* we can express e in states and inputs only.

For a flow storage the corresponding equation is

$$e(t) = \frac{1}{\beta} \int^t f(\tau) d\tau$$

with the following natural choices of state.

$$x = e \;\Rightarrow\; \dot{x} = \frac{1}{\beta} f \tag{6.7}$$

or

$$x = \beta e \;\Rightarrow\; \dot{x} = f \tag{6.8}$$

If we introduce state equations of the form (6.5) – (6.8) for all I and C elements, we get a state-space description as soon as all flows of C elements and all efforts of I elements are expressed in states and inputs. The flow of an I element and the effort of a C element are given by the states, and they are also regarded as outputs by the causality of the bonds. We can then do calculations *as if* the I elements were flow sources and the C elements effort sources. We can then propagate the signals through the graph using the causality marks as local information whether a signal is to be regarded as an input or an output. When the whole graph has been filled, we know in particular the right sides of (6.5)-(6.8).

We illustrate the method using the bond graph of Example 6.3. With the causality of Figure 6.33, we have the graph of Figure 6.34a, where we have marked those variables that are known when the I element is regarded as a flow source (the current i) and the C element as an effort source (the voltage v_c). At the s junction we now have the flow i as input. Then all the other flows are given as outputs ($= i$). In

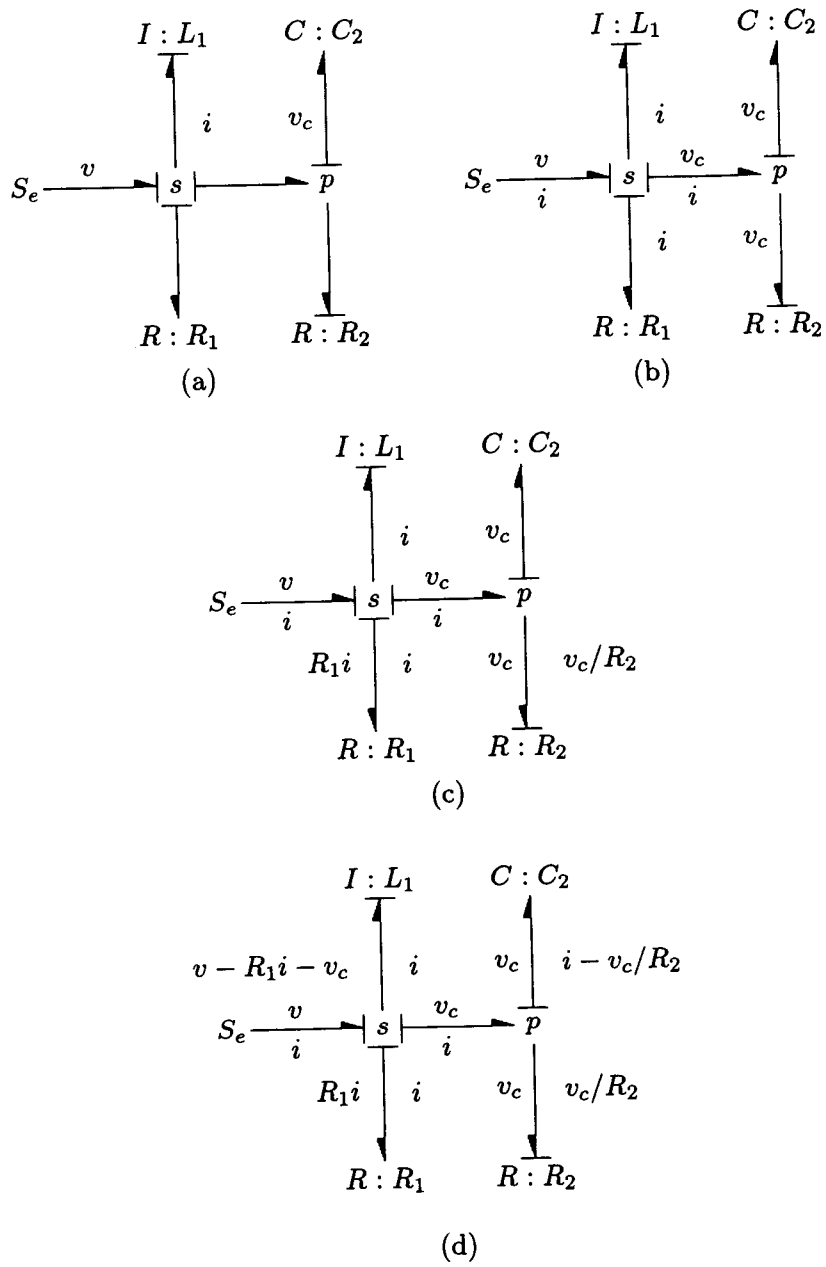


Figure 6.34: Relations among variables for an electrical system.

the same way, v_c at the p junction determines all other efforts to be equal to v_c. The result is shown in Figure 6.34b.

We see that we now know the inputs of the R elements with values R_1 and R_2 so that their outputs $R_1 i$ and v_c/R_2 can be calculated (see Figure 6.34c).

Since we now know the three efforts that are inputs to the s junction, the fourth one, which is an output, can be computed to be $v - R_1 i - v_c$. At the p junction we analogously calculate the current into the capacitor to be $i - v_c/R_2$. The complete graph is shown in Figure 6.34d.

If we introduce the state variables $x_1 = i$ and $x_2 = v_c$, we get directly, by reading the bonds at the I and C elements, the equations

$$\dot{x}_1 = \frac{1}{L_1}(v - R_1 i - v_c)$$
$$\dot{x}_2 = \frac{1}{C_2}(i - \frac{v_c}{R_2})$$

or equivalently

$$\dot{x}_1 = -\frac{R_1}{L_1}x_1 - \frac{1}{L_1}x_2 + \frac{1}{L_1}v$$
$$\dot{x}_2 = \frac{1}{C_2}x_1 - \frac{1}{R_2 C_2}x_2$$

This is a state space description with input v. The procedure of this example can be formalized into an algorithm. It is then possible to translate a bond graph into state equations automatically on a computer. We will discuss this in Section 7.4. However, the informal reasoning that we used in this example is usually sufficient for bond graphs that are small enough for hand calculations.

6.8 Ill-posed Modeling Problems and Bond Graphs

Sometimes the causality rules lead to conflicts that cannot be resolved. Often this means that the problem is in some sense ill-posed. We give some examples.

Consider the simple circuit of Figure 6.35a. With the causality

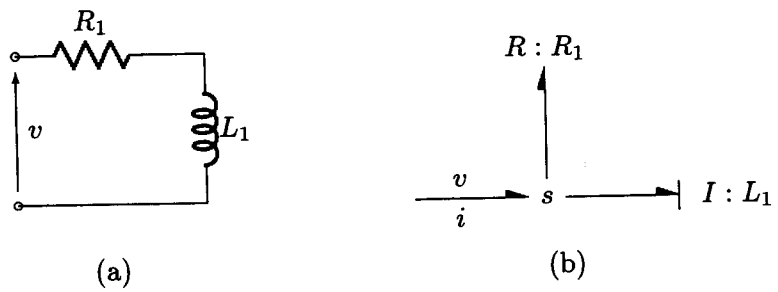


Figure 6.35: Electrical circuit.

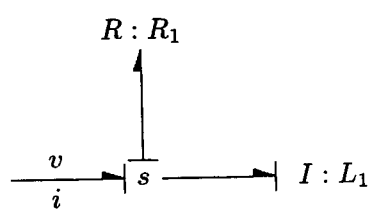


Figure 6.36: Electrical circuit with complete causality

of an I element marked, we get the bond graph of Figure 6.35b. The rules of an s junction now make it necessary to complete the graph as in Figure 6.36. We see that we have to choose v as the input. If we choose the current i we get an inconsistent diagram (if we do not want to move the causality of the inductor to make it an element that differentiates the signal). Hence we note that one cause of problems can be a *bad choice of input signal.* Physically, the problem with the current as input is related to the fact that the current is also the natural choice of state variable for the inductor. We like to think of inputs as variables that can be changed momentarily, but it is not natural to be able to change the state (and thereby the stored energy) of a system momentarily.

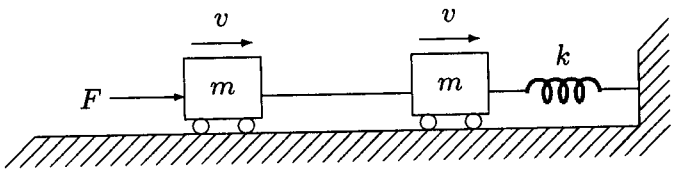
Now consider the mechanical system of Figure 6.37a. The accompanying graph is shown in Figure 6.37b. We have marked causalities except on the bond between the s junctions. For this bond there is no possibility to get consistence with the rules. The problem is related to the fact that the states associated with the masses are not independent of each other (the velocity is the same for both masses).

In this case our physical insight tells us that the problem can be solved by lumping the masses together as in Figure 6.38a, giving the bond graph of Figure 6.38b.

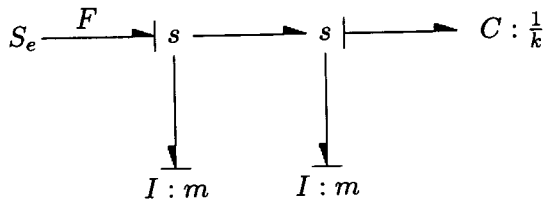
Another possibility would be to introduce a spring between the masses (see Figure 6.39a). The bond graph, shown in Figure 6.39b, then has no causality conflict. If we try to make the connection between the masses "almost rigid" by making k' large, there will be a problem with stiff differential equations when simulating the system (see Chapter 11).

6.9 Controlled Elements

In many cases the elements of a bond graph are variable. Variable R elements occur for instance in electrical systems (potentiometers), in hydraulic systems (valves), and mechanical systems (brakes). In some cases the variable resistance can be regarded as an input. In other cases it might be determined by variables in some other part of the bond graph. We denote this a *controlled* R element. All types of elements can in principle be controlled. The most common cases except

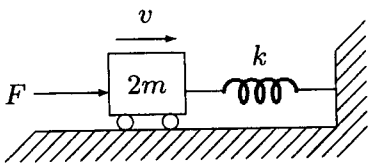


(a)

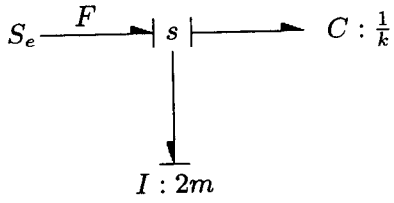


(b)

Figure 6.37: An ill-posed mechanical problem.

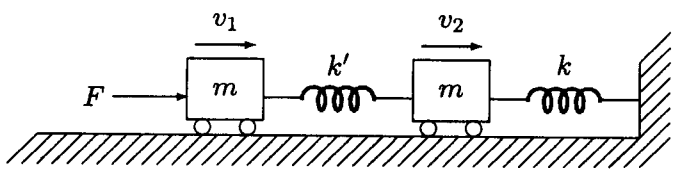


(a)

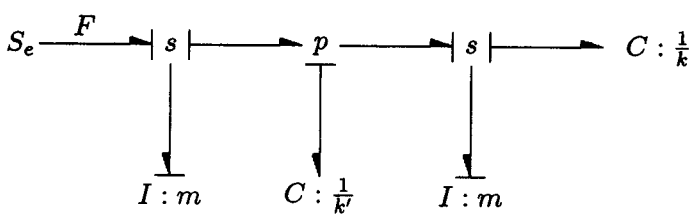


(b)

Figure 6.38: A mechanical system with the masses lumped into one.



(a)



(b)

Figure 6.39: A mechanical system with a spring between the masses.

Figure 6.40: A controlled source and a controlled R element.

R elements are probably controlled sources and controlled transformers and gyrators. In the simple cases we will discuss here, the control implies a one-way dependence without any significant transfer of energy. It is marked as shown in Figure 6.40. The whole arrow marks a signal flow without energy transfer. If the control is via an effort variable, we get the graph of Figure 6.41a. The figure shows that e_0 of the effort source depends on the common effort e of the p junction. The dependence, which can be dynamic, is symbolized by the block G. [If G is a linear system $G = G(p)$ in the differentiation operator p, we have $e_0(t) = G(p)e(t)$.] Since we assume that the signal flow of the whole arrows implies no energy transfer, the flow out of the p junction this way must be zero. When writing down the flow balance, we should therefore include only the ordinary bonds:

$$f_1 = f_2 + f_3$$

If we instead have a flow as input to G, we get the diagram of Figure 6.41b. In this case we get

$$e_1 = e_2 + e_3$$

since no effort is transferred by the signal arrow.

As an example we can study the flow system of Figure 6.42. Here we describe a level control where the pressure at the bottom of the tank is taken as a measure of the level. The regulator Reg controls the valve, which is thus a controlled R element.

If the coefficients n, r of a transformer or gyrator are controlled from some other variable, we use the term *modulated* transformer or gyrator. Analogously to the preceding example, above we use the notation of

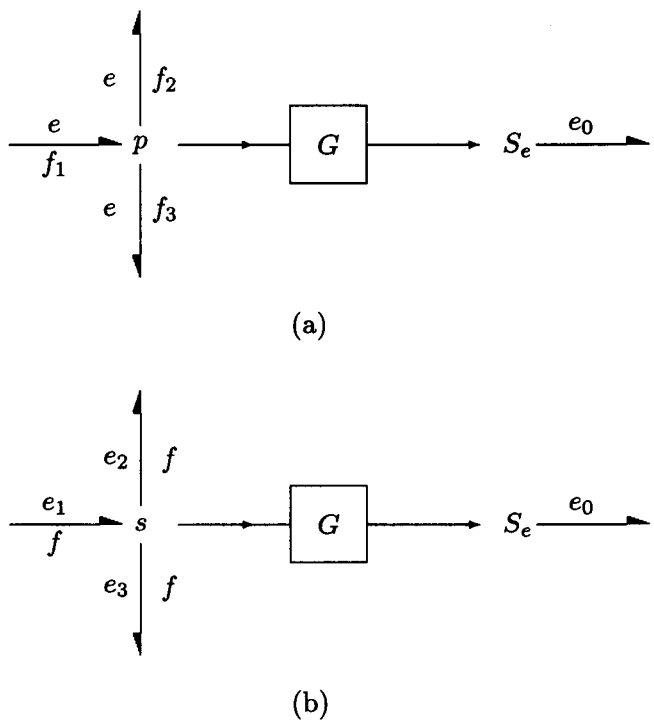


Figure 6.41: (a) Effort-controlled and (b) flow-controlled source.

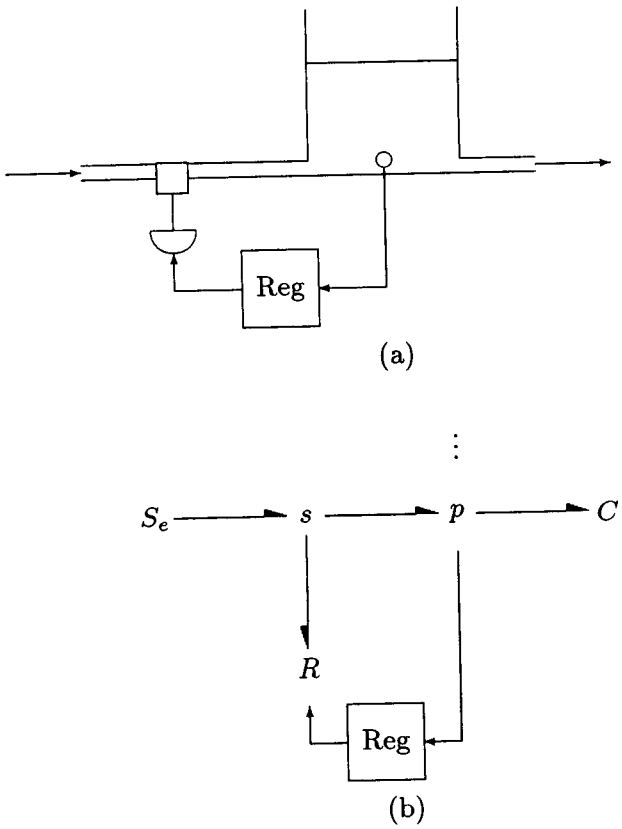


Figure 6.42: Flow system with a controlled R element.

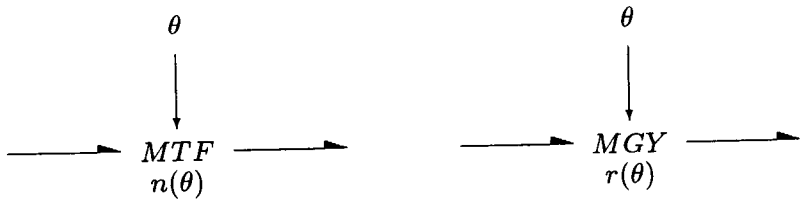


Figure 6.43: Modulated transformer and modulated gyrator.

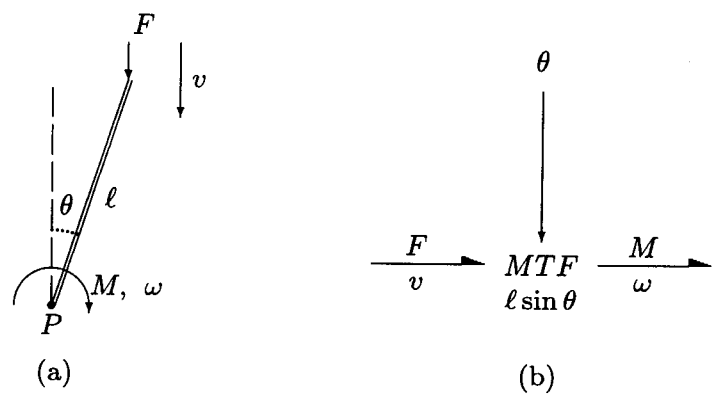


Figure 6.44: Example of a modulated transformer.

Figure 6.43, where θ denotes a variable determined at some other place in the system (or an input variable).

A typical example is the relationship between rotation and translation. Such a case is is shown in Figure 6.44a. A bar turns around the point P. Between the moment M and the force F and between the velocity v and the angular velocity ω we have the relations

$$M = (\ell \sin\theta)F, \quad v = (\ell \sin\theta)\omega$$

We can regard these equations as a modulated transformer with $n = \ell \sin\theta$ as shown in Figure 6.44b. This example can be used to describe the model of a pendulum.

Example 6.4 Model of a Pendulum

Consider Figure 6.45a. We can regard gravitation as a constant input. As we just saw, the pendulum can be regarded as a modulated transformer, transforming the gravitational force into a moment. The energy supplied by gravity is stored as rotational energy (an I element with moment of inertia = $m\ell^2$), but also lost to friction (an R element). If we assume that the moment caused by friction is proportional to the angular velocity

$$M_f = b\omega$$

we get the bond graph of Figure 6.45b. Here the relationship between θ

and ω is also marked ($\omega = \dot{\theta}$). If we fill in the relationships generated by the graph, we finally get Figure 6.45c. Introducing ω as a state variable we get $m\ell^2\ \dot{\omega} = -mg\ell \sin\theta - b\omega$, which gives the state-space description

$$\begin{aligned} \dot{\omega} &= -\tfrac{g}{\ell}\sin\theta - \tfrac{b}{m\ell^2}\omega \\ \dot{\theta} &= \omega \end{aligned}$$

□

6.10 Further Remarks

We have presented the basis of the theory of bond graphs. This theory has been extended in many ways. We list some of these extensions.

1. For electrical and mechanical systems there is a special systematic technique. It is outlined in Section 6.12.

2. The bonds can represent vector-valued variables. This is of great interest in mechanical applications.

3. Simple thermal systems can be handled using *pseudo bond graphs*, in which the heat flow is the flow variable and the temperature effort variable.

4. General thermodynamic problems can be handled with bond graphs in which the entropy flow is the flow variable and the temperature the effort variable.

5. We have assumed that the storage of flow and effort only affects one variable at a time. More complicated situations are possible and can be treated within the framework of bond graphs.

For details regarding these extensions we refer to the bibliography at the end of Part II. Finally, we note some differences regarding notation.

1. We have marked series and parallel junctions with an s and a p, respectively. It is also common to use 1 and 0 (see Figures 6.46 and 6.47).

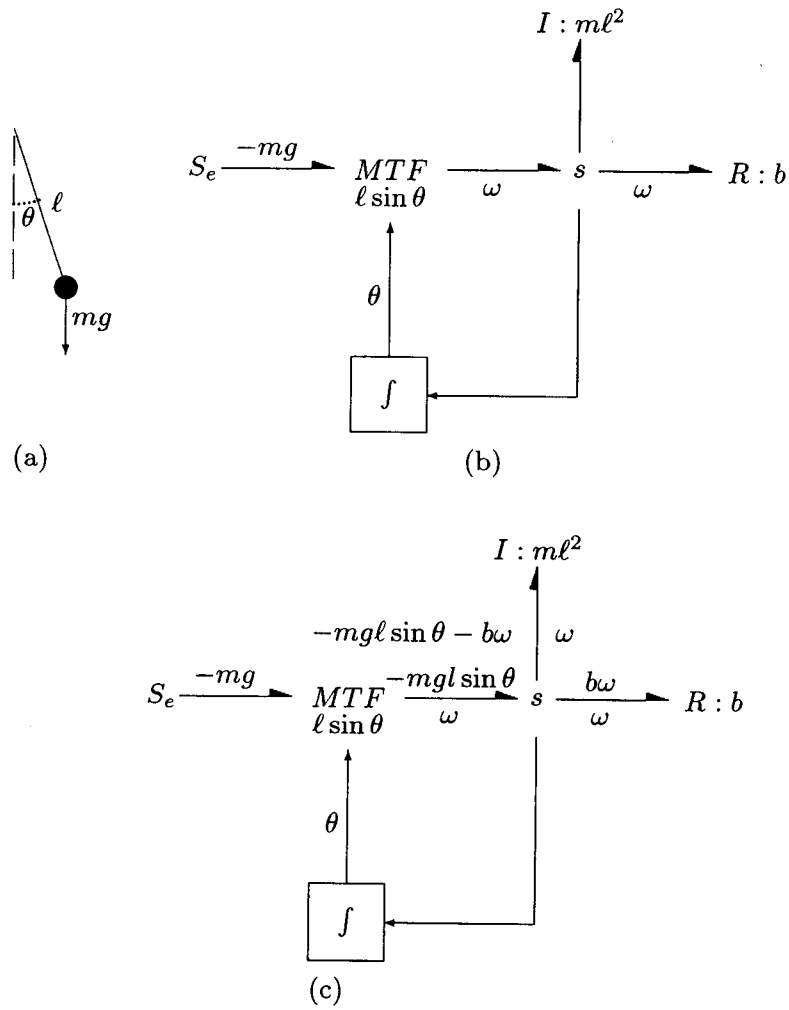


Figure 6.45: Description of a pendulum.

Figure 6.46: Series junction.

Figure 6.47: Parallel junction.

2. From a purely mathematical point of view it does not matter if the variables denoting flow and effort change places. For mechanical systems we could use force and moment as flow variables and velocity and angular velocity as effort variables. This is in fact done by many authors. The only difference is that I elements are turned into C elements, s junctions into p junctions, some transformers into gyrators, and vice versa.

6.11 Conclusions

We have seen that bond graphs represent a very systematic method for the modeling of systems in physics and engineering. It is an advantage that the method is based on a central physical concept — energy — and that the same methodology is used for different types of systems. Another advantage is that we work locally, "one junction at a time." In this way, complex systems are in principle no more difficult than simple ones. The bond graph is also a complete description of a dynamic system. It can automatically be translated into state-space equations. Perhaps the main disadvantage with bond graphs is that we exclude

nontechnical applications.

6.12 Appendix

A Systematic Method of Translating Electrical Circuits into Bond Graphs

Most electrical diagrams can easily be converted into bond graphs by grouping components as series and parallel connections. However, we could also use the following completely systematic procedure:

1. Introduce a p junction for every point with a well-defined potential.

2. Introduce an s junction with a C, I, or R element attached for every component. Introduce an s junction with a source attached for every input.

3. Use the fact that grounded points have zero voltage (and consequently do not contribute to the summation at s junctions) for removing certain bonds.

4. Use the simplification rules of Section 1 for removing and merging junctions.

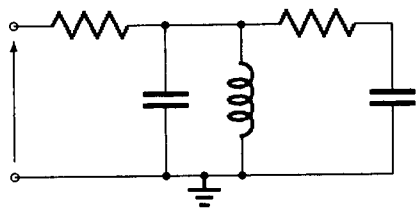
Example 6.5 Simplification of an Electric Circuit.
Consider the circuit of Figure 6.48(a). Steps 1 and 2 give Figure 6.48(b). Note that the structure of the circuit diagram is preserved. Step 3 gives Figure 6.48(c) and step 4 gives (d). □

A Systematic Method for Mechanical Systems

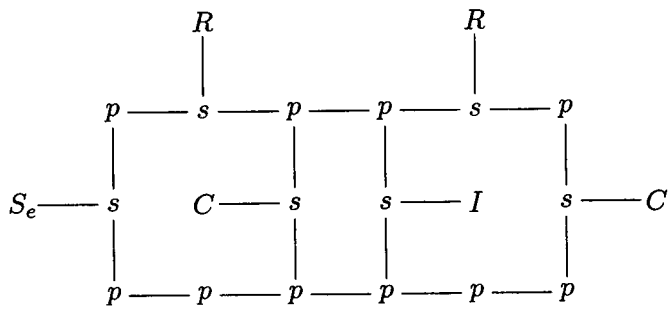
For mechanical systems we can use the following procedure:

1. Introduce an s junction for every point with a well-defined velocity.

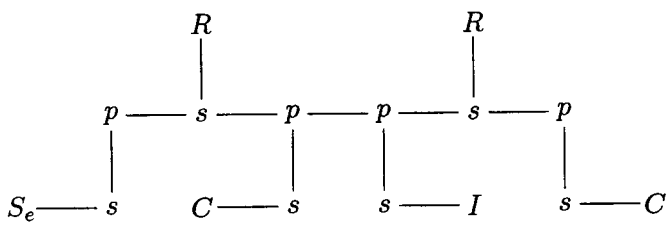
2. Introduce s junctions for velocity *differences* and p junctions to form these differences.



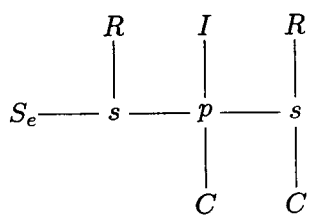
(a)



(b)



(c)



(d)

Figure 6.48: Systematic procedure for an electrical bond graph.

3. Introduce I elements at those s junctions that are associated with masses.

4. Introduce C and R elements.

5. Make suitable simplifications.

When dealing with the first item it is sometimes convenient to introduce an s junction for a fictitious point with velocity zero. This junction is then eliminated from the graph when it is simplified.

Example 6.6 Simplification of a Mechanical System
Consider the mechanical system Figure 6.49(a). The first two items give Figure 6.49(b). (We have introduced the velocity differences needed for the resistive and capacitive elements.) With I, R, and C elements introduced, we get Figure 6.49(c), and finally a simplification gives (d), where also a source has been introduced. □

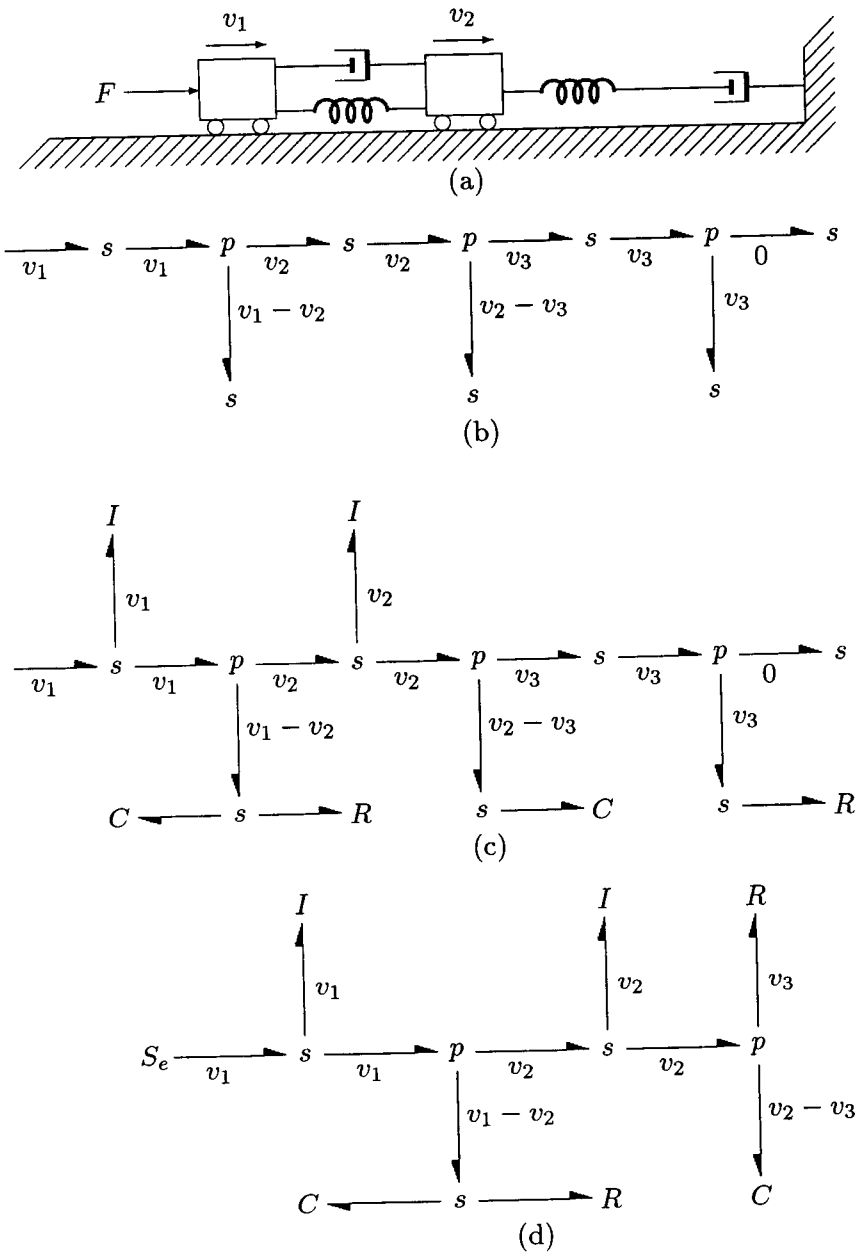


Figure 6.49: Systematic procedure for a mechanical bond graph.

Chapter 7

Computer-aided Modeling

In earlier chapters we discussed different systematic approaches to modeling. We assumed that the goal is a state-space description.

$$\begin{aligned} \dot{x} &= f(x, u) \\ y &= h(x, u) \end{aligned} \tag{7.1}$$

We will see in Chapter 11 on simulation that this is natural, considering how numeric methods work. Most simulation programs assume that the user supplies the right side of (7.1) in some form.

We have seen that the road from the physical description to the state-space form (7.1) can contain many complicated calculations even for comparatively simple physical systems. It is thus natural to investigate to what extent a computer can assist in this work. We will discuss some aspects of this. First we will discuss computer algebra in general, then we will present two systematic methods of arriving at a state-space description, one of them algebraic and the other based on bond graphs. In Chapter 11 we will discuss how some of these features are incorporated into simulation languages.

7.1 Computer Algebra and Its Applications to Modeling

Computer algebra uses methods for manipulation of mathematical formulas as opposed to numerical calculations. There are a number of commercially available programs, for example, Macsyma, Maple, Reduce, Axiom, and Mathematica. Some examples of the capabilities of a typical computer algebra program follow:

- Algebraic expansions

$$(x+y)^2 = x^2 + 2xy + y^2$$

- Factorizations

$$x^3 - y^3 = (x-y)(x^2 + xy + y^2)$$

- Symbolic differentiation

$$\frac{\partial}{\partial z}(x^2 z + \sin yz + \mathrm{a}\tan z) = x^2 + y\cos yz + \frac{1}{1+z^2}$$

- Symbolic integration

$$\int \sqrt{1+x^2}dx = \frac{1}{2}(\operatorname{arcsinh} x + x\sqrt{x^2+1})$$

Obviously, such a program is of interest for almost all engineering and scientific applications. We will here discuss some applications in modeling.

Using the biological example of Section 2.2, equations (2.2a) and (2.2b), we will show how Macsyma can be used as a tool for modeling. Figure 7.1 shows the result of a simple computer session. The right sides of the differential equations are in `f1` and `f2`. The stationary points are determined using `solve` to solve the nonlinear system of equations defined by `f1` and `f2`. This gives three solutions displayed in (d5). The command (c6) gives the linearization, as shown in (d6). By substituting the different solutions of (d5) we get the linearizations of the different stationary points, (d7), (d8) and (d9). Eliminating the time variable from (2.2a) and (2.2b) gives the differential equation of (c11). It can be solved using the `ode` command giving an explicit expression for the solution curves in the `n1-n2`-plane, (d12). In this expression `%c` is an arbitrary constant of integration.

```
(c3) f1;

(d3)        2 n1 - n1 (n2 + n1)
(c4) f2;
(d4)         n2 - n2 (n2 + n1)
(c5) solve([f1,f2],[n1,n2]);

(d5)      [[n1 = 0, n2 = 0], [n1 = 2, n2 = 0], [n1 = 0, n2 = 1]]

(c6) a:matrix([diff(f1,n1),diff(f1,n2)],[diff(f2,n1),diff(f2,n2)]);

                    [ - n2 - 2 n1 + 2      - n1       ]
(d6)                [                                 ]
                    [     - n2        - 2 n2 - n1 + 1 ]

(c7) a1:subst(d5[1],a);
                                  [ 2  0 ]
(d7)                              [      ]
                                  [ 0  1 ]

(c8) a2:subst(d5[2],a);
                                 [ - 2  - 2 ]
(d8)                             [          ]
                                 [  0   - 1 ]

(c9) a3:subst(d5[3],a);
                                 [   1   0 ]
(d9)                             [         ]
                                 [ - 1  - 1 ]

(c10) depends(n2,n1);
(d10)     [n2(n1)]

(c11) diffeq:diff(n2,n1)=f2/f1;
                            dn2   n2 - n2 (n2 + n1)
(d11)                       --- = -------------------
                            dn1  2 n1 - n1 (n2 + n1)

(c12) ode(diffeq,n2,n1);
                                      2
                             2 n1 n2 + n1  - 2 n1
(d12)                      - ------------------- = %c
                                        2
                                   2 n2
```

Figure 7.1: A computer algebraic analysis of a biological system: c, user commands; d, replies by the computer.

```
(c2) depends(h,t);

(d2)       [h(t)]

(c3) tank:diff(h,t)=(-a*sqrt(2*g*h)+u)/aa;

                          dh   u - sqrt(2) a sqrt(g h)
(d3)                      -- = ----------------------
                          dt          aa

(c4) ode(tank,h,t);


(d4) - (aa log(sqrt(2) a sqrt(g) sqrt(h) - u) u - aa u

                                                   2
             + sqrt(2) a aa sqrt(g) sqrt(h))/(a  g) = t + %c
```

Figure 7.2: Analytical solution for the tank example.

7.2 Analytical Solutions

Up to now it has been regarded as more or less self-evident that the state equations have to be solved numerically, apart from some very simple cases. We will discuss numerical methods in Chapter 11. However, the rapid development of computer algebra has changed the situation in recent years. We saw in the example of Section 7.1 that the differential equations admitted an explicit solution. Also there are more analytically solvable differential equations than we would first expect. The fact that there might be algebraic manipulations involved to get the solutions is of less importance when they can be done in computer algebra. It is thus well worth investigating whether analytic solutions can be found for the model under consideration. We give one more example for the tank system of Section 2.3 (see Figure 7.2). Here we have defined the differential equation in (c3) and solved it with the command `ode` giving the result of (d4), where `%c` is an arbitrary constant.

Even when no complete solution can be presented, there might be

partial results that are interesting. For a second-order system

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2) \\ \dot{x}_2 &= f_2(x_1, x_2) \end{aligned} \tag{7.2}$$

the solution algorithm sometimes generates results of the form

$$F(x_1, x_2) = C \tag{7.3}$$

where C is a constant. (In fact this was the result we got in the example of Figure 7.1.) Sometimes it is possible to continue from this expression to explicit solutions of the form

$$\begin{aligned} x_1(t) &= \varphi_1(t) \\ x_2(t) &= \varphi_2(t) \end{aligned}$$

Even when this in not possible, equation (7.3) will give essential information. The function F of (7.3) is often called an *integral* of the system. Geometrically, (7.3) is the equation of the path in the $x_1 - x_2$ plane. The information that is not present is the velocity along the path. An example is the pendulum of Example 6.4 with the equation

$$\begin{aligned} \dot{\theta} &= \omega \\ \dot{\omega} &= -\tfrac{g}{\ell} \sin\theta \end{aligned}$$

(here $b = 0$). In this case

$$\frac{1}{2}\omega^2 - \frac{g}{\ell}\cos\theta = C$$

is an integral of the system. Physically it represents the total energy (the sum of kinetic and potential energy) of the system. Figure 7.3 shows a plot of the θ-ω plane for the case $g/\ell = 1$.

7.3 Algebraic Modeling

In many modeling cases, physical knowledge will provide a number of equations describing the system. These equations are usually not in the desired state-space form. An important task is then to try to transform the equations into a more convenient form, a task for which computer algebra is often of great help.

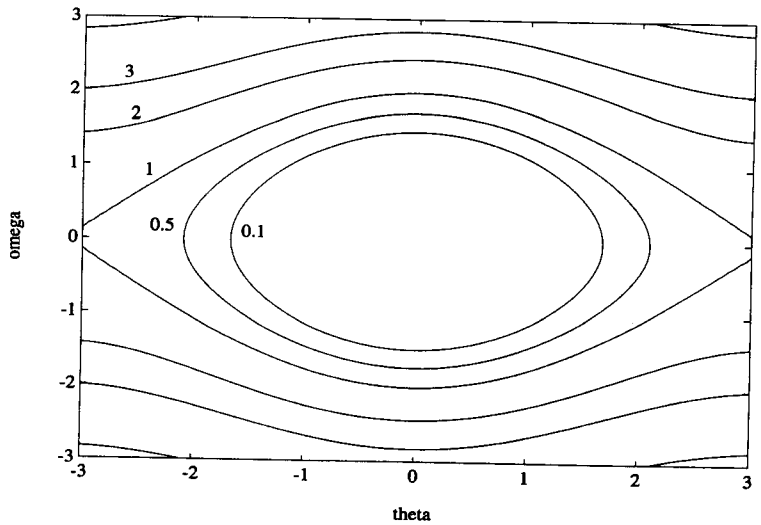


Figure 7.3: The θ-ω plane of a pendulum.

Introduction of State Variables for Higher-Order Differential Equations

Sometimes the modeling will give differential equations with high-order derivatives. Consider the case of a single such equation

$$F(y, \dot{y}, \ldots, y^{n-1}, y^{(n)}; u) = 0$$

where u is the input. Let us introduce the variables

$$x_1 = y, \;\; x_2 = \dot{y}, \ldots, \;\; x_n = y^{(n-1)}$$

We then get the equations

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ F(x_1, x_2, \ldots, x_n, \dot{x}_n; u) &= 0 \end{aligned}$$

This is a state-space description provided $\dot{x}_n$ can be solved from the last equation.

Example 7.1
Let

$$y^{(3)^2} - \dot{y}^2 y^4 - 1 = 0$$

be given [here $y^{(3)}$ denotes the third-order derivative of y with respect to time]. With $x_1 = y$, $x_2 = \dot{y}$, and $x_3 = \ddot{y}$, we get

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3^2 - x_2^2 x_1^4 - 1 &= 0 \end{aligned}$$

Here the last equation can be solved for $\dot{x}_3$ giving

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= \pm\sqrt{1 + x_2^2 x_1^4} \end{aligned}$$

Note that we have to treat two cases if we do not know the sign of $y^{(3)} = \dot{x}_3$ from the physical context. □

Systems of Higher-Order Differential Equations

Now consider the slightly more complicated situation in which there are two higher-order differential equations in two variables

$$\begin{aligned} F(y, \dot{y}, \ldots, y^{(n)}; v, \dot{v}, \ldots, v^{(m-1)}; u) &= 0 \\ G(y, \dot{y}, \ldots, y^{(n-1)}; v, \dot{v}, \ldots, v^{(m)}; u) &= 0 \end{aligned}$$

[We use the notation $y^{(j)} = d^j y/dt^j$.] Here we can introduce the variables

$$x_1 = y, \;\; x_2 = \dot{y}, \ldots, \;\; x_n = y^{(n-1)}$$

$$x_{n+1} = v, \;\; x_{n+2} = \dot{v}, \ldots, \;\; x_{n+m} = v^{(m-1)}$$

and we get

$$
\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_{n-1} &= x_n \\
F(x_1,\ x_2, \ldots, x_n, \dot{x}_n; x_{n+1}, \ldots, x_{n+m}; u) &= 0 \\
\dot{x}_{n+1} &= x_{n+2} \\
&\vdots \\
\dot{x}_{n+m-1} &= x_{n+m} \\
G(x_1,\ x_2, \ldots, x_n; x_{n+1}, \ldots, x_{n+m}, \dot{x}_{n+m}; u) &= 0
\end{aligned}
$$

This is a state-space description provided $\dot{x}_n$ and $\dot{x}_{n+m}$ can be solved for in F and G, respectively.

Example 7.2

$$
\begin{aligned}
\ddot{y} + \dot{v}^2 + y &= 0 \\
\dot{y}^2 + \ddot{v} + vy &= 0
\end{aligned}
$$

Here the variables

$$x_1 = y,\ \ x_2 = \dot{y},\ \ x_3 = v,\ \ x_4 = \dot{v}$$

directly give the state-space description

$$
\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -x_1 - x_4^2 \\
\dot{x}_3 &= x_4 \\
\dot{x}_4 &= -x_1 x_3 - x_2^2
\end{aligned}
$$

□

Example 7.3
Let us modify Example 7.2 somewhat.

$$\ddot{y} + v^{(3)} + \dot{v}^2 + y = 0 \tag{7.4}$$

$$\dot{y}^2 + \ddot{v} + vy = 0 \tag{7.5}$$

We cannot use the technique of Example 7.2 directly, since the highest derivatives of v and y are in the same equation. One possibility is to differentiate (7.5) and subtract it from (7.4), eliminating $v^{(3)}$.

$$\ddot{y}+v^{(3)}+\dot{v}^2+y-(2\dot{y}\ddot{y}+v^{(3)}+v\dot{y}+\dot{v}y) = (1-2\dot{y})\ddot{y}+\dot{v}^2-\dot{v}y-v\dot{y}+y = 0$$

We can let this equation represent our system together with (7.5) and choose the state variables as in Example 7.2. This gives the result

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= \frac{-x_4^2 + x_4x_1 + x_2x_3 - x_1}{1-2x_2} \\
\dot{x}_3 &= x_4 \\
\dot{x}_4 &= -x_1x_3 - x_2^2
\end{aligned}$$

□

A technique of the form suggested in Example 7.3 can often be used to eliminate variables in static relationships.

Example 7.4
Consider the system

$$\ddot{y} + \ddot{v} + \dot{y}\dot{v} = 0 \tag{7.6}$$

$$\frac{y^2}{2} + \frac{v^2}{2} - 1 = 0 \tag{7.7}$$

A differentiation of (7.7) gives

$$y\dot{y} + v\dot{v} = 0 \tag{7.8}$$

One more differentiation gives

$$\dot{y}^2 + y\ddot{y} + \dot{v}^2 + v\ddot{v} = 0 \tag{7.9}$$

Multiplying (7.6) by v and subtracting (7.9), eliminates the $v\ddot{v}$ terms. This gives

$$v\ddot{y} + v\dot{y}\dot{v} - \dot{y}^2 - y\ddot{y} - \dot{v}^2 = 0$$

If this equation is multiplied by v^2,

$$v^2(v-y)\ddot{y} - v^2\dot{y}^2 + v^2\dot{y}(v\dot{v}) - (v\dot{v})^2 = 0$$

then we see that $v\dot{v}$ can be eliminated with the help of (7.8). The result is

$$v^2(v-y)\ddot{y} - v^2\dot{y}^2 - v^2y\dot{y}^2 - y^2\dot{y}^2 = 0$$

Finally, v can be eliminated with the help of (7.7), giving an equation in y only.

$$(2-y^2)(\sqrt{2-y^2}-y)\ddot{y} - (2-y^2)\dot{y}^2 - (2-y^2)y\dot{y}^2 - y^2\dot{y}^2 = 0$$

Introducing

$$x_1 = y, \quad x_2 = \dot{y}$$

finally gives

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= -\frac{(x_1^3 - 2x_1 - 2)x_2^2}{(2-x_1^2)^{3/2} + x_1^3 - 2x_1}\end{aligned}$$

which is a state-space description. □

Example 7.4 is interesting because it shows that we can eliminate one of the variables from the two equations (7.6) and (7.7), although both are nonlinear and contain derivatives of the variables.

The calculations of the examples can in principle be done by hand. For slightly more complex problems there is, however, hardly any alternative to computer algebra. We can show that the methods of Example 7.4 can be generalized to an arbitrary number of equations in an arbitrary number of variables with derivatives of arbitrarily high order, provided all equations are polynomial in the variables and their derivatives. (It is even possible to have partial derivatives of the variables.) The general algorithm is complicated but can be automated using computer algebra.

7.4 An Automatic Translation of Bond Graphs to Equations

To understand the principle behind an automatic generation of equations from bond graphs, we consider Figure 7.4. We have marked the causality that is without conflict. Let us introduce the state $x = \alpha f_2$ for the I element, giving

$$\dot{x} = e_2$$

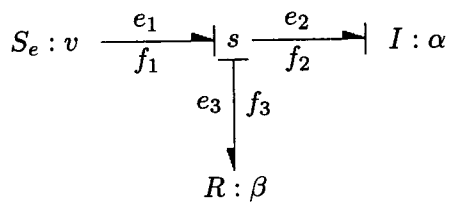


Figure 7.4: Bond graph for generation of state equations.

We can now use the following reasoning. Imagine a list of equations in which all e_i and f_i are computed from the given v and x. First on the list we can write

$$e_1 = v$$

where we use the given input. We can put the computation of the other variable of the bond f_1 last on the list. Then we can be sure that whatever is needed for the calculation will be available. This gives us the following list:

$$\begin{aligned} e_1 &= v \\ &\vdots \\ f_1 &= f_2 \end{aligned}$$

The last equation could also be $f_1 = f_3$. Since the equation is last, both f_2 and f_3 are clearly known from previous calculations at this point.

Since there are no more sources, we turn our attention to storage elements. From the I element we get the equation

$$f_2 = \frac{x}{\alpha}$$

which can be evaluated for given x. The paired variable e_2 is used in $\dot{x} = e_2$ and is an output of the junction. We then get

$$\dot{x} = e_2 = e_1 - e_3$$

Placing this equation second to last in the list assures that e_1 and e_3 are calculated before. The list is now

$$\begin{aligned} e_1 &= v \\ f_2 &= \tfrac{1}{\alpha}x \\ &\vdots \\ \dot{x} &= e_2 = e_1 - e_3 \\ f_1 &= f_2 \end{aligned}$$

We can now investigate what further variables have been defined by our first two equations. From the input f_2 of the junction we know the flows of the other bonds, in particular f_3.

$$f_3 = f_2$$

Let us place the equation for the other variable of the bond

$$e_3 = \beta f_3$$

on the first free space from below.

Since all bonds are covered the complete list is

$$\begin{aligned} e_1 &= v \\ f_2 &= \tfrac{1}{\alpha}x \\ f_3 &= f_2 \\ e_3 &= \beta f_3 \\ \dot{x} &= e_2 = e_1 - e_3 \\ f_1 &= f_2 \end{aligned}$$

Starting from the input v and the state x, all variables are evaluated in the proper order. In particular $\dot{x}$ is evaluated from x and v.

By performing successive substitutions from below, we can get a compact state space description:

$$\dot{x} = e_1 - e_3 = e_1 - \beta f_3 = e_1 - \beta f_2 = e_1 - \frac{\beta}{\alpha}x = v - \frac{\beta}{\alpha}x$$

Our example can be generalized into a general algorithm. Instead of one list filled from the top and the bottom, we could use two lists, the forward and the backward list.

Algorithm for Equation Sorting

1. Choose a source and write its input signal in the forward list and the equation of the other bond variable in the backward list.

2. Check adjacent bonds. If some variable is defined in terms of already calculated variables, write its equation in the forward list and the equation of the other bond variable in the backward list. Do this as far into the bond graph as possible.

3. Repeat steps 1 and 2 until all sources have been treated.

4. Choose an I element and write the equation $f_i = \frac{1}{\alpha_i} x_i$ (or alternatively $f_i = x_i$) in the forward list and

$$\dot{x}_i = e_i = \ldots$$

[or alternatively $\dot{x}_i = \frac{1}{\alpha} e_i = \frac{1}{\alpha}(\ldots)$] in the backward list.

5. Do the analogy of step 2.

6. Repeat steps 4 and 5 until all I elements have been processed.

7. Do the analogy of steps 4, 5, and 6 for all C elements [$e_i = \frac{1}{\beta_i} x_i$ or $e_i = x_i$ to the forward list and $\dot{x}_i = f_i = \ldots$ or $\dot{x}_i = \frac{1}{\beta_i} f_i = \frac{1}{\beta_i}(\ldots)$ to the backward list].

8. Reverse the order of the backward list and put it after the forward list.

We illustrate the algorithm using the dc motor from Section 6.5; see Figure 7.5.

We choose the following state variables

$$x_1 = \int^t v_2 d\tau = L_1 i_2, \quad x_2 = \int^t M_2 d\tau = J\omega_2$$

We then get the following lists. Note that equations defining the effort and flow of each bond are in the same row.

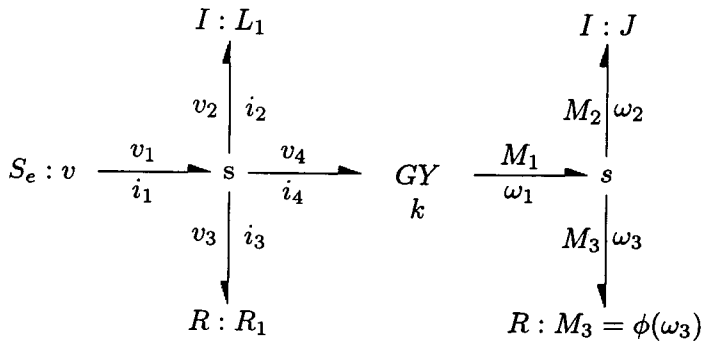


Figure 7.5: Bond graph for a dc motor.

Step	Forward list	Backward list
1	$v_1 = v$	$i_1 = i_2$
2	$i_2 = \frac{1}{L_1}x_1$	$\dot{x}_1 = v_2 = v_1 - v_3 - v_4$
2	$i_3 = i_2$	$v_3 = R_1 i_3$
2	$i_4 = i_2$	$v_4 = k\omega_1$
2	$M_1 = k i_4$	$\omega_1 = \omega_2$
4	$\omega_2 = \frac{1}{J}x_2$	$\dot{x}_2 = M_2 = M_1 - M_3$
5	$\omega_3 = \omega_2$	$M_3 = \varphi(\omega_3)$

The complete list of equations is obtained by putting the reverse back-

ward list after the forward list giving

$$
\begin{aligned}
v_1 &= v \\
i_2 &= \tfrac{1}{L_1} x_1 \\
i_3 &= i_2 \\
i_4 &= i_2 \\
M_1 &= k i_4 \\
\omega_2 &= \tfrac{1}{J} x_2 \\
\omega_3 &= \omega_2 \\
M_3 &= \varphi(\omega_3) \\
\dot{x}_2 &= M_2 = M_1 - M_3 \\
\omega_1 &= \omega_2 \\
v_4 &= k\omega_1 \\
v_3 &= R_1 i_3 \\
\dot{x}_1 &= v_2 = v_1 - v_3 - v_4 \\
i_1 &= i_2
\end{aligned}
$$

This is a complete description of the dynamic system. Eliminating all variables that are not states gives

$$
\begin{aligned}
\dot{x}_1 &= v_1 - v_3 - v_4 = v_1 - R_1 i_3 - v_4 \\
&= v_1 - R_1 i_3 - k\omega_1 = v_1 - R_1 i_3 - k\omega_2 \\
&= v_1 - R_1 i_3 - \tfrac{k}{J} x_2 = v_1 - R_1 i_2 - \tfrac{k}{J} x_2 \\
&= v_1 - \tfrac{R_1}{L_1} x_1 - \tfrac{k}{J} x_2 = v - \tfrac{R_1}{L_1} x_1 - \tfrac{k}{J} x_2 \\
\dot{x}_2 &= M_1 - M_3 = M_1 - \varphi(\omega_3) \\
&= M_1 - \varphi(\omega_2) = M_1 - \varphi(\tfrac{1}{J} x_2) \\
&= k i_4 - \varphi(x_2/J) = k i_2 - \varphi(x_2/J) \\
&= \tfrac{k}{L_1} x_1 - \varphi(x_2/J)
\end{aligned}
$$

We see that it is possible to use the list in two ways. Reading it from above we eventually evaluate the time derivative of each state variable, given the states and inputs. We will see that this is exactly what is needed in a simulation program. However we also see that it is possible to obtain a compact description of the state equations by reading the list from below and substituting. This compact equation sometimes shows the structure of the mathematical relationship more clearly. Of course it is also possible to use this description in the simulation program.

If we compare this algorithm with the one for causality, we see that they have essentially the same structure. In fact we can regard the

causality marking as a check as to whether it would be possible to write down equations in a straightforward manner, without actually doing so.

7.5 Conclusions

We have seen that it is possible to use various forms of computer support for modeling. In particular the use of computer algebra is interesting. Its use will probably increase as the capabilities of the languages increase. For practical modeling the possibility of directly translating bond graphs to equations is of great interest. It represents a direct step from a physical representation to a purely mathematical one.

Bibliography for Part II

Bond graphs are described in

R. C. Rosenberg and D. C. Karnopp: *Introduction to Physical System Dynamics.* McGraw-Hill Book Company, New York, 1983.

Various systematic modeling methods (including bond graphs) are presented in

P. E. Wellstead: *Physical System Modelling.* Academic Press, New York, 1979.

F. E. Cellier: *Continuous System Modelling*, Springer-Verlag, New York, 1991.

A number of modeling examples are given by

H. Nicholson: *Modeling of Dynamical Systems.* Peter Peregrinus Ltd., 1980.

D. J. G. James and J. J. McDonald: *Case Studies in Mathematical Modelling.* Stanley Thornes (Publishers) Ltd., 1981.

Some medical applications are described in

Matthew Witten (ed): *Mathematical Models in Medicine.* Pergamon Press, New York, 1987.

Part III

Identification

System Identification: Use of Data in Modeling

It is impossible to sit at a desk and figure out how the world works: it also has to be studied. In our case this means that we have to use experimental data of some kind during the modeling work. In earlier chapters we used such data indirectly by analyzing the system in terms of subsystems, whose functions are well known through experience and classical experiments.

This is often not enough. There can be system constants whose values we do not know. There can also be subsystems of such characteristics that it is difficult to describe their function by known physical laws. In such cases, data from the system have to be used to complete the model. These data consist of measurements of variables in the system: outputs, inputs and possibly also disturbances. Such measurements can be used to understand how the system works, to describe partial systems, and to compute values of system constants.

The technique to build and complement models from measurements is called *(system) identification.* In the following three chapters, we will describe possibilities and limitations in system identification.

There are, in principle, three different ways to use identification methods for modeling purposes.

1. Make simple experiments to facilitate phase 1 in the modeling (structuring the problem).

2. Build models to describe how the outputs depend on the inputs that are not based on any physical insight of what is happening inside the system. Such models are often linear. There are two ways of attacking this problem:

- **2a.** Building models as arbitrary, linear systems by estimating their impulse response or frequency function.
- **2b.** Estimating *ready made models* of the type (A.18), where the number n and the parameters a_i and b_i are fitted to observed data.

3. Use data to determine unknown parameters in a model obtained from physical modeling, according to Part II. We then have a *tailor-made model* in which to estimate parameters.

The main tool for problem 1 is transient analysis, which is described in more detail in Section 8.1. Problem 2a can be solved by correlation analysis, frequency analysis, or spectral analysis, which are described in the remainder of Chapter 8. These methods, like transient analysis, do not directly estimate any model parameters. They are therefore called *nonparametric identification methods.* Problems 2b and 3 both result in parameter estimation in dynamic models, and this general problem is discussed in Chapter 9. Finally, Chapter 10 contains a user-oriented account of the possibilities the identification offers for modeling.

Chapter 8

Estimating Transient Response, Spectra, and Frequency Functions

8.1 Experiments for the Structuring Phase: Transient Analysis

The first step in modeling is to decide which quantities and variables are important to describe what happens in the system. We called this the *structuring phase* in Section 4.2. It is then also necessary to decide, or guess, how the variables affect each other, which time constants are important, and which relationships can approximately be described as static ones (compare Section 4.6).

It is a rather demanding task for the modeler to answer these questions. Considerable knowledge and insights about the system will be required. Often simple experiments on the real system have to be carried out to support the work in this phase. A simple and common kind of experiment that shows how and in what time span various variables affect each other is called *step-response analysis* or *transient analysis*. In such experiments the inputs vary (typically one at the time) as a step: $u(t) = u_0$, $t < t_0$; $u(t) = u_1$, $t \geq t_0$. The other measurable variables in the system are recorded during this time. We thus study the *step response* of the system, using terminology from

Appendix A. An alternative would be to study the impulse response of the system by letting the input be a pulse of short duration. From such measurements, information of the following nature can be found:

1. The variables affected by the input in question. This makes it easier to draw block diagrams for the system and to decide which influences can be neglected.

2. The time constants of the system. This also allows us to decide which relationships in the model can be described as static (that is, they have significantly faster time constants than the time scale we are working with; see Section 4.6).

3. The characteristic (oscillatory, poorly damped, monotone, and the like) of the step responses, as well as the levels of static gains. Such information is useful when studying the behavior of the final model in simulation. Good agreement with the measured step responses should give a certain confidence in the model.

Example 8.1 Tank Dynamics

Mixing tanks are common in process industry. Their purpose is to smooth variations in concentration in a liquid by letting it pass through a big tank, where it is mixed. At a paper mill (Skärblacka, Sweden) three identical mixing vessels, coupled together as indicated in Figure 8.1, are used to smooth the concentration of pulp (see Section 4.3). The dynamics of this system were investigated by an impulse response experiment. The actual concentration of the pulp could not be manipulated since the experiment had to be done during normal operation.

The problem was solved in the following way: A bucket of water with 2070 grams of radioactive lithium, with short half-life, was poured into the first mixing vessel at point A in Figure 8.1. The radioactivity was then measured at point B during 5 hours. This radioactivity will obviously be proportional to the concentration of lithium, after correction for the half-life and background radiation. Figure 8.2 shows the measurements. Even if the measurements are disturbed by a fair amount of noise, a clear picture of a typical time response is obtained.

To correctly scale the impulse response we argue as follows: Let both the input and the output have the unit mg/liter. *Then the coefficients*

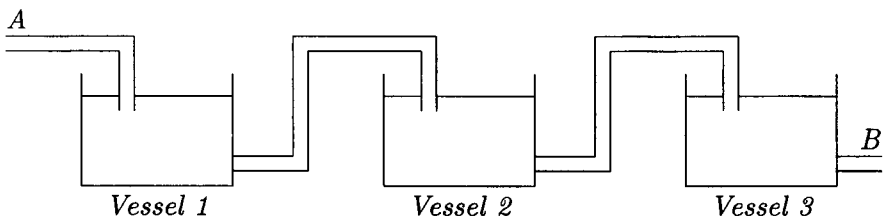


Figure 8.1: Three coupled mixing vessels.

of the impulse response will be dimensionless quantities. The total flow through the system was about 8300 liters/minute during the time of the experiment. The sudden addition of 2070 grams of lithium then corresponds to an impulse $u(t) = u_0\delta(t)$ *with* $u_0 = 2070$ grams/minute $= (2070/8300) * 10^3 \approx 250$ milligrams per liter. *Consequently we must divide the lithium concentration with this number to get the impulse response. This is shown in the lower plot of Figure 8.2. In this figure we also show the impulse response of the system*

$$G(s) = \left(\frac{1}{s\tau + 1}\right)^3 \quad \textit{for } \tau = 72 \textit{ min} \tag{8.1}$$

This gives a reasonable fit to the measurements and can be used as a model for further considerations with good approximation. Note that we have used a certain amount of physical insight in (8.1): The three cascaded mixers are identical, and therefore $G(s)$ *must be the product of three identical transfer functions. Moreover, the static gain must be 1. (Everything that is poured into the tanks will eventually leave them.) We may also note that the tanks each have a volume of* $600\,m^3$. *With a mean flow of* 8300 *liters/minute this gives, for a perfectly mixed tank, a theoretical time constant (mean residence time) of* $600/8.3 \approx 72$ *minutes. This is in excellent agreement with the result of the transient analysis.* □

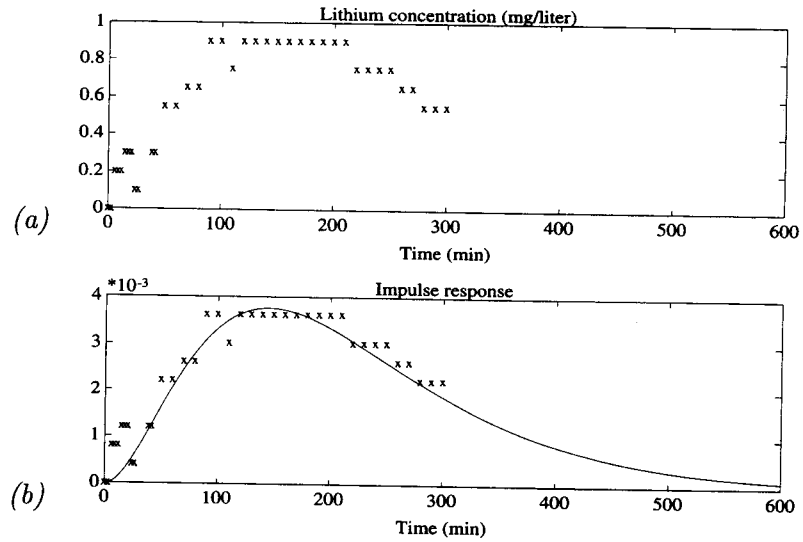


Figure 8.2: (a) Lithium concentration at point B. (b) Cross: the experimental impulse response. Solid line: a fitted impulse response.

Transient Analysis: Summary

- Transient analysis is an excellent method to get a quick and easy insight into cause and effect relationships, time delays, time constants, and static gains.
- Transient analysis is probably the most widely used method of identification in industrial practice.
- A drawback is that the obtained information is somewhat limited.
- Practical limits in the amplitude of the input, together with disturbances and measurement errors may make it difficult to determine quantitative models with a reasonable degree of accuracy.

8.2 Correlation Analysis

It is not necessary to use an impulse as input to directly estimate the impulse response of a system. Consider a sampled data system with the impulse response $\{g_k\}$:

$$y(t) = \sum_{k=0}^{\infty} g_k u(t-k) + v(t) \tag{8.2}$$

Let $\{u(t)\}$ be a signal that is a realization of a stochastic process with zero mean value and covariance function $R_u(\tau)$ [see (3.52) – (3.53)]:

$$R_u(\tau) = Eu(t)u(t-\tau) \tag{8.3}$$

and assume that $\{u(t)\}$ and $\{v(t)\}$ are uncorrelated. The cross covariance function between u and y is then

$$\begin{aligned} R_{yu}(\tau) = Ey(t)u(t-\tau) &= \sum_{k=0}^{\infty} g_k Eu(t-k)u(t-\tau) \\ + Ev(t)u(t-\tau) &= \sum_{k=0}^{\infty} g_k R_u(\tau-k) \end{aligned} \tag{8.4}$$

If the input is white noise,

$$R_u(\tau) = \begin{cases} \lambda, \tau = 0 \\ 0, \tau \neq 0 \end{cases}$$

we obtain

$$R_{yu}(\tau) = \lambda g_\tau \tag{8.5}$$

The cross covariance function $R_{yu}(\tau)$ will thus be proportional to the impulse response. Of course, this function is not known, but it can be estimated in an obvious way from observed inputs and outputs as the corresponding sample mean:

$$\hat{R}_{yu}^N(\tau) = \frac{1}{N}\sum_{t=1}^{N} y(t)u(t-\tau) \tag{8.6}$$

In this way we also obtain an estimate of the impulse response:

$$\hat{g}_\tau^N = \frac{1}{\lambda}\hat{R}_{yu}^N(\tau) \tag{8.7}$$

If the input is non white, it would be possible to estimate its covariance function as $\hat{R}_u^N(\tau)$, analogous to (8.6), and then solve for g_k from (8.4) where R_u and R_{uy} have been replaced by the corresponding estimates. However, a better and more common way is the following: First note that if both input and output are filtered through the same filter

$$y_F(t) = L(q)y(t) \qquad u_F(t) = L(q)u(t) \tag{8.8}$$

then the filtered signals will be related by the same impulse response as in (8.2):

$$y_F(t) = \sum_{k=1}^{\infty} g_k u_F(t-k) + v_F(t) \tag{8.9}$$

We can now choose the filter L so that the signal $\{u_F(t)\}$ will be as white as possible. Such a filter is called a *whitening filter*. It is often computed by describing $u(t)$ as an AR process (see Section 3.7): $A(q)u(t) = e(t)$. The polynomial $A(q) = L(q)$ will then be estimated using the least squares method. [See (9.27) and (9.35)]. Let y=u and u=0, and choose the order na to 4–8, and nb to 0). We can now use the estimate (8.7) applied to the filtered signals. The algorithm, which we call CRA (for *correlation analysis*), can thus be summarized as in the equation box (8.11).

Example 8.2 Estimation of Impulse Response
Consider the same system as in Example 3.14

$$G(s) = \frac{1}{s^2 + 2s + 1} \tag{8.10}$$

The input is piecewise constant over the sample interval $T = 0.5$ second. Figure 3.20 shows the impulse response of a continuous time system together with that of the sampled system. Input and output were observed during 250 seconds. The variance of the measurement noise was about 0.1. Let us define the signal to noise ratio *as the ratio between the input's contribution and the noise's contribution to the output variance. Then the signal to noise ratio in the data is about 6 (18 dB). A portion of the data is shown in Figure 8.3. The procedure CRA gives an estimate of the impulse response according to Figure 8.4.* □

Algorithm CRA (8.11)

1. Collect data $y(k),\ u(k),\quad k = 1, \ldots, N$
2. Subtract sample means from each signal:

$$\bar{y}(k) = y(k) - \frac{1}{N}\sum_{t=1}^{N} y(t),\ \bar{u}(t) = u(k) - \frac{1}{N}\sum_{t=1}^{N} u(t)$$

3. Form the signals

$$y_F(t) = L(q)\bar{y}(t) \qquad\qquad u_F(t) = L(q)\bar{u}(t)$$

4. Form the estimates

$$\hat{R}^N_{y_F u_F}(\tau) = \frac{1}{N}\sum_{t=1}^{N} y_F(t)u_F(t-\tau)$$

$$\hat{\lambda}_N = \frac{1}{N}\sum_{t=1}^{N} u_F^2(t)$$

5. The impulse response estimate is now

$$\hat{g}^N_\tau = \frac{\hat{R}^N_{y_F u_F}(\tau)}{\hat{\lambda}_N}$$

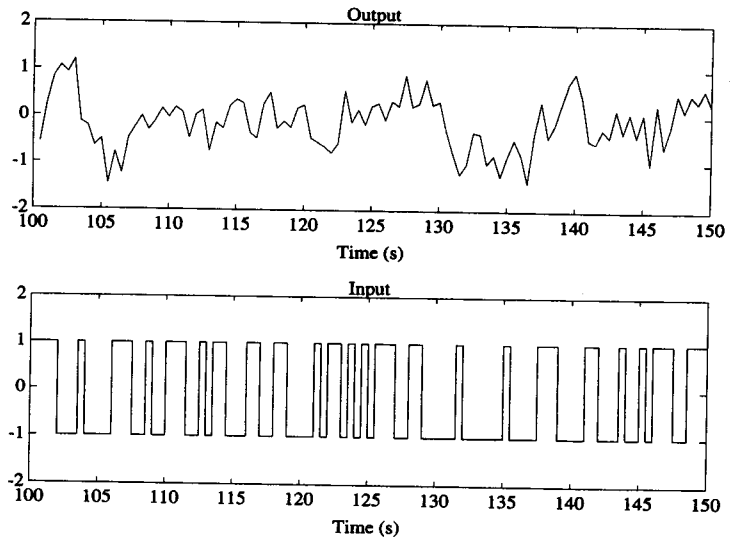


Figure 8.3: Input–output data from the system (8.10).

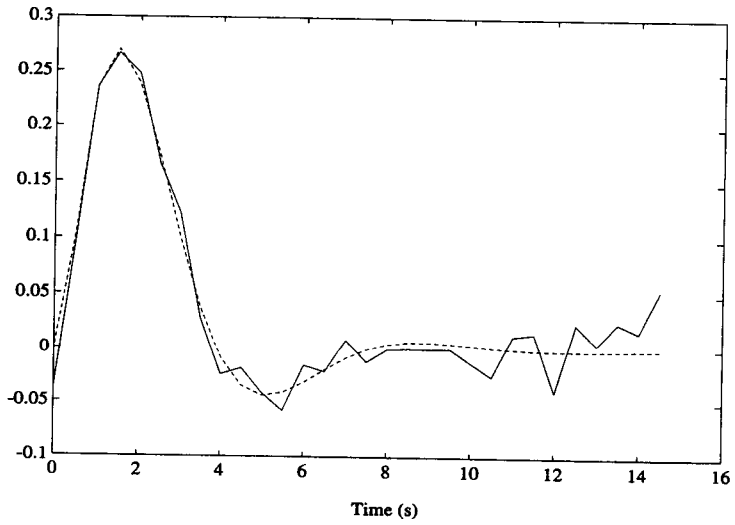


Figure 8.4: Solid line: estimated impulse response. Dashed line: true impulse response of the system (8.10) sampled at 0.5 s. Observe that the curves are obtained by linear interpolation between times 0, 0.5, 1, 1.5, ... s.

Correlation Analysis: Basic Properties

- Like transient analysis, correlation analysis gives a quick insight into time constants and time delays.
- No special inputs are required, and poor signal to noise ratios can basically be compensated for by longer data records
- The result is a table or a graph that cannot be used for simulation directly.
- Correlation analysis, as described here, assumes that the input is uncorrelated with the disturbances [see equation (8.4)]. This means that correlation analysis will not work properly when the data are collected from the system under output feedback.

8.3 Frequency Analysis

A linear system is uniquely determined by its impulse response or its frequency response $G(i\omega)$ (the Laplace transform of the impulse response evaluated at $s = i\omega$). While transient and correlation analysis aim at direct estimates of the impulse response, there are several techniques to directly estimate the frequency response. We will first describe *frequency analysis.*

If a linear system has the transfer function $G(s)$ and the input is

$$u(t) = u_0 \cos \omega t, \tag{8.12}$$

then the output after possible transients have faded away (see Appendix A) will be

$$y(t) = y_0 \cos(\omega t + \varphi) \tag{8.13}$$

where

$$y_0 = |G(i\omega)| \cdot u_0 \tag{8.14}$$

$$\varphi = \arg G(i\omega) \tag{8.15}$$

If the system is driven by the input (8.12) for a certain u_0 and ω_1 and we measure y_0 and φ from the output signal, it is possible to determine the complex number $G(i\omega_1)$ using (8.14)–(8.15). By repeating this

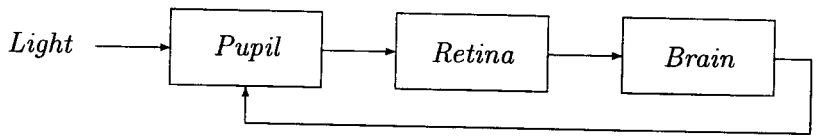


Figure 8.5: Diagram for the adaptation of the pupil

procedure for a number of different ω, we can get a good estimate of the function $G(i\omega)$. This method is called *frequency analysis*. Sometimes it is possible to see or measure u_0, y_0, and φ directly from graphs of the input and output signals. (See for example Figure 8.6.) Most of the time, however, there will be noise and irregularities that make it difficult to determine φ directly. A suitable procedure is then to correlate the output with $\cos\omega t$ and $\sin\omega t$ in the way that is evident later from equation (8.21). This procedure is called *frequency analysis with the correlation method.*

Example 8.3 Eye Dynamics

It is well known that the pupil of an eye reacts to incoming light so that the light intensity at the retina is more or less independent of the outside brightness. We also know that this reaction is not immediate, but dynamic. It takes about a second or so before the eye adapts to new light conditions. Let us construct a model for how the pupil reacts to incoming light. Consider a system according to Figure 8.5. The dynamic properties of this system depend on how the nerve impulses that register the light at the retina are processed and sent to the pupil muscle. They also depend on the reaction time of the pupil muscle. It is not easy to write down reliable equations for this process. Instead we will use an experiment and build a model using frequency analysis. The experiments are described in the work carried out by L. Stark in 1959.

A complication of the experiment in this case is that there is always feedback from the output to the input in the system in Figure 8.5. The

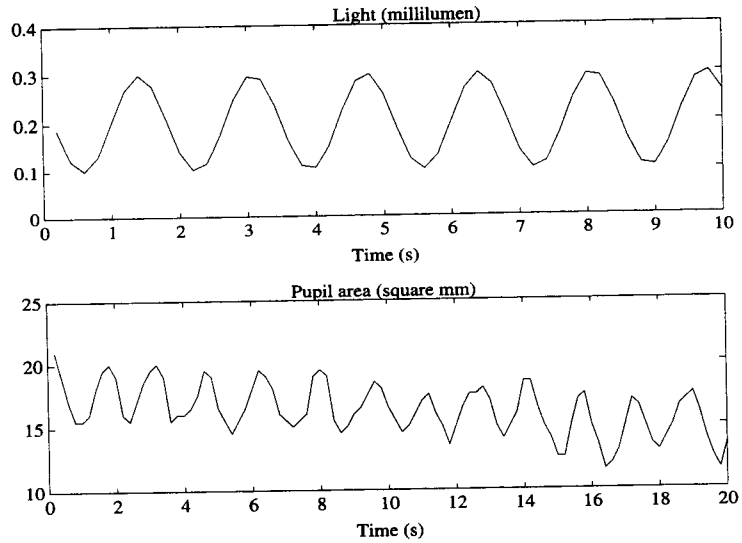


Figure 8.6: Inputs and outputs from pupil experiment.

area of the pupil will affect the light intensity at the retina; this indeed is the purpose of the reaction. To disable the feedback during the experiment, an input ray with a very small area was used and aimed at the center of the pupil. The intensity was then varied as a sinusoid, and the area of the pupil was measured. These measurements were made using a wide infrared light beam, also aimed at the pupil. From the reflected intensity it was possible to compute the area of the pupil. Figure 8.6 shows corresponding inputs and outputs at the frequency $\omega = 3.75$ *rad/s. The output is not exactly sinusoidal, which shows that the system is not exactly linear and/or that errors affect the measurements. It can still be described as a sinusoid with reasonable approximation, and it is not difficult to determine the gain* $|G(i\omega)|$ *and the phase delay* $\arg G(i\omega)$ *from the figure.*

By repeating the experiment for a number of frequencies and graphing $\log|G(i\omega)|$ *and* $\arg G(i\omega)$ *as functions of* $\log\omega$*, the points in the diagram in Figure 8.7 were obtained. These points could then be adjusted to transfer function curves for linear systems. In Figure 8.7 they have been adjusted to the transfer function*

$$G(s) = e^{-0.28s}\frac{0.19}{(1+0.09s)^3} \tag{8.16}$$

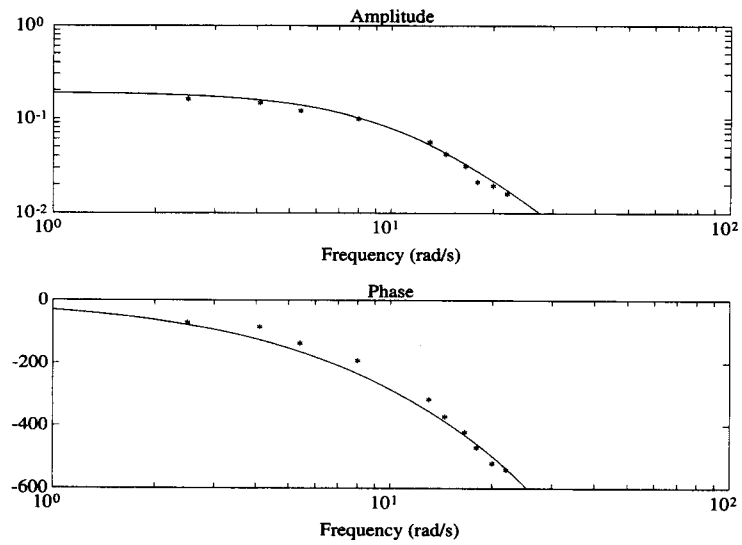


Figure 8.7: Experimentally determined transfer function for the pupil system.

The factor $e^{-0.28s}$ *means that there is a pure time delay of 0.28 s before anything at all happens [the Laplace transform of* $u(t-T)$ *is* $e^{-sT}U(s)$*]. The factor* $\frac{0.19}{(1+0.09s)^3}$ *is a third order system that describes how the output reacts after the time delay. The step response of the model (8.16) is shown in Figure 8.8.*

This model can be physiologically interpreted as a time delay corresponding to the time to transmit and process the information in nerves and synapses, while the third-order system corresponds to the dynamics of the muscle. □

Example 8.4 Submarine dynamics

In the end of the 1940s the Swedish company ASEA constructed a regulator for maintaining depth for Swedish submarines. It was then important to determine the dynamics of the system depicted in Figure 8.9.

Parts of the dynamics from elevator to the actual depth of the submarine are quite difficult to model from physical equations. One reason is that the response of the submarine to changes in this rudder depends

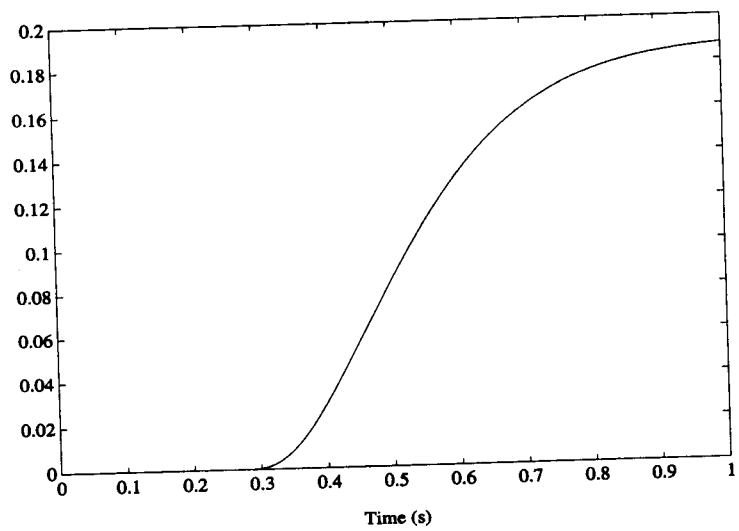


Figure 8.8: Step response for the system (8.16).

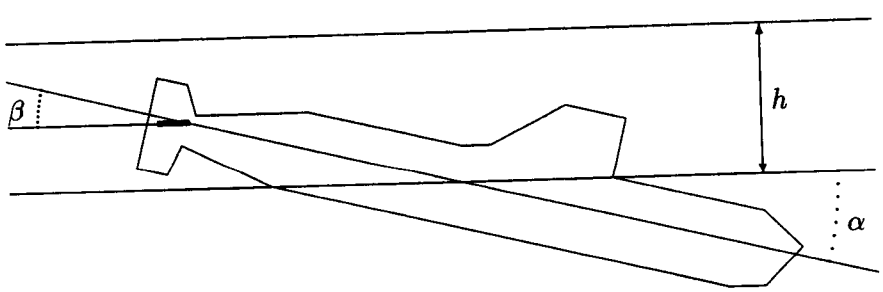


Figure 8.9: Depth control for a submarine.

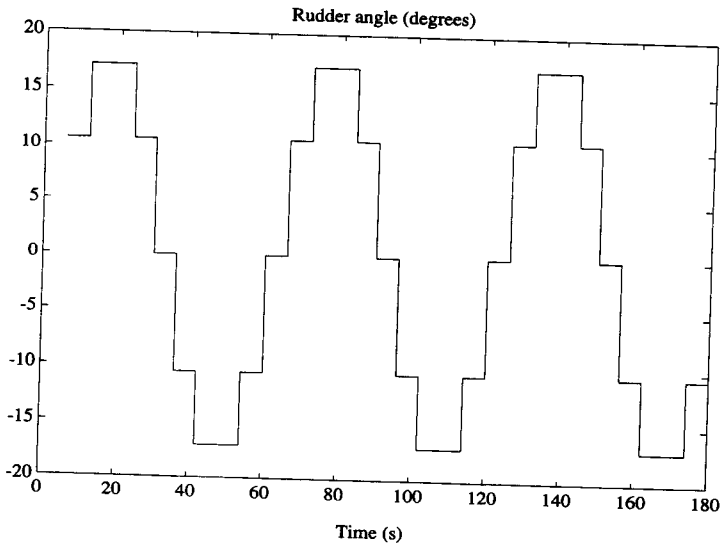


Figure 8.10: The rudder signal during the depth-control experiment.

on its shape in a complicated fashion. A model was thus constructed using frequency analysis. The input (the elevator) was varied in a near sinusoidal way by giving the helmsman orders for the previously computed rudder values. Since the dynamics with reasonable approximation are linear, the submarine traversed the water like a sine wave. The depth was registered and Figures 8.10 and 8.11 show the corresponding values of elevator, pitch angle, and depth. From repeated experiments with different frequencies a plot of the transfer function of the system is obtained as in Figure 8.12. In this case the purpose of the model was to determine a regulator for the system. This was done directly from the diagram in Figure 8.12, and no explicit mathematical description of the model was needed. □

Frequency Analysis. Basic Properties

The method of frequency analysis is often used to build models of systems. The following advantages and disadvantages can be pointed out:

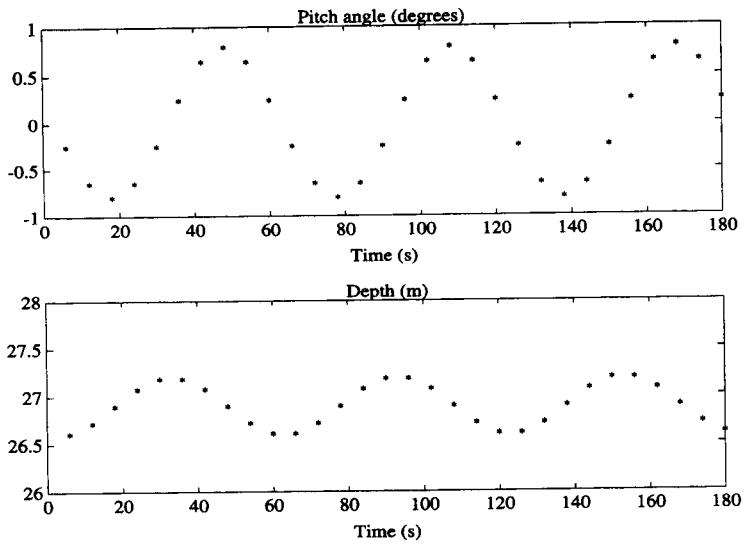


Figure 8.11: Pitch angle and depth of the submarine during the experiment.

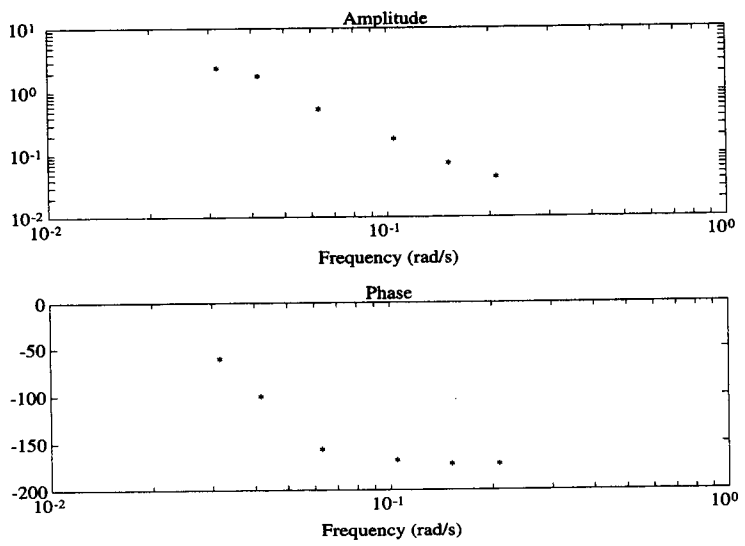


Figure 8.12: Experimentally determined transfer function from rudder angle to pitch angle.

Advantages

- Easy to use and requires no complicated data processing.
- Requires no structural assumptions about the system, other than it being linear.
- It is easy to concentrate on frequency ranges of special interest (for example around resonance frequencies).

Disadvantages

- The basic result is a table or a graph of the function $G(i\omega_k)$ $k = 0, \ldots, M$. This cannot be used directly for simulation.
- Many systems, especially in a process industry, cannot be experimented with freely. Frequency analysis may need long periods of experimentation if $G(i\omega)$ has to be determined at many frequencies.

8.4 Fourier Analysis

Consider a linear system that can be described by the transfer function $G(s)$. If the input has finite energy, the following relationship holds for the Fourier transforms of the input and the output [see (A.4)]:

$$Y(\omega) = G(i\omega)U(\omega) \tag{8.17}$$

If these were known, the frequency function $G(i\omega)$ could thus be *computed* as

$$G(i\omega) = \frac{Y(\omega)}{U(\omega)} \tag{8.18}$$

Normally we have access to $y(t)$, $u(t)$ only over a finite interval $0 \leq t \leq S$. We could then use

$$Y_S(\omega) = \int_0^S y(t)e^{-i\omega t}dt, \quad U_S(\omega) = \int_0^S u(t)e^{-i\omega t}dt \tag{8.19}$$

and form the *estimate*

$$\hat{\hat{G}}_S(i\omega) = \frac{Y_S(\omega)}{U_S(\omega)}. \tag{8.20}$$

We call $\hat{\hat{G}}$ the *empirical transfer function estimate* (ETFE), since it is formed directly from data, without any other model assumptions than linearity.

If $u(t) = u_0 \cos \omega_* t$, we have that

$$U_S(\omega) = \frac{u_0 S}{2} \qquad \text{for } S = \frac{k\pi}{\omega_*}; k = 1, 2, \ldots$$

The estimate (8.20) then is

$$\hat{\hat{G}}_S(i\omega_*) = \frac{2}{u_0 S}\left(\int_0^S y(t)\cos(\omega_* t)dt - i\int_0^S y(t)\sin(\omega_* t)dt\right) \tag{8.21}$$

which is how the frequency function normally is computed using frequency analysis.

If only sampled values of u and y, $(u(kT),\ y(kT),\ k = 1, \ldots, N)$, are available, which is the normal situation, the following approximations in (8.19) are natural:

$$Y_S(\omega) = T\sum_{k=1}^{N} y(kT)e^{-i\omega kT} \tag{8.22a}$$

$$U_S(\omega) = T\sum_{k=1}^{N} u(kT)e^{-i\omega kT} \tag{8.22b}$$

Here T is the sampling interval and $S = N \cdot T$. Note that (8.22) can be efficiently computed at $\omega = r \cdot 2\pi/N$, $r = 0, \ldots, N-1$, using the FFT (fast Fourier transform). N is then first adjusted so that it becomes a power of 2.

Well, how good an estimate is (8.20)? This question is answered by the following result:

Theorem 8.1. Assume that a system is given by

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau + v(t)$$

Assume that

$$|u(t)| \le c_u \quad \text{and} \quad \int_0^\infty \tau|g(\tau)|d\tau = c_g$$

Let

$$G(s) = \int_0^\infty g(\tau)e^{-s\tau}d\tau$$

Then

$$|\hat{\hat{G}}_S(i\omega) - G(i\omega)| \le \frac{2c_u \cdot c_g}{|U_S(\omega)|} + \frac{|V_S(\omega)|}{|U_S(\omega)|} \tag{8.23}$$

Here $\hat{\hat{G}}_S(i\omega)$ is given by (8.19)–(8.20) and $V_S(\omega)$ is the Fourier transform of the disturbance $v(t)$ over the time interval $[0, S]$. The proof is given in the appendix to this chapter.

For a signal with infinite energy the Fourier transform typically has a magnitude

$$|U_S(\omega)| \approx \sqrt{S} \cdot \text{const}$$

If the signal contains a pure sinusoid with frequency ω_0, we have

$$|U_S(\omega_0)| \approx S \cdot \text{const}$$

Theorem 8.1 thus shows that if the input contains pure sinusoids (and the disturbance signals do not) the transfer function will be estimated with arbitrary accuracy at these frequencies, as the time interval tends to infinity. For inputs that do not contain pure sinusoids, estimate (8.20) has an error for large S that is equal to the noise to signal ratio $V_S(\omega)/U_S(\omega)$ at the frequency in question.

The fact that we in practice use (8.22) rather than the time continuous signals in (8.19) gives further discrepancies between the empirical transfer function estimate and the true G, in addition to what is described by (8.23). For short sampling intervals T compared to the system dynamics, this difference is, however, small as we saw in Example 3.14 (see Figure 3.21).

Example 8.5 Empirical Transfer Function Estimates
Consider the same system and the same data as in Example 8.2. (See Figure 8.3.) The ETFE $\hat{\hat{G}}_{500}(e^{i\omega})$ was formed according to (8.20), (8.22) ($T = 0.1$). The result is shown in Figure 8.13. We see that the estimate gives a rather good picture of the frequency function up to about 1 rad/s. At higher frequencies it is not of much use. Compare also the difference between the transfer function for the continuous time system (8.10) and the one for the corresponding sampled data system, sampled with 0.5 s. This was shown in Figure 3.21. □

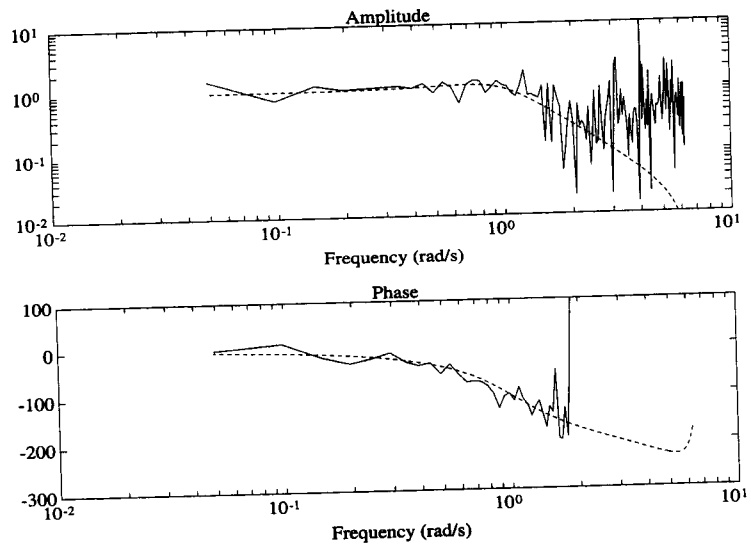


Figure 8.13: Solid line: Empirical transfer function estimation according to (8.20), (8.22). Dashed line: The true transfer function sampled at 0.5 s.

Fourier Analysis: Basic Properties

- Easy and efficient to use (especially if FFT is applied).
- Good estimates of $G(i\omega)$ are obtained at frequencies where the input has pure sinusoids.
- Otherwise, the estimate is a wildly fluctuating graph, which only gives a rough picture of the true frequency function.

8.5 Estimation of Signal Spectra

The spectrum $\Phi_u(\omega)$ for a signal $\{u(t)\}$ was defined in Section 3.8 as its average frequency content. We have formalized $\Phi_u(\omega)$ as the absolute square of the Fourier transform of the input, as a normalized version of this, as well as the expected value of this, depending on the actual nature of the signal. See Appendix C.

To build models from observations of the signal, it is essential to be able to estimate the spectrum from observed, sampled data. Assume

therefore that we have observed

$$u(t), \quad t = T, 2T, \ldots, NT$$

and from these data we want to estimate $\Phi_u(\omega)$. The sampling interval is thus T. From the data we can of course only estimate the spectrum $\Phi_u^{(T)}(\omega)$ for the sampled data signal. If T is small compared to the frequency contents in the underlying time continuous signal, there is not much difference between $\Phi_u^{(T)}(\omega)$ and $\Phi_u^c(\omega)$ according to Poisson's summation result (C.12). See also Section 3.9.

For simplicity we will in the sequel of this section use $T = 1$. The sampling interval is thus *one time unit.* All frequencies will then have the unit *radians/time unit.* When the estimate has been formed it is easy to translate the result to any other time unit: See (8.35).

The Periodogram

Based on the definition of the spectrum it is natural to estimate it as

$$\hat{\Phi}_N(\omega) = \frac{1}{N}|U_N(\omega)|^2 \tag{8.24}$$

where

$$U_N(\omega) = \sum_{k=1}^{N} u(k)e^{-i\omega k} \tag{8.25}$$

This estimate is called the *periodogram* of u. Figure 8.14 shows the periodograms of the signals in Figure 3.15.

Periodograms from measured signals often show the following three typical properties:

1. Pure sinusoids in the signal show up as clear-cut peaks in the periodogram.

2. Otherwise, the periodogram is wildly fluctuating.

3. Smoothing the periodogram by eye gives a reasonable picture of the frequency contents of the signal.

There is reason to understand what causes the fluctuations. One explanation is as follows: Suppose that the signal u is a realization of a

therefore that we have observed

$$u(t), \quad t = T, 2T, \ldots, NT$$

and from these data we want to estimate $\Phi_u(\omega)$. The sampling interval is thus T. From the data we can of course only estimate the spectrum $\Phi_u^{(T)}(\omega)$ for the sampled data signal. If T is small compared to the frequency contents in the underlying time continuous signal, there is not much difference between $\Phi_u^{(T)}(\omega)$ and $\Phi_u^c(\omega)$ according to Poisson's summation result (C.12). See also Section 3.9.

For simplicity we will in the sequel of this section use $T = 1$. The sampling interval is thus *one time unit.* All frequencies will then have the unit *radians/time unit.* When the estimate has been formed it is easy to translate the result to any other time unit: See (8.35).

The Periodogram

Based on the definition of the spectrum it is natural to estimate it as

$$\hat{\Phi}_N(\omega) = \frac{1}{N}|U_N(\omega)|^2 \tag{8.24}$$

where

$$U_N(\omega) = \sum_{k=1}^{N} u(k)e^{-i\omega k} \tag{8.25}$$

This estimate is called the *periodogram* of u. Figure 8.14 shows the periodograms of the signals in Figure 3.15.

Periodograms from measured signals often show the following three typical properties:

1. Pure sinusoids in the signal show up as clear-cut peaks in the periodogram.

2. Otherwise, the periodogram is wildly fluctuating.

3. Smoothing the periodogram by eye gives a reasonable picture of the frequency contents of the signal.

There is reason to understand what causes the fluctuations. One explanation is as follows: Suppose that the signal u is a realization of a

therefore that we have observed

$$u(t), \quad t = T, 2T, \ldots, NT$$

and from these data we want to estimate $\Phi_u(\omega)$. The sampling interval is thus T. From the data we can of course only estimate the spectrum $\Phi_u^{(T)}(\omega)$ for the sampled data signal. If T is small compared to the frequency contents in the underlying time continuous signal, there is not much difference between $\Phi_u^{(T)}(\omega)$ and $\Phi_u^c(\omega)$ according to Poisson's summation result (C.12). See also Section 3.9.

For simplicity we will in the sequel of this section use $T = 1$. The sampling interval is thus *one time unit.* All frequencies will then have the unit *radians/time unit.* When the estimate has been formed it is easy to translate the result to any other time unit: See (8.35).

The Periodogram

Based on the definition of the spectrum it is natural to estimate it as

$$\hat{\Phi}_N(\omega) = \frac{1}{N}|U_N(\omega)|^2 \tag{8.24}$$

where

$$U_N(\omega) = \sum_{k=1}^{N} u(k)e^{-i\omega k} \tag{8.25}$$

This estimate is called the *periodogram* of u. Figure 8.14 shows the periodograms of the signals in Figure 3.15.

Periodograms from measured signals often show the following three typical properties:

1. Pure sinusoids in the signal show up as clear-cut peaks in the periodogram.

2. Otherwise, the periodogram is wildly fluctuating.

3. Smoothing the periodogram by eye gives a reasonable picture of the frequency contents of the signal.

There is reason to understand what causes the fluctuations. One explanation is as follows: Suppose that the signal u is a realization of a

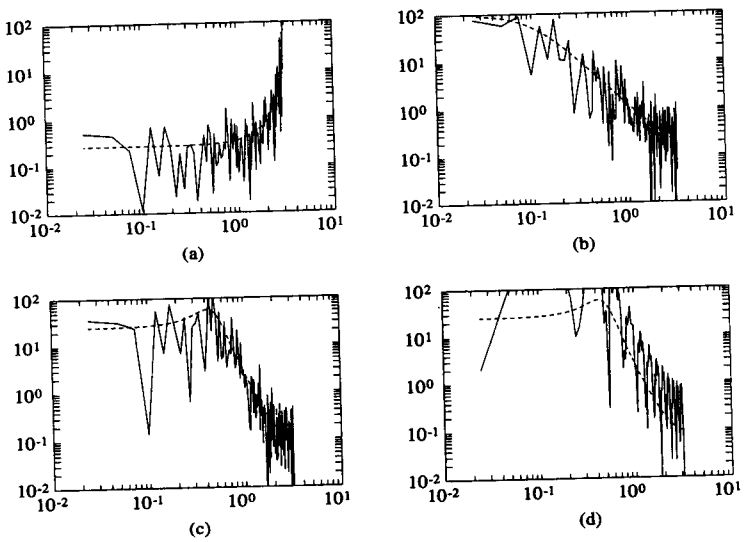


Figure 8.14: Solid lines: periodogram estimates of spectra for the signals in Figure 3.15. Dashed lines: true spectra (compare with Figure 3.17).

stochastic process with the spectral density $\Phi_u(\omega)$. The periodogram $\hat{\hat{\Phi}}_N$ is formed from $u(t), t = 1, \ldots, N$, and thus is itself a stochastic variable with a certain mean and variance. In fact we have

$$E\hat{\hat{\Phi}}_N(\omega) = \Phi_u(\omega) + R_N^{(1)} \tag{8.26}$$

$$E(\hat{\hat{\Phi}}_N(\omega) - \Phi_u(\omega))^2 = (\Phi_u(\omega))^2 + R_N^{(2)} \tag{8.27}$$

$$E(\hat{\hat{\Phi}}_N(\omega_1) - \Phi_u(\omega_1))(\hat{\hat{\Phi}}_N(\omega_2) - \Phi_u(\omega_2)) = R_N^{(3)} \tag{8.28}$$

$$|\omega_1 - \omega_2| \geq \frac{2\pi}{N}$$

$$R_N^{(i)} \to 0 \text{ as } N \to \infty \qquad i = 1, 2, 3$$

Here $R_N^{(i)}$ denote remainder terms that tend to zero as the number of data increases. E is expectation with respect to the stochastic process u. These calculations are carried out in the appendix to this chapter.

We see from these expressions that for large N the periodogram is unbiased, but its variance does not decay to zero as N increases, but is proportional to $\Phi_u^2(\omega)$. Furthermore, the estimates of the periodogram

at different frequencies are uncorrelated. The periodogram for a realization of stochastic processes is thus a fluctuating graph, which varies around the true spectral density with a standard deviation that is as large as the spectral density itself. The observation that periodograms for many real-life signals are of this nature can be seen as an indication that stochastic processes form a good mathematical model for such signals.

Frequency resolution is another concept of importance for spectral estimates. By this is meant how fine details in a spectrum can be detected by a certain method. In principle we can reason as follows: To see a frequency (or a frequency difference) that is less than Δ radians/second, it must have had an opportunity to "show itself." This means that we must observe the signal during at least $2\pi/\Delta$ seconds, so that it has gone through at least a full cycle. For the periodogram based on N data we have that the *frequency resolution is* $2\pi/N$. This follows from the fact that the Fourier transform (8.25) gives the discrete Fourier transform (DFT) at the frequency values $\omega = 2\pi\ell/N, \quad \ell = 1, \ldots, N$. Between these frequencies the Fourier transform consists of (trigonometrically) interpolated values.

The problem with the periodogram is that it has a high variability, while the frequency resolution is good. We will now examine different methods to reduce the variance of the spectral estimate. The price, however, will turn out to be worse frequency resolution.

Averaged Periodograms: Welch's Method

A classical way to reduce the variance of an estimate is to form averages of a number of independent estimates. In our case an obvious idea would be to split the signal into a number, say R, of segments, each of length M and possibly overlapping. For each of these, periodograms $\hat{\hat{\Phi}}_M^{(k)}(\omega)$, $k = 1, \ldots, R$ are formed. The spectral estimate is then taken as the average of these periodograms:

$$\hat{\Phi}_N(\omega) = \frac{1}{R} \sum_{k=1}^{R} \hat{\hat{\Phi}}_M^{(k)}(\omega) \tag{8.29}$$

By selecting the length of the segments to be powers of 2, the calculations of the periodograms can be efficiently done using FFT.

Since the different periodograms are essentially uncorrelated (if segments do not overlap) the variance of $\hat{\Phi}_N(\omega)$ in (8.29) is reduced by a factor R. The price for this is worse frequency resolution in the estimate. According to our earlier discussion, it will increase from $1/N$ radians/time unit (N = the original data record length) to $1/M = R/N$ radians/time unit ($M = \frac{N}{R}$ = the length of the nonoverlapping segments). The trade-off between variance and frequency resolution is thus determined by the number of segments R.

Estimating spectra using (8.29) is also known as *Welch's method.*

Smoothed Periodograms: Blackman-Tukey's Method

We noted earlier that the periodograms are unbiased with uncorrelated estimates for neighboring frequencies. It is therefore natural to smooth the fluctuating graph of a periodogram by averaging over a number of neighboring frequencies:

$$\hat{\Phi}_N(\omega) = \int_{-\pi}^{\pi} W_\gamma(\omega - \xi)\hat{\hat{\Phi}}_N(\xi)d\xi \tag{8.30}$$

[Here we have that $\int_{-\pi}^{\pi} W_\gamma(\omega)d\omega = 1$.] Here $W_\gamma(\omega)$ is a window function that typically is centered around $\omega = 0$. We shall use the number γ to describe the "width" of the window. For reasons that soon will be clear, we will let the width be inversely proportional to γ. A simple window with width $1/\gamma$ would be rectangular:

$$W_\gamma(\xi) = \begin{cases} \gamma, \text{ if } |\xi| < \frac{1}{2\gamma} \\ 0, \text{ else} \end{cases}$$

The width of the window then corresponds to the frequency resolution of the smoothed estimate $\hat{\Phi}_N(\omega)$. Normally, other windows than the rectangular one will be used. In that way more weight can be given to the center value. See Figure 8.15 and equations (8.36)–(8.38) to follow.

The actual algorithm to realize (8.30) is best implemented in the time domain. In the appendix to this chapter it is shown that (8.30) also can be written as

$$\hat{\Phi}_N(\omega) = \sum_{k=-\gamma}^{\gamma} w_\gamma(k)\hat{R}_u^N(k)e^{-i\omega k} \tag{8.31}$$

where

$$w_\gamma(k) = \int_{-\pi}^{\pi} W_\gamma(\xi) e^{i\xi k} d\xi \tag{8.32}$$

and

$$\hat{R}_u^N(k) = \frac{1}{N} \sum_{t=1}^{N} u(t+k)u(t) \tag{8.33}$$

This spectral estimation method is known as the The *Blackman-Tukey approach.* The procedure can be summarized as follows:

Procedure: Blackman–Tukey's Spectral Estimate (8.34)

1. Choose time window $w_\gamma(k)$. See for example (8.36).
2. Choose window size γ (discussed later).
3. Compute $\hat{R}_u^N(k)$ for $k = 0, \ldots, \gamma$ according to (8.33).
4. Form $\hat{\Phi}_N(\omega)$ as in (8.31).

In the preceding expressions we have assumed that $w_\gamma(k)$ is chosen so that it is zero for $|k| > \gamma$. This means that there are special requirements of the choice of window $W_\gamma(\xi)$. (A rectangular window would be impossible for this reason.) In (8.33) we have also assumed that $u(t) = 0$ when t is outside the interval $[1, N]$.

Transformations to Correct Sampling Interval

In the preceding expressions we have assumed the sampling interval to be one time unit. The frequency unit ω is thus radians/sampling interval. The spectral density $\hat{\Phi}_N(\omega)$ has the dimension power per frequency unit and we have used the unit power · sampling interval/radian.

Suppose now that the sample interval is T seconds. To express the frequency ξ in the unit radians/second and the spectral density $\hat{\Phi}_N^0(\omega)$ in the unit power per radians/second we simply do the following:

$$\hat{\Phi}_N^0(\xi) = T \cdot \hat{\Phi}_N(\xi \cdot T), \qquad \frac{-\pi}{T} \le \xi \le \frac{\pi}{T} \tag{8.35}$$

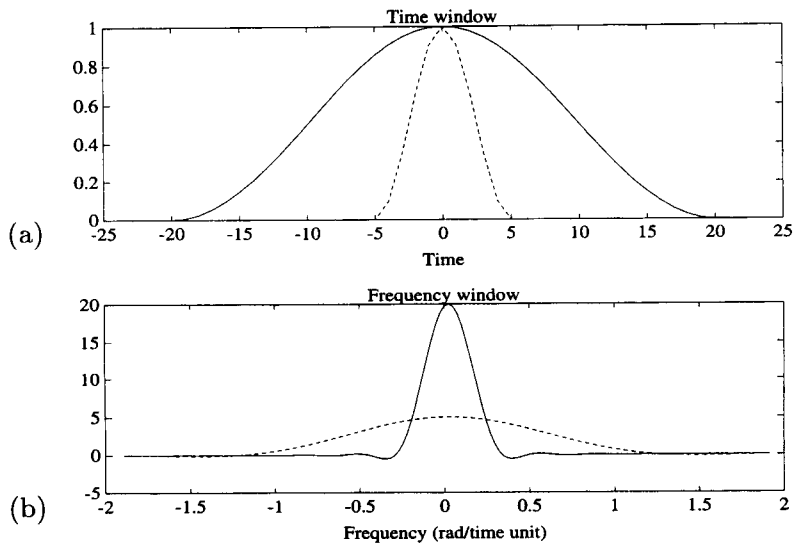


Figure 8.15: The Hamming window. (a) The time window $w_\gamma(k)$. (b) Its Fourier transform $W_\gamma(\omega)$. Solid line: $\gamma = 20$. Dashed line: $\gamma = 5$.

Choice of Window Functions

The Fourier transform pair $w_\gamma(k) \leftrightarrow W_\gamma(\xi)$ determines the properties of the spectral estimate. For the given width γ of the time window $w_\gamma(k)$ it is desirable to have as "nice" (narrow and high) a frequency function $W_\gamma(\xi)$ as possible. There is no optimal solution to this problem, but the most common window used in spectral analysis is the *Hamming window*:

$$\begin{array}{ll} w_\gamma(k) = \frac{1}{2}(1 + \cos\frac{\pi k}{\gamma}) & |k| < \gamma \\ w_\gamma(k) = 0 & |k| \geq \gamma \end{array} \tag{8.36}$$

Its Fourier transform is

$$W_\gamma(\omega) = \frac{1}{2}D_\gamma(\omega) + \frac{1}{4}(D_\gamma(\omega - \frac{\pi}{\gamma}) + D_\gamma(\omega + \frac{\pi}{\gamma})) \tag{8.37}$$

where

$$D_\gamma(\omega) = \frac{\sin(\gamma + \frac{1}{2})\omega}{\sin\omega/2} \tag{8.38}$$

See Figure 8.15. The effective width of the frequency window, which

gives the frequency resolution, can be measured as

$$(\int_{-\pi}^{\pi} \xi^2 W_\gamma(\xi) d\xi)^{1/2} = M(\gamma) \tag{8.39}$$

For the Hamming window we have

$$M(\gamma) \approx \frac{1}{\sqrt{2}} \cdot \frac{\pi}{\gamma} \tag{8.40}$$

which is, naturally enough, inversely proportional to the width of the time window γ.

In the periodogram the frequency resolution is π/N radians/time unit. Within the indicated window width we consequently have

$$\frac{M(\gamma)}{\pi/N} \approx \frac{\pi \cdot N}{\pi\sqrt{2} \cdot \gamma} = \frac{N}{\gamma\sqrt{2}}$$

independent periodogram estimates, and therefore the variance in the average quantity (8.30) is reduced by about the same factor. Then, according to (8.40) and (8.27) this variance will be $\approx \sqrt{2}\gamma \cdot \frac{1}{N}(\Phi_u(\omega))^2$. In summary, we consequently have the following:

1. The frequency resolution of $\hat{\Phi}_N(\omega)$ is $\approx \pi/(\gamma\sqrt{2})$ radians/sampling interval.

(8.41)

2. The variance of $\hat{\Phi}_N(\omega)$ is $\approx \sqrt{2} \cdot \frac{\gamma}{N}(\Phi_u(\omega))^2$.

Choice of Window Size

The choice of γ according to (8.41) is a pure trade-off between frequency resolution and variance (variability). For a spectrum with narrow resonance peaks it is thus necessary to choose a large value of γ and accept a higher variance. For a more flat spectrum, smaller values of γ will do well. In practice a number of different values of γ are tried out. Often we start with a small value of γ and increase it successively until a spectrum is found that balances the trade-off between frequency resolution (true details) and variance (random fluctuations). A typical value for spectra without narrow resonances is γ= 20–30.

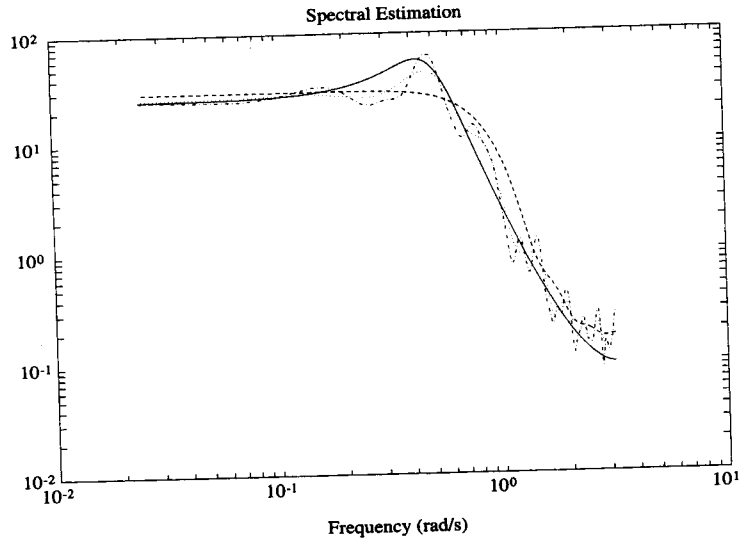


Figure 8.16: Spectral estimation by the method (8.34) for data in Figure 3.15c. Solid line: true spectrum. Dashed line: $\gamma = 10$. Dotted line: $\gamma = 30$. Dash–dotted line: $\gamma = 60$.

Example 8.6 Spectral Estimates
The Blackman–Tukey procedure was applied to data in Figure 3.15c. A number of different values of γ were tested and the results are shown in Figure 8.16. We see that the resonance peak is entirely lost for $\gamma = 10$, but for larger values of γ a reasonable picture of the true spectrum is obtained. □

Cross Spectra

Estimation of cross spectra is entirely analogous to the procedure of estimating spectra that we have just described. From sampled values

$$y(k) \quad \text{and} \quad u(k) \quad k = 1, \ldots, N$$

the cross-covariance function is formed as

$$\hat{R}_{yu}^{N}(\tau) = \frac{1}{N} \sum_{k=1}^{N} y(k)u(k-\tau) \tag{8.42}$$

from which the spectral estimate is computed:

$$\hat{\Phi}_{yu}^N(\omega) = \sum_{\ell=-\gamma}^{\gamma} \hat{R}_{yu}^N(\ell) \cdot w_\gamma(\ell) e^{-i\ell\omega} \tag{8.43}$$

Here $w_\gamma(\ell)$ is the same window function as in (8.31), and the same aspects for choosing γ are valid also in this case.

8.6 Estimating Transfer Functions Using Spectral Analysis

Suppose that we have the following relationship between output y, input u, and disturbance v:

$$y(t) = G(p)u(t) + v(t) \tag{8.44}$$

If $u(t)$ and $v(t)$ are mutually uncorrelated, the following expressions hold for spectra and cross spectra [See (3.63)–(3.64)]:

$$\Phi_{yu}(\omega) = G(i\omega)\Phi_u(\omega) \tag{8.45}$$

$$\Phi_y(\omega) = |G(i\omega)|^2\Phi_u(\omega) + \Phi_v(\omega) \tag{8.46}$$

From the spectral estimates (8.31) and (8.43), we then have a natural estimate of the frequency function (8.45):

$$\hat{G}_N(i\omega) = \frac{\hat{\Phi}_{yu}^N(\omega)}{\hat{\Phi}_u^N(\omega)} \tag{8.47}$$

Furthermore, the disturbance spectrum can be estimated from (8.46) as

$$\hat{\Phi}_v^N(\omega) = \hat{\Phi}_y^N(\omega) - \frac{|\hat{\Phi}_{yu}^N(\omega)|^2}{\hat{\Phi}_u^N(\omega)} \tag{8.48}$$

To compute these estimates, the following steps are carried out:

Algorithm SPA (8.49)

1. Collect data $y(k)$, $u(k)$ $k = 1, \ldots, N$. Subtract the corresponding sample means.
2. Choose width of the lag window $w_\gamma(k)$.
3. Compute $\hat{R}_y^N(k)$, $\hat{R}_u^N(k)$, and $\hat{R}_{yu}^N(k)$ for $|k| \leq \gamma$ according to (8.33).
4. Form the spectral estimates $\hat{\Phi}_y^N(\omega)$, $\hat{\Phi}_u^N(\omega)$, and $\hat{\Phi}_{yu}^N(\omega)$ according to (8.31) and analogous expressions.
5. Form (8.47) and possibly also (8.48).

The user only has to choose γ. A good value for systems without sharp resonances is $\gamma = 20$ to 30. This may have to be modified according to the discussion in Section 8.5. Larger values of γ may be required for systems with narrow resonances.

From sampled data it is actually the sampled frequency function $G_T(e^{i\omega T})$ that is estimated [see (3.80)]. With a suitable choice of sampling interval this function does not differ very much from $G(i\omega)$ in the frequency region of interest [compare (3.81)]. Furthermore, experience shows that the estimate $\hat{G}_N(i\omega)$ is unreliable at high frequencies anyway.

We may also note that with $\gamma = N$ we essentially obtain $\hat{\Phi}_u^N(\omega) = |U_N(\omega)|^2$ and $\hat{\Phi}_{yu}^N(\omega) = Y_N(\omega)\overline{U_N(\omega)}$ (the periodogram estimates), which gives $\hat{G}_N(i\omega) = Y_N(\omega)/U_N(\omega)$, that is, the empirical transfer function estimate (8.20).

Quality of the Estimates

The estimates $\hat{G}_N$ and $\hat{\Phi}_w^N$ are formed entirely from estimates of spectra and cross spectra. Their properties will therefore be inherited from the properties of the spectral estimates, as they were summarized in (8.41). The corresponding results also hold for cross spectra.

For the Hamming window with width γ, the frequency resolution

will be about

$$\frac{\pi}{\gamma\sqrt{2}} \qquad \text{radians/time unit} \tag{8.50}$$

This means that details in the true frequency function that are finer than this expression will be smeared out in the estimate. It is also possible to show that the estimate's variances satisfy

$$\text{Var } \hat{G}_N(i\omega) \approx 0.7 \cdot \frac{\gamma}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)} \tag{8.51}$$

and

$$\text{Var } \hat{\Phi}_v^N(\omega) \approx 0.7 \cdot \frac{\gamma}{N} \cdot \Phi_v^2(\omega) \tag{8.52}$$

[Variance" here refers to taking expectation over the noise sequence $v(t)$.] Note that the relative variance in (8.51) typically increases dramatically as ω tends to the Nyquist frequency. The reason is that $|G(i\omega)|$ typically decays rapidly, while the noise-to-signal ratio $\Phi_v(\omega)/\Phi_u(\omega)$ has a tendency to increase as ω increases. In a Bode diagram the estimates will thus show considerable fluctuations at high frequencies. Moreover, the constant frequency resolution (8.50) will look thinner and thinner at higher frequencies in a Bode diagram due to the logarithmic frequency scale.

Example 8.7 Estimating a Frequency Function with Spectral Analysis

Consider again the system and data from Example 8.2. The procedure SPA applied to these data with different values of γ will give estimates of the system's frequency function, which are shown in Figure 8.17. We see that $\gamma = 10$ is too small a value, in that the (modest) resonance peak does not show up at all. The values $\gamma = 30$ and 60 give a good picture of the frequency function up to 2 to 3 rad/s. Compare with the empirical transfer function estimate in Figure 8.13! □

Spectral Analysis: Summary

- Spectral analysis is a very common method for analysis of signals and systems.
- It is general, assuming only that the system is linear, and requires no specific input signals.

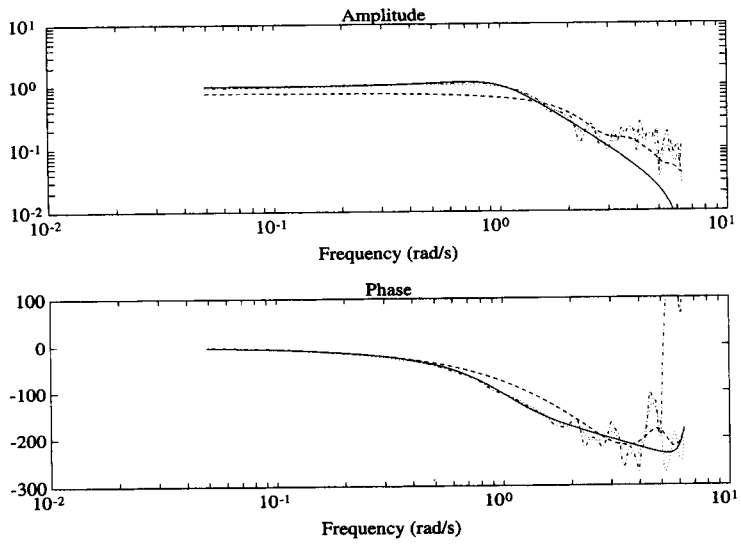


Figure 8.17: Spectral estimation of frequency functions according to the procedure SPA for different values of γ. Solid line: the true frequency function for the system sampled with $T = 0.5$. Dashed line: $\gamma = 10$. Dotted line: $\gamma = 30$. Dash–dotted: $\gamma = 60$.

- Spectral analysis does not work for systems that operate under output feedback (when the input is partly determined by old outputs). The reason is that the assumption that u and v are uncorrelated is violated in that case; that is, (8.45)–(8.46) do not hold. See Example 10.1.

- After adjustment of the window size γ, it is usually possible to get a good picture of the system's or signal's frequency properties.

- The result is a plot, the frequency function, or the spectrum, and it cannot be used directly for simulation.

- The method should be complemented with parametric modeling as described in Chapter 9.

8.7 Summary

The nonparametric identification methods of this chapter aim at direct estimates of the system's transient response or frequency response. As such they are very valuable initially when the structure of a model is not yet clear. Simple step and impulse response experiments give important insight into relationships between measured variables, their static relationships, and dominating time constants.

Spectral analysis is the most commonly used method for estimating frequency functions. Fourier analysis can be viewed as a special case (very wide lag windows), which in turn contains frequency analysis as a special case (sinusoidal input). The procedure for spectral analysis is summarized in the algorithm SPA in Section 8.6. The most essential influence by the user over the estimates is the choice of window size γ. This will determine the trade-off between frequency resolution and variability of the estimate. See (8.50)–(8.52). Reasonable choices of γ give good, but preliminary, insight into the dominating frequency properties of the system.

8.8 Appendix

Proof of Theorem 8.1

From (8.19) and the expression for $y(t)$ in the theorem we have

$$\begin{aligned} Y_S(\omega) &= \int_0^S y(t)e^{-i\omega t}dt \\ &= \int_{\tau=0}^{\infty}\int_{t=0}^{S} g(\tau)u(t-\tau)d\tau e^{-i\omega t}dt + V_S(\omega) = [t-\tau=\sigma] \\ &= \int_{\tau=0}^{\infty} g(\tau)e^{-i\tau\omega}\left[\int_{\sigma=-\tau}^{S-\tau} u(\sigma)e^{-i\sigma\omega}d\sigma\right]d\tau + V_S(\omega) \end{aligned}$$

Now we have

$$\begin{aligned} |U_S(\omega) - \int_{\sigma=-\tau}^{S-\tau} u(\sigma)e^{-i\sigma\omega}d\sigma| &\le |\int_{S-\tau}^{S} u(\sigma)e^{-i\sigma\omega}d\sigma| + |\int_{-\tau}^{0} u(\sigma)e^{-i\sigma\omega}d\sigma| \\ &\le 2\cdot\tau\cdot c_u \end{aligned}$$

Inserting this into the preceding expression we obtain

$$|Y_S(\omega) - G(i\omega)U_S(\omega)| \leq |\int_{\tau=0}^{\infty} |g(\tau)e^{-i\tau\omega}| \cdot 2 \cdot \tau c_u d\tau| + |V_S(\omega)|$$
$$\leq 2 \cdot c_u \cdot c_g + |V_S(\omega)|$$

Proof of (8.26)–(8.28)

We introduce the assumption that

$$\sum_{\tau=-\infty}^{\infty} |\tau R_u(\tau)| = c_u < \infty$$

Let $U_N(\omega)$ be defined by (8.25), and let us compute

$$\begin{aligned}
E\frac{1}{N}U_N(\omega)U_N(-\xi) &= \frac{1}{N}E\sum_{k=1}^{N} u(k)e^{-ik\omega} \cdot \sum_{\ell=1}^{N} u(\ell)e^{i\ell\xi} \\
&= \frac{1}{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} Eu(k)u(\ell)e^{i(\ell\xi - k\omega)} \\
&= \frac{1}{N}\sum_{k=1}^{N}\sum_{\ell=1}^{N} R_u(k-\ell)e^{-i\xi(k-\ell)} \cdot e^{i(\xi-\omega)k} \\
&= \begin{bmatrix} \text{make the change of variables} \\ k - \ell = \tau \end{bmatrix} \\
&= \frac{1}{N}\sum_{k=1}^{N} e^{i(\xi-\omega)k} \sum_{\tau=k-N}^{k-1} R_u(\tau)e^{-i\tau\xi}
\end{aligned}$$

Now note the following:

$$\frac{1}{N}\sum_{k=1}^{N} e^{i(\xi-\omega)k} = \begin{cases} 1, \text{ if } \xi = \omega \\ 0, \text{ if } |\xi - \omega| = \frac{2\pi r}{N}, \quad r = 1, 2, \ldots \end{cases}$$

This is easily seen by using the expression for the sum of the geometric series. We also have

$$\sum_{\tau=k-N}^{k-1} R_u(\tau)e^{-i\tau\xi} = \Phi_u(\xi) - \sum_{\tau=-\infty}^{k-N-1} R_u(\tau)e^{-i\xi\tau} - \sum_{\tau=k}^{\infty} R_u(\tau)e^{-i\xi\tau}$$

Now consider

$$|\frac{1}{N}\sum_{k=1}^{N} e^{i(\xi-\omega)k} \cdot \sum_{\tau=-\infty}^{k-N-1} R_u(\tau)e^{-i\xi\tau}|$$

$$\leq \frac{1}{N}\sum_{k=1}^{N}\sum_{\tau=-\infty}^{k-N-1} |R_u(\tau)| \leq \begin{bmatrix}\text{change order}\\ \text{of summation}\end{bmatrix} \leq$$

$$\leq \frac{1}{N}\sum_{\tau=-\infty}^{-1} |\tau| \cdot |R_u(\tau)| \leq \frac{1}{N}c_u$$

In the same way it is shown that

$$|\frac{1}{N}\sum_{k=1}^{N} e^{i(\xi-\omega)k} \cdot \sum_{\tau=k}^{\infty} R_u(\tau)e^{-i\xi\tau}| \leq \frac{1}{N}c_u$$

All this together gives

$$E\frac{1}{N}U_N(\omega)U_N(-\xi) = \begin{cases} \Phi_u(\omega) + R_N^{(1)} & \text{if } \xi = \omega \\ R_N^{(1)} & \text{if } |\xi - \omega| = \frac{k2\pi}{N}, k = 1, 2, \ldots \end{cases} \tag{8.53}$$

$$|R_N^{(1)}| \leq \frac{c_u}{N} \tag{8.54}$$

In particular we obtain the result (8.26).

For (8.27)–(8.28) we restrict ourselves to the special case that $\{u\}$ is a Gaussian process. This also means that Re $[U_N(\omega)]$ and Im $[U_N(\omega)]$ are Gaussian random variables (since they are formed as sums of Gaussian variables). With

$$\text{Re}[U_N(\omega)] = \frac{1}{2}\left(U_N(\omega) + \overline{U_N(\omega)}\right) = \frac{1}{2}(U_N(\omega) + U_N(-\omega))$$

and

$$\text{Im}[U_N(\omega)] = \frac{1}{2}\left(U_N(\omega) - U_N(-\omega)\right)$$

it is easy to verify, using (8.53), that

$$E\frac{1}{N}(\text{Re}[U_N(\omega)])^2 = \frac{1}{2}\Phi_u(\omega) + R_N^{(1)}$$

$$E\frac{1}{N}(\mathrm{Im}[U_N(\omega)])^2 = \frac{1}{2}\Phi_u(\omega) + R_N^{(1)}$$

$$E\frac{1}{N}\mathrm{Re}[U_N(\omega)] \cdot \mathrm{Im}[U_N(\omega)] = R_N^{(1)}$$

$$|R_N^{(1)}| \leq \frac{c_u}{N}$$

This means that

$$Z_N = \frac{2\hat{\hat{\Phi}}_N(\omega)}{\Phi_u(\omega)} = \frac{2}{N\Phi_u(\omega)}\left\{(\mathrm{Re}[U_N(\omega)])^2 + (\mathrm{Im}[U_N(\omega)])^2\right\}$$

is the sum of squares of two (up to R_N) independent Gaussian variables. Therefore, Z_N is $\chi^2(2)$ and has a $\chi^2(2)$ distribution with the variance 4 (up to R_N). Consequently, (8.27) holds.

From (8.53) it also follows that the periodogram is independent (up to R_N) for different frequencies, which gives (8.28):

$$E(\hat{\hat{\Phi}}_N(\omega_1) - \Phi_u(\omega_1))(\hat{\hat{\Phi}}_N(\omega_2) - \Phi_u(\omega_2))$$

$$= E(\hat{\hat{\Phi}}_N(\omega_1) - \Phi_u(\omega_1)) \cdot E(\hat{\hat{\Phi}}_N(\omega_2) - \Phi_u(\omega_2)) \approx R_N$$

Remark. Our heuristic way to treat "up to R_N" is formalized by projecting away the dependent part of $\mathrm{Im}[U_N(\omega)]$ (and similarly in the other expressions). Its magnitude is bounded by R_N and therefore it cannot change the result by a value larger than this.]

Proof of (8.31)

We start from (8.30):

$$\hat{\Phi}_N(\omega) = \int_{-\pi}^{\pi} W_\gamma(\omega - \xi)\hat{\hat{\Phi}}_N(\xi)d\xi$$

This is a convolution between the functions $W_\gamma(\omega)$ and $\hat{\hat{\Phi}}_N(\omega)$. It will then, according to the rules for Fourier transformation, be written as the Fourier transform of the product of these two functions' inverse Fourier transforms:

$$\hat{\Phi}_N(\omega) = \sum_{k=-\infty}^{\infty} (w_\gamma(k) \cdot \phi_N(k)) \cdot e^{-i\omega k}$$

Here $w_\gamma(k)$ is defined by (8.32) and $\phi_N(k)$ is the inverse transform of $\hat{\hat{\Phi}}_N(\omega)$. But [let $u(k) = 0$ for $k < 1$ and $k > N$]

$$\begin{aligned}
\hat{\hat{\Phi}}_N(\omega) &= \frac{1}{N}|U_N(\omega)|^2 = \frac{1}{N}\sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} u(k)u(\ell)e^{-i(k-\ell)\omega} \\
&= \begin{bmatrix} \text{change} \\ k-\ell=\tau \end{bmatrix} = \frac{1}{N}\sum_{\tau=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} u(k)u(k-\tau)e^{-i\tau\omega} \\
&= \sum_{\tau=-\infty}^{\infty}\left(\frac{1}{N}\sum_{k=-\infty}^{\infty} u(k)u(k-\tau)\right)\cdot e^{-i\tau\omega} = \sum_{\tau=-\infty}^{\infty}\hat{R}_u^N(\tau)e^{-i\tau\omega}
\end{aligned}$$

This means that $\phi_N(k) = \hat{R}_u^N(k)$ according to (8.33) and consequently (8.31) has been proved.

Chapter 9

Parameter Estimation in Dynamic Models

It is a well-known statistical problem to estimate parameters in different models. Such methods are also of great importance when building models of dynamic systems. The lack of knowledge about specific details or about general subsystems in the system will then be represented using parameters, whose numerical values have to be determined by statistical methods. In this chapter we shall give a basic account of such methods.

There are two different kinds of parameterized models:

1. **Tailor-made models**, for which the model is constructed from basic physical principles and the parameters represent unknown values of system parameters that, at least in principle, have a physical interpretation.

2. **Ready-made models** are families of flexible models of general applicability. The parameters in such models have no direct physical interpretation, but are only used as vehicles to describe the properties of the input–output relationships of the system. Such models are also known as *black-box models*.

9.1 Tailor-made Models

When the relationships between the variables in the systems are written down in phase 2 of the modeling process, it is normally found that a number of system constants have unknown values. It could for example be the outlet area of the head box in equation (4.11) or the moment of inertia of the electric motor in Example 6.1.

The resulting state-space model (4.1) will be of the form

$$\frac{d}{dt}x(t) = f(x(t), u(t), \theta) \tag{9.1a}$$

$$y(t) = h(x(t), u(t), \theta) \tag{9.1b}$$

where the parameter vector θ contains the unknown system parameters. If we have d such parameters, this vector can be written as

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_d \end{pmatrix} \tag{9.2}$$

Our task is now to determine the values of these parameters using measured data. In some cases this can be done by conventional physical experimentation and measurement methods. This approach of determining the parameter values will not be discussed further here, since it is entirely problem dependent and goes back directly to measurement technology and physics.

Example 9.1 A DC Motor

Consider the electric motor in Example 6.1 and Figure 6.20. The input u is the applied voltage v and the output y is the angular position of the motor shaft. Moreover ω is the angular velocity; that is, $\frac{d}{dt}y(t) = \omega$. From the bond-graph in Figure 6.20 we obtain

$$u(t) = R_1 i(t) + L_1 \frac{d}{dt} i(t) + k\omega(t)$$

$$J \cdot \frac{d}{dt}\omega(t) = ki(t) - f\omega$$

where we assumed the function $\varphi(\omega)$ to be linear: $\varphi(\omega) = f\omega$.

Neglecting the inductance L_1 and introducing the state variables

$$x(t) = \begin{bmatrix} y(t) \\ \omega(t) \end{bmatrix}$$

gives, after some calculations,

$$\frac{d}{dt}x(t) = \begin{bmatrix} 0 & 1 \\ 0 & -1/\tau \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \beta/\tau \end{bmatrix} u(t) \tag{9.3}$$

$$y(t) = [1 \quad 0]x(t)$$

where

$$\tau = \frac{JR_1}{fR_1 + k^2} \qquad \beta = \frac{k}{fR_1 + k^2} \tag{9.4}$$

In the modeling procedure we have used four different system constants: J, R_1, k, and f. We see that the dynamics of the system still only depends on the two parameters τ and β, which have been calculated from the original constants. Should we know two of the four system constants, we still have no direct use of that in (9.3).

The model (9.3) is thus parameterized by two parameters

$$\theta = \begin{pmatrix} \tau \\ \beta \end{pmatrix}$$

and can also be written

$$\begin{aligned} \frac{d}{dt}x(t) &= A(\theta)x(t) + B(\theta)u(t) \\ y(t) &= C(\theta)x(t) \end{aligned} \tag{9.5}$$

$$A(\theta) = \begin{bmatrix} 0 & 1 \\ 0 & -1/\theta_1 \end{bmatrix}, \quad B(\theta) = \begin{bmatrix} 0 \\ \theta_2/\theta_1 \end{bmatrix}, \quad C(\theta) = [1 \quad 0]$$

The transfer function is

$$G(s,\theta) = C(\theta)[sI - A(\theta)]^{-1}B(\theta) = \frac{\theta_2}{s(1 + s\theta_1)} = \frac{\beta}{s(1 + s\tau)} \tag{9.6}$$

The model is consequently a general first-order model, followed by an integrator. In spite of the rather detailed modeling work, we have still only obtained a structure in which the angular position is the integral

of the angular velocity, which in turn is determined from the applied voltage via a first-order system. This could have been realized directly from physics.

Note also that β is the static gain from voltage to angular velocity. If it is possible to measure the angular velocity, it is thus rather easy to determine the parameter β by a step response experiment and thus reduce the model uncertainty to the parameter $\theta_1 = \tau$. □

We saw in the example that (9.5) will be a special case of (4.1). To stress that the output of the model will depend on the parameter vector θ we write

$$\frac{d}{dt}x(t) = f(x(t), u(t), \theta) \tag{9.7a}$$

$$\hat{y}(t|\theta) = h(x(t), u(t), \theta) \tag{9.7b}$$

$\hat{y}(t|\theta)$ is thus the predicted value of the output at time t, according to the model.

If in our model we assume that measurement noise is affecting the output,

$$y(t) = h(x(t), u(t), \theta) + e(t) \tag{9.8}$$

and that this noise can be described as white noise, the value of $e(t)$ cannot be predicted. In that case it is still $\hat{y}(t|\theta)$ in (9.7b) that constitutes the model's prediction, or guess, of $y(t)$.

When more sophisticated modeling of the disturbances is required, the Kalman filter can be used for computing $\hat{y}(t|\theta)$.

Example 9.2 Model with Kalman Filter

The Kalman filter is a way of estimating the state in a model like (9.5) by using observed values of $u(t)$ and $y(t)$. If the input is known and $y(t)$ is observed only for $t = kT$, $k = 1, 2, \ldots$, the Kalman filter for the time continuous model is given by

$$\dot{\hat{x}}(t) = A(\theta)\hat{x}(t) + B(\theta)u(t) \qquad kT < t < (k+1)T \tag{9.9}$$

$$\hat{x}(kT+) = \hat{x}(kT-) + K(\theta)[y(kT) - C(\theta)\hat{x}(kT-)] \tag{9.10}$$

Here $\hat{x}(kT-)$ is the state estimate just before the observation of $y(kT)$ and $\hat{x}(kT+)$ its value just after this observation. The state estimate $\hat{x}$ will thus change instantaneously at the time instants kT. The Kalman gain $K(\theta)$ is a vector or a matrix that depends in a rather complicated

way on the properties of the disturbance signal. These properties are often unknown, and then $K(\theta)$ can equally well be parameterized directly. The prediction of the output according to the model will now be

$$\hat{y}(t|\theta) = C(\theta)\hat{x}(t) \qquad kT < t < (k+1)T$$
$$\hat{y}(kT|\theta) = C(\theta)\hat{x}(kT-)$$

□

In this section we have used only time continuous models, since this is the most common case in physical modeling. The model (9.7) and other corresponding expressions could of course equally well have been given in discrete time, with obvious modifications.

9.2 Linear, Ready-made Models

Sometimes we are faced with systems or subsystems that cannot be modeled based on physical insights. The reason may be that the function of the system or its construction is unknown or that it would be too complicated to sort out the physical relationships. It is then possible to use standard models, which by experience are known to be able to handle a wide range of different system dynamics. Linear systems constitute the most common class of such standard models. From a modeling point of view these models thus serve as *ready-made models*: tell us the size (model order), and it should be possible to find something that fits (to data).

A Family of Transfer Function Models

Normally, ready-made models are described in discrete time, since data are collected in sampled form. If a time continuous model is required, it is always possible to transform the time discrete model into such a model. A general, linear, time discrete model can be written as

$$y(t) = \eta(t) + w(t) \tag{9.11}$$

Here $w(t)$ is a disturbance term and $\eta(t)$ is the noise-free output, which in turn can be written as

$$\eta(t) = G(q,\theta)u(t) \tag{9.12}$$

If $G(q,\theta)$ is a rational function of the shift operator q,

$$G(q,\theta) = \frac{B(q)}{F(q)} = \frac{b_1 q^{-nk} + b_2 q^{-nk-1} + \cdots + b_{nb} q^{-nk-nb+1}}{1 + f_1 q^{-1} + \cdots + f_{nf} q^{-nf}}, \quad (9.13)$$

then (9.12) is a shorthand notation for the relationship

$$\eta(t) + f_1\eta(t-1) + \cdots + f_{nf}\eta(t-nf)$$

$$b_1 u(t-nk) + \cdots + b_{nb}(t-(nb+nk-1)) \quad (9.14)$$

There is also a time delay of nk samples. In this section we will assume, for simplicity, that the sampling interval T is one time unit.

In the same way the disturbance term can be written

$$w(t) = H(q,\theta)e(t) \quad (9.15)$$

with

$$H(q,\theta) = \frac{C(q)}{D(q)} = \frac{1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}}{1 + d_1 q^{-1} + \cdots + d_{nd} q^{-nd}} \quad (9.16)$$

where $e(t)$ is white noise. [Compare with equation (3.41)!]

The model (9.11) can now be summarized as

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \quad (9.17)$$

The parameter vector θ thus contains the coefficients b_i, c_i, d_i, and f_i of the transfer functions. This ready-made model is thus described by five structural parameters: nb, nc, nd, nf, and nk. When these have been chosen, it remains to adjust the parameters b_i, c_i, d_i, and f_i to data. How this is done will be described in Section 9.3. The ready-made model (9.13)–(9.17) is known as the *Box-Jenkins (BJ) model*, after the statisticians G. E. P. Box and G. M. Jenkins.

An important special case is when the properties of the disturbance signals are not modeled, and the noise model $H(q)$ is chosen to be $H(q) \equiv 1$; that is, $nc = nd = 0$. This special case is known as an *output error (OE) model* since the noise source $e(t) = w(t)$ will then be the difference (error) between the actual output and the noise-free output.

A common variant is to use the same denominator for G and H:

$$F(q) = D(q) = A(q) = 1 + a_1 q^{-1} + \cdots + a_{na} a^{-na} \quad (9.18)$$

Multiplying both sides of (9.17) by $A(q)$ then gives

$$A(q)y(t) = B(q)u(t) + C(q)e(t) \tag{9.19}$$

This ready-made model is known as the *ARMAX model.* The name is derived from the fact that $A(q)y(t)$ represents an AutoRegression and $C(q)e(t)$ a Moving Average of white noise, while $B(q)u(t)$ represents an eXtra input (or with econometric terminology, an eXogenous variable).

Physically, the difference between ARMAX and BJ models is that the noise and input are subjected to the same dynamics (same poles) in the ARMAX case. This is reasonable if the dominating disturbances enter early in the process (together with the input). Consider for example an airplane where the disturbances from wind gusts give rise to the same type of forces on the airplane as the deflections of rudders.

Finally, we have the special case of (9.19) that $C(q) \equiv 1$, that is, $nc = 0$

$$A(q)y(t) = B(q)u(t) + e(t) \tag{9.20}$$

which with the same terminology would be called an *ARX model.*

Figure 9.1 shows the most common model structures.

To use these ready-made models, decide on the orders $na, nb, nc, nd,$ nf, and nk and let the computer pick the best model in the class thus defined. The obtained model is then scrutinized, and it might be found that other order must also be tested.

A relevant question is how to use the freedom that the different model structures give. Each of the BJ, OE, ARMAX, and ARX structures offer their own advantages, and we will discuss them in Section 10.3.

Prediction

Starting with model (9.17), it is possible to predict what the output $y(t)$ will be, based on measurements of $u(s)$, $y(s)$ $s \leq t-1$. The signal $e(t)$ which represents white noise cannot be predicted, since it is independent of everything that has happened before. It is easiest to calculate the prediction for the OE-case, $H(q,\theta) \equiv 1$, when we obtain the model

$$y(t) = G(q,\theta)u(t) + e(t)$$

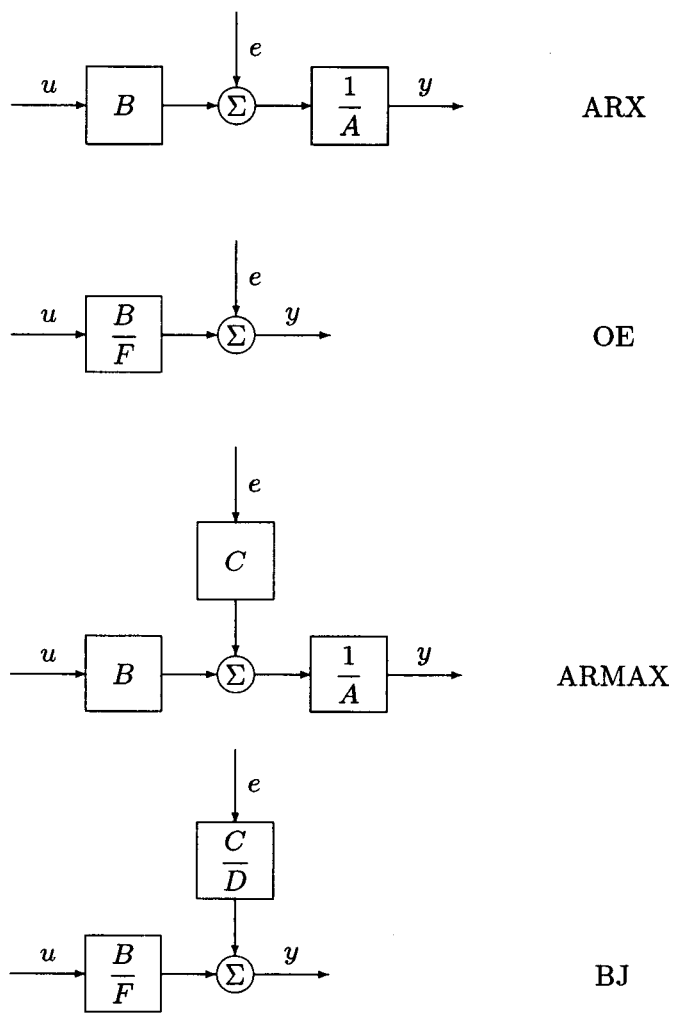


Figure 9.1: Model structures.

with the natural prediction

$$\hat{y}(t|\theta) = G(q,\theta)u(t) \tag{9.21}$$

From the ARX case (9.20) we obtain

$$\begin{aligned} y(t) = &-a_1 y(t-1) - \cdots - a_{na} y(t-na) \\ &+ b_1 u(t-nk) + \cdots + b_{nb} u(t-nk-nb+1) + e(t) \end{aligned} \tag{9.22}$$

and the prediction (delete $e(t)$!)

$$\begin{aligned} \hat{y}(t|\theta) = &-a_1 y(t-1) - \cdots - a_{na} y(t-na) \\ &+ b_1 u(t-nk) + \cdots + b_{nb} u(t-nk-nb+1) \end{aligned} \tag{9.23}$$

Note the difference between (9.21) and (9.23). In the OE model the prediction is based entirely on the input $\{u(t)\}$, whereas the ARX model also uses old values of the output.

In the general case of (9.17) the prediction can be deduced in the following way: Divide (9.17) by $H(q,\theta)$:

$$H^{-1}(q,\theta)y(t) = H^{-1}(q,\theta)G(q,\theta)u(t) + e(t)$$

or

$$y(t) = [1 - H^{-1}(q,\theta)]y(t) + H^{-1}(q,\theta)G(q,\theta)u(t) + e(t) \tag{9.24}$$

We see that (assume $nc \geq nd$)

$$\begin{aligned} 1 - H^{-1}(q,\theta) &= 1 - \frac{D(q)}{C(q)} = \frac{C(q) - D(q)}{C(q)} \\ &= \frac{(c_1 - d_1)q^{-1} + \cdots + (c_{nc} - d_{nc})q^{-nc}}{1 + cq^{-1} + \cdots + c_{nc}q^{-nc}} \end{aligned}$$

The expression $[1 - H^{-1}(q,\theta)]y(t)$ thus only contains old values of $y(s)$, $s \leq t-1$. The right side of (9.24) is thus known at time $t-1$, with the exception of $e(t)$. The prediction of $y(t)$ is simply obtained (9.24) by deleting $e(t)$:

$$\hat{y}(t|\theta) = [1 - H^{-1}(q,\theta)]y(t) + H^{-1}(q,\theta)G(q,\theta)u(t) \tag{9.25}$$

This is a general expression for how ready-made models predict the next value of the output, given old values of y and u. It is also easy to verify that (9.21) and (9.23), respectively, are obtained for the special cases OE and ARX.

Linear Regression

Both tailor-made and ready-made models describe how the predicted value of $y(t)$ depends on old values of y and u and on the parameters θ. We denote this prediction by

$$\hat{y}(t|\theta)$$

See (9.7) and (9.25). In general this can be a rather complicated function of θ. The estimation work is considerably easier if the prediction is a linear function of θ:

$$\hat{y}(t|\theta) = \theta^T \varphi(t) \tag{9.26}$$

Here θ is a column vector that contains the unknown parameters, while $\varphi(t)$ is a column vector formed by old inputs and outputs. Such a model structure is called a *linear regression*: The vector $\varphi(t)$ is the *regression vector* and its components are called *regressors*.

The ARX model (9.20) is the most common example of (9.26) in our context. If we define

$$\theta = [a_1\ a_2 \cdots a_{na} \quad b_1 \cdots b_{nb}]^T \tag{9.27}$$

$$\varphi(t) = [-y(t-1) \cdots - y(t-na) \quad u(t-nk) \cdots u(t-nk-nb+1)]^T$$

we notice that (9.23) corresponds to (9.26).

Linear regression models can also be obtained in several other ways. See Section 10.3.

9.3 Fitting Parameterized Models to Data

The Principle: Minimize the Prediction Errors

For each value of the parameter vector θ, our model provides us with a guess, a prediction of $y(t)$: s value at time $t-1$, that is,

$$\hat{y}(t|\theta) \tag{9.28}$$

regardless if we have tailor-made models as in (9.7) or ready-made models as in (9.25) or linear regression as in (9.26). At time t we can evaluate how good this prediction is by calculating the prediction error

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

When we have collected input and output data over a period $t = 1, \ldots, N$, it is possible to evaluate how well the model with the parameter value θ can describe the performance of the system. We can form the number

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta) \tag{9.29}$$

as a measure of how well the parameter value θ performs. It is natural to choose the value of θ that minimizes (9.29):

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta) \tag{9.30}$$

(arg min denotes the minimizing argument).

Several variants of (9.29) can be considered. If the system has several outputs, a quadratic norm of the vector $\varepsilon(t, \theta)$ can be chosen. In general we may use any arbitrary positive, scalar-valued function $\ell(\varepsilon)$ as a measure and minimize

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \ell(\varepsilon(t, \theta)) \tag{9.31}$$

Here we have motivated the estimation method (9.30) from a pragmatic point of view: Choose the model that best describes (= predicts) observed data. A long list of statistical and information theoretical criteria support the use of the criterion (9.30) as a good choice for parameter estimation. The function (9.29) is, for example, the negative logarithm of the likelihood function for the estimation problem if the noise is supposed to be Gaussian.

In general (9.30)–(9.31) give the *maximum likelihood* (ML) estimate of θ if $\ell(\cdot)$ is chosen as

$$\ell(\varepsilon) = \log f_e(\varepsilon) \tag{9.32}$$

where $f_e(x)$ is the probability density function (pdf) of the noise $e(t)$ in (9.17).

It is also often interesting to estimate the variance of the noise source $e(t)$. With $\hat{\theta}_N$ determined, a natural estimate of λ is obtained as

$$\hat{\lambda}_N = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \hat{\theta}_N) \tag{9.33}$$

Computing the Estimate When the Model Is a Linear Regression

Now consider the special case when the model is a linear regression

$$\hat{y}(t|\theta) = \theta^T \varphi(t)$$

We assume that θ is a d-dimensional column vector. The error will be

$$\varepsilon(t, \theta) = y(t) - \theta^T \varphi(t)$$

and the quadratic criterion (9.29) can be written

$$\begin{aligned} V_N(\theta) &= \frac{1}{N}\sum_{k=1}^{N}(y(t) - \theta^T\varphi(t))^2 = \frac{1}{N}\sum_{t=1}^{N} y^2(t) - \frac{1}{N}\sum_{t=1}^{N} 2\theta^T\varphi(t)y(t) \\ &+ \frac{1}{N}\sum_{t=1}^{N}\theta^T\varphi(t)\varphi^T(t)\theta = \frac{1}{N}\sum_{t=1}^{N} y^2(t) - 2\theta^T f_N + \theta^T R_N \theta \end{aligned}$$

where

$$f_N = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t) \qquad \text{(a } d\text{-dimensional column vector)}$$

$$R_N = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t). \qquad \text{(a } d \times d \text{ matrix)} \tag{9.34}$$

If R_N is invertible, the preceding expression can be written as

$$\begin{aligned} V_N(\theta) = &\frac{1}{N}\sum_{t=1}^{N} y^2(t) - f_N^T R_N^{-1} f_N \\ &+ (\theta - R_N^{-1} f_N)^T R_N (\theta - R_N^{-1} f_N) \end{aligned}$$

The last term is always positive since the matrix R_N is positive semidefinite. The smallest possible value of $V_N(\theta)$ is obtained when this term is zero, that is, when

$$\theta = \hat{\theta}_N = R_N^{-1} f_N \tag{9.35}$$

The least squares estimate $\hat{\theta}_N$ is thus computed by (9.34) and (9.35). In practice, inversion of the matrix R_N is avoided for numerical reasons and $\hat{\theta}_N$ is solved from a linear equation system.

Note that the elements in the matrix R_N and the vector f_N for the ARX model (9.27) are all of the type

$$\frac{1}{N}\sum_{t=1}^{N} y(t-j)u(t-k), \quad \frac{1}{N}\sum_{t=1}^{N} y(t-j)y(t-k)$$

or

$$\frac{1}{N}\sum_{t=1}^{N} u(t-j)u(t-k)$$

The estimate $\hat{\theta}_N$ is thus formed from the estimates of y's and u's covariance functions.

Iterative Search for Minimum

For many model structures the function $V_N(\theta)$ in (9.29) is a rather complicated function of θ, and the minimizing value $\hat{\theta}_N$ must then be computed by numerical search for the minimum. The most common methods for this are based on the Newton–Raphson method. This a method to solve the equation

$$g(x) = 0$$

numerically for x by iteratively selecting

$$x^{(i+1)} = x^{(i)} - [g'(x^{(i)})]^{-1} g(x^{(i)}) \tag{9.36}$$

Here $g'(x)$ is the derivate of $g(x)$ with respect to x.

A step length parameter μ is often used in (9.36) to adjust the update so that $x^{(i+1)}$ is guaranteed to be better than $x^{(i)}$:

$$x^{(i+1)} = x^{(i)} - \mu[g'(x^{(i)}]^{-1} g(x^{(i)}) \tag{9.37}$$

This method can be used to minimize (9.29). We search for a minimum by solving

$$\frac{d}{d\theta} V_N(\theta) = 0 \tag{9.38}$$

Since $V_N(\theta)$ is a real-valued function of d, its gradient is a d- dimensional column vector. Applying (9.36) gives

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - \mu^{(i)}[V_N''(\hat{\theta}^{(i)})]^{-1} V_N'(\hat{\theta}^{(i)}) \tag{9.39}$$

Here $V_N''(\theta)$ is the second derivative (the Hessian) of $V_N(\theta)$ with respect to θ (a $d \times d$ matrix) and $V_N'(\theta)$ is the gradient (a $d \times 1$ vector). The step length $\mu^{(i)}$ is determined so that $V_N(\hat{\theta}^{(i+1)}) < V_N(\hat{\theta}^{(i)})$.

The computation of these expressions depends on the model structure used. More detailed formulas for a general case are shown in the appendix to this chapter. We will conclude this section with two examples of parameter estimation in dynamic systems.

Example 9.3 Estimation of Linear Ready-made Models
Consider the same data that were used in Examples 8.2, 8.3, and 8.5. (See Figure 8.3.) They are fitted to second-order BJ, OE, ARMAX and, ARX models, which give the following models for the dynamics:

$$\mathbf{BJ} : \hat{G}(q) = \frac{0.090q^{-1} + 0.1075q^{-2}}{1 - 1.399q^{-1} + 0.596q^{-2}}$$

$$\mathbf{OE} : \hat{G}(q) = \frac{0.0916q^{-1} + 0.1079q^{-2}}{1 - 1.394q^{-1} + 0.590q^{-2}}$$

$$\mathbf{ARMAX} : \hat{G}(q) = \frac{0.0907q^{-1} + 0.1070q^{-2}}{1 - 1.397q^{-1} + 0.5918q^{-2}}$$

$$\mathbf{ARX} : \hat{G}(q) = \frac{0.1064q^{-1} + 0.2498q^{-2}}{1 - 0.5122q^{-1} - 0.1580q^{-2}}$$

The noise models are

$$\mathbf{BJ} : \hat{H}(q) = \frac{1 + 0.107q^{-1} - 0.886q^{-2}}{1 + 0.166q^{-1} - 0.829q^{-2}}$$

$$\mathbf{OE} : \hat{H}(q) = 1$$

$$\mathbf{ARMAX} : \hat{H}(q) = \frac{1 - 1.438q^{-1} + 0.599q^{-2}}{1 - 1.397q^{-1} + 0.5918q^{-2}}$$

$$\mathbf{ARX} : \hat{H}(q) = \frac{1}{1 - 0.5122q^{-1} - 0.1580q^{-2}}$$

The corresponding values of the loss function (9.33) are

$$\begin{aligned} \mathbf{BJ} &: \hat{\lambda} = 0.1006 \\ \mathbf{OE} &: \hat{\lambda} = 0.1018 \\ \mathbf{ARMAX} &: \hat{\lambda} = 0.1010 \\ \mathbf{ARX} &: \hat{\lambda} = 0.1734 \end{aligned}$$

The frequency functions for the different models are depicted in Figure 9.2. Data had been generated by the time-discrete model

$$y(t) = \frac{0.1044q^{-1} + 0.0883q^{-2}}{1 - 1.4138q^{-1} + 0.6065q^{-2}}u(t) + e(t); \ \lambda = 0.100$$

[This is what we obtain when we sample the system (8.10) with the sampling interval 0.5.] The OE-model is thus the correct structure, but we see that we also obtain a very good estimate of the frequency function in the BJ and ARMAX structures. However, the ARX model gives a bad estimate, since a correct description of both the dynamics and the noise properties is impossible in this structure.

□

Example 9.4 Estimating the Output Coefficient for the Outflow Area in a Paper Machine

In Section 4.2 we described a model for the head box to a paper machine. The model was obtained from equations (4.12)-(4.16). This contains several system parameters, which were listed in that section. Most of them can be determined easily, like the total volume of the head box, its cross section, the outflow area, and so on. But what we called the output coefficient for the outflow area C in equation (4.14) is a parameter not known directly.

In this example we will discuss how C can be estimated from measurements from the system. The difficulties in estimating C depend on which variables can be measured. If the fluid level h in the head box, the pressure p_o, and the outplow q can be measured, then C can easily be determined from the static *relation (4.11). But if the outflow q, the input M (the fluid input), and Q (air input) are the only measured variables the estimation problem is more difficult. We have to use the dynamic model (4.12)-(4.16) in order to estimate C. We use*

$$M(t) = M_0 \qquad \textit{constant}$$

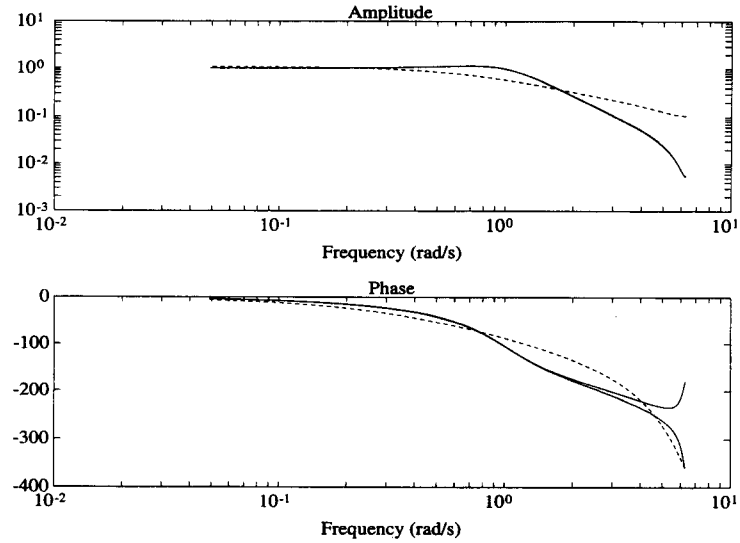


Figure 9.2: Bode plot for estimated models in Example 9.3. Solid line: the true system. Dashed line: ARX model. Other curves: BJ, OE and ARMAX models (these almost coincide with the true system).

$$Q(t) = \begin{cases} Q + \Delta Q \\ \qquad\qquad \textit{random changes between these levels} \\ Q - \Delta Q \end{cases}$$

as input signals during the experiment.

The outflow $q(t)$ is measured at times $t_k = kT$, where $T = 1$ s. Figure 9.3 shows related values of the inputs and outputs.

(In this example these were simulated with noise and the values $V = 10\,m^3$, $A = 10\,m^2$, $R = 267.4\,J/kg\cdot K$, $T = 293\,K$, $a_1 = 10^{-3}\,m^2$, $a_2 = 0.1\,m^2$, $p_0 = 1.013 \cdot 10^5\,N/m^2$, $\rho_2 = 10^3\,kg/m^3$, $g = 9.81\,m/s^2$, $C = 0.8$, $Q = 0.71\,m^3/s$, and $M = 0.34\,kg/s$.)

This gives a mean level of h and h_{eff}, which are $0.5\,m$ and $4\,m$ respectively. If we assume that all system parameters except C are known and given by the preceding values, the model will be (4.12)–(4.13).

$$\frac{d}{dt}x_1(t,C) = Q - 0.1C[1.962x_1 + \frac{78.348x_2}{10 - x_1} - 101.3]^{\frac{1}{2}}$$

$$\frac{d}{dt}x_2(t,C) = M - [(\frac{156.4x_2}{10 - x_1} - 202.6)\frac{x_2}{10 - x_1}]^{\frac{1}{2}} \tag{9.40}$$

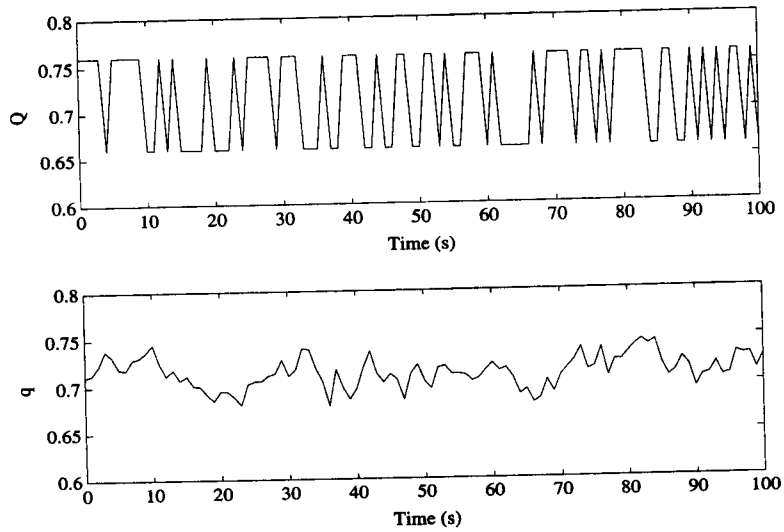


Figure 9.3: Measured input–output data for the head box

$$\hat{q}(t|C) = 0.1C[1.962x_1 + \frac{78.348x_2}{10 - x_1} - 101.3]^{\frac{1}{2}}$$

If we compare the output $\hat{q}(t|C)$ *for this model with the measured value* $q(kT), k = 1, ..., 100$, *we can form the following criterion:*

$$V_{100}(C) = \frac{1}{100}\sum_{k=1}^{100}(\hat{q}(kT|C) - q(kT))^2 \tag{9.41}$$

This function is shown in Figure 9.4.

We see that the criterion is minimized for $C \approx 0.802$. *This is in good agreement with the true value* $C = 0.8$. □

9.4 Model Properties

The method to minimize the prediction error (9.30) is a general method to estimate parameters. Which properties will these estimates have? How good are the resulting models? These questions will be treated in this section.

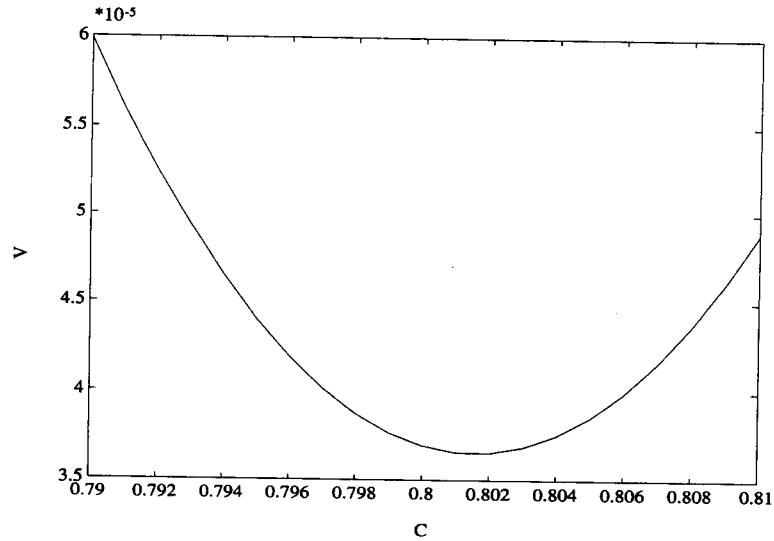


Figure 9.4: The criterion as a function of C.

Model Quality

What do we mean by a good model? This should mean that it is close to the true description. In practice, no "true" descriptions are available, and the model quality must thus be judged on different grounds. We note a number of basic facts:

1. *Model quality is related to model use.* A model can for example be excellent for control design but inadequate for simulation.

2. *Model quality is typically related to the ability of the model to reproduce the behavior of the system.* This usually means that the model's simulated or predicted output is in good agreement with the outputs produced by the system.

3. *Model quality is also related to the model's stability,* that is, how well the model can be reproduced from different measured data sets. Regardless of the statistical formalism that can be developed around these facts, it is obvious that it is necessary to question the resulting model if it varies much with the segment of data it was computed from. Conversely, confidence is gained in a model obtained with small variations from different measurement data under varied conditions and maybe with different identification methods.

We will return to these general aspects of model quality in Section

10.4. Here we will first develop some formal results in connection with the third aspect.

Bias and Variance

Model inadequacy can take two, principally different shapes. One is the model error that arises because of noise influence on the measurements and system. If an experiment is repeated with exactly the same input, the output will not be exactly the same because the noise is not reproduced. Because of this the resulting model will be different. Such model variations are called *variance errors*. Variance errors can typically be reduced by using longer measurement sequences.

The other model inadequacies originate from deficiencies in the model structure. The model is simply not capable of describing the system, even if it should be fitted to noise-free data. Such errors will be called systematic errors, or *bias error*. Bias errors are noticed as variations in the model when it is fitted to data that have been collected during different conditions (even if the measurement intervals are long enough to make the variance error insignificant). The reason is that different conditions (operating point, input characteristics and so on) bring out different aspects of the system's properties and the model is fitted to the dominating system properties.

A good model according to statement 3 above is thus one that has both small variance and bias error.

Convergence of the Estimate – Bias Error

Consider the parameter estimate $\hat{\theta}_N$ from (9.30). If we want to concentrate on the bias error, the obvious question is what will happen when N, the number of measurement data, tends to infinity. The variance error should then be negligible.

The answer is the following. If the noise that influences the system can be described as a stationary stochastic process, then the prediction error $\varepsilon(t, \theta)$ for each value of θ is a stationary process. Let the variance of ε be denoted by

$$E\varepsilon^2(t, \theta) = \bar{V}(\theta) \tag{9.42}$$

If $\varepsilon(t, \theta)$ is a sequence of independent stochastic variables (white noise),

then the law of large numbers in its simplest form would imply that

$$\frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\theta) \to E\varepsilon^2(t,\theta) = \bar{V}(\theta) \quad \text{as} \quad N \to \infty \tag{9.43}$$

This convergence happens with probability one (w.p.1), that is, the probability that the event occurs (9.43) is 1. Now, $\varepsilon(t,\theta)$ are not independent, but under very general conditions the dependence decreases so fast that (9.43) still holds. The convergence is also uniform in the parameter θ. This implies that

$$\boxed{\hat{\theta}_N \to \theta^* = \arg\min_{\theta} \bar{V}(\theta) \text{ as } N \to \infty \qquad (9.44)}$$

This result is completely general and contains all information of the bias error. The estimate $\hat{\theta}_N$ converges to the value that minimizes $E\varepsilon^2(t,\theta)$. If we cannot get an exact model (which gives white prediction errors), we will at least obtain the best approximation that is available within the parametrized model, the one that minimizes the prediction error variance. This is an important robustness property of the estimate.

For linear models the result is best interpreted in the frequency domain. Assume that we have a linear model

$$y(t) = G(q,\theta)u(t) + H_*(q)e(t) \tag{9.45}$$

where $H_*(q)$ is a fixed (θ-independent) model for the noise term. Assume that the true system is given by

$$y(t) = G_0(q)u(t) + w(t) \tag{9.46}$$

Using expression (9.25), we can compute the prediction for (9.45). The difference $\varepsilon(t,\theta) = y(t) - \hat{y}(t|\theta)$ gives

$$\begin{aligned}\varepsilon(t,\theta) &= H_*^{-1}(q)[y(t) - G(q,\theta)u(t)] \\ &= H_*^{-1}(q)[G_0(q) - G(q,\theta)]u(t) + H_*^{-1}(q)w(t)\end{aligned}$$

The spectrum for $\varepsilon(t,\theta)$ will then be, according to (3.63) (if u and w are independent),

$$\Phi_\varepsilon(\omega,\theta) = |G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \frac{\Phi_u(\omega)}{|H_*(e^{i\omega})|^2} + \frac{\Phi_w(\omega)}{|H_*(e^{i\omega})|^2} \tag{9.47}$$

Parseval's formula (C.21) gives

$$E\varepsilon^2(t,\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_\varepsilon(\omega,\theta)d\omega$$

and thus yields

$$\boxed{\theta^* = \lim_{N\to\infty} \hat{\theta}_N = \arg\min_\theta \int_{-\pi}^{\pi} |G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \frac{\Phi_u(\omega)}{|H_*(e^{i\omega})|^2} d\omega \qquad (9.48)}$$

since the last term in (9.47) is independent of θ.

The estimate thus converges to the value θ^*, which makes the frequency function of the model $G(e^{i\omega},\theta^*)$ as close as possible to the true $G_0(e^{i\omega})$ measured in a quadratic frequency norm with the weighting function

$$\frac{\Phi_u(\omega)}{|H_*(e^{i\omega})|^2} \qquad (9.49)$$

Note especially that by choice of Φ_u and H_* we can control the frequency bands where the fit is best. This will give a good illustration of how the bias error depends on the experimental conditions — in this case the input spectrum.

If there is a value θ_0 such that

$$G(e^{i\omega},\theta_0) \equiv G_0(e^{i\omega})$$

we see from (9.48) that $\theta^* = \theta_0$ independently of $\Phi_u(\omega)$ and $H_*(e^{i\omega})$ as long as $\Phi_u(\omega)$ is different from zero for sufficiently many ω.

Another general result also follows directly from (9.44). Assume, in the general case, that there is a value θ_0 such that

$$y(t) - \hat{y}(t|\theta_0) = \varepsilon(t,\theta_0) = e(t) = \text{ white noise, with variance } \lambda \quad (9.50)$$

Then we get from (9.44)

$$\begin{aligned}\bar{V}(\theta) &= E\varepsilon^2(t,\theta) = E(y(t) - \hat{y}(t|\theta))^2 \\ &= E(\hat{y}(t|\theta_0) - \hat{y}(t|\theta) + e(t))^2 \\ &= E(\hat{y}(t|\theta_0) - \hat{y}(t|\theta))^2 + \lambda \qquad (9.51)\end{aligned}$$

since $e(t)$ is independent of all old data. We see that $\theta = \theta_0$ minimizes $\bar{V}(\theta)$. From this follows

$$\boxed{\hat{\theta}_N \to \theta_0 \quad \text{when} \quad N \to \infty \qquad (9.52)}$$

under the condition

$$\hat{y}(t|\theta_0) \equiv \hat{y}(t|\theta) \Rightarrow \ \theta = \theta_0 \tag{9.53}$$

We will discuss the condition (9.53) later under "Identifiability."

Variance Error

We will discuss the variance error of the estimate in the case that the bias error is zero, that is, when the assumption (9.50) is valid. In the appendix to this chapter we will show that

$$P_N = E(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T \approx \frac{1}{N}\lambda \cdot \bar{R}^{-1} \tag{9.54}$$

where

$$\bar{R} = E\psi(t, \theta_0)\psi^T(t, \theta_0) \tag{9.55}$$

$$\psi(t, \theta) = \frac{d}{d\theta}\hat{y}(t|\theta) \quad (\text{a}\, d \times 1 \text{ vector}) \tag{9.56}$$

The covariance matrix for $\hat{\theta}_N$ is thus proportional to the noise intensity λ and inversely proportional to the number of measurement points. Less noise and more data mean a better estimate, which is completely natural. The covariance matrix of $\hat{\theta}_N$ is also proportional to the inverse of the covariance matrix for $\psi(t, \theta)$, the gradient of the prediction with respect to θ. This is also natural. The quality of $\hat{\theta}_N$ depends on how sensitive the prediction is regarding θ. If the prediction $\hat{y}(t|\theta)$ changes only a little when a certain component of θ is changed, the corresponding component in $\psi(t, \theta_0)$ will be small $\bar{R}$ will be small, and consequently the uncertainty in the θ component will be large.

It is true that the matrix $\bar{R}$ is unknown to us, but it can easily be estimated. The gradient $\psi(t, \theta)$ can be computed for each given value of θ — this is also normally done in the minimizing algorithm when

$\hat{\theta}_N$ is computed. We use the estimates

$$\hat{R}_N = \frac{1}{N}\sum_{t=1}^{N}\psi(t,\hat{\theta}_N)\psi^T(t,\hat{\theta}_N) \tag{9.57}$$

$$\hat{\lambda}_N = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\hat{\theta}_N)$$

and can then estimate $\hat{\theta}_N$'s covariance matrix as

$$\hat{P}_N = \frac{1}{N}\hat{\lambda}_N\hat{R}_N^{-1} \tag{9.58}$$

It can also be shown that the distribution for the stochastic variable $\hat{\theta}_N$ converges to a normal distribution (with mean θ_0 and covariance matrix P_N). This is written

$$\boxed{\sqrt{N}(\hat{\theta}_N - \theta_0) \in AsN(0, \lambda\bar{R}^{-1})} \tag{9.59}$$

("belongs to the asymptotic normal distribution with mean 0 and covariance matrix $\lambda\bar{R}^{-1}$").

With the help of (9.59) we answer questions of the type "How large is the probability that $\hat{\theta}_N$ differs from θ_0 by more than 10%?" For this we use standard tables for the normal distribution.

The result (9.54) is mainly used to judge the variance error in a computed estimate $\hat{\theta}_N$. We then use (9.57)–(9.58). The result can of course also be used to compute the effect on an estimate of different design variables.

Example 9.5 Variance of ARX Parameters

Consider a system described by

$$y(t) - 0.9y(t-1) = u(t-1) + e(t) \tag{9.60}$$

The input u is white noise with variance μ, and the noise $\{e(t)\}$ is white with variance λ. By squaring (9.60)and taking the expected value, we get

$$(1 + 0.81)R_y(0) - 1.8R_y(1) = \mu + \lambda$$

By multiplying (9.60) by $u(t)$ and taking the expected value, we obtain

$$R_{yu}(0) = Ey(t)u(t) = 0$$

[$u(t)$ is independent of $y(t-1), u(t-1)$, and $e(t)$]. By multiplying (9.60) by $y(t-1)$ and taking the expected value, we get

$$R_y(1) - 0.9R_y(0) = 0$$

This gives

$$R_y(0) = \frac{\mu + \lambda}{0.19}$$

We use the ARX model

$$y(t) + ay(t-1) = bu(t-1) + e(t)$$

to identify (9.60). The predictor is

$$\hat{y}(t|\theta) = -ay(t-1) + bu(t-1)$$

which gives

$$\psi(t,\theta) = \begin{pmatrix} -y(t-1) \\ u(t-1) \end{pmatrix}$$

and

$$\bar{R} = E\psi(t,\theta)\psi^T(t,\theta) = \begin{pmatrix} R_y(0) & R_{yu}(0) \\ R_{yu}(0) & R_u(0) \end{pmatrix} = \begin{pmatrix} \frac{\mu+\lambda}{0.19} & 0 \\ 0 & \mu \end{pmatrix}$$

The variance in the estimate $\hat{a}_N, \hat{b}_N$ is then, according to (9.54),

$$\text{Var}\ (\hat{a}_N) \approx \frac{1}{N}\frac{0.19\lambda}{\mu+\lambda}, \quad \text{Var}\ (\hat{b}_N) \approx \frac{1}{N}\frac{\lambda}{\mu}$$

Here we see directly how the input variance μ influences the accuracy of the estimate.

□

In a ready-made model of the type (9.17) the primary interest may not be in the variance of the parameter, but in the variance of the frequency function. Starting with (9.54), we can estimate the variance for $G(e^{i\omega}, \hat{\theta}_N)$. This gives a rather complicated expression. However, the simple relationship

$$\boxed{\text{Var}(G(e^{i\omega}, \hat{\theta}_N)) \approx \frac{n}{N} \cdot \frac{\Phi_w(\omega)}{\Phi_u(\omega)} \qquad (9.61)}$$

holds approximately, where n is the model order, N the number of data, and Φ_w and Φ_u spectra for w and u in (9.15) and (9.17). Notice the similarity to (8.51).

Identifiability

If we use a tailor-made parameterization it is interesting to know if the chosen parameters can be determined from these data at all.

Consider for example the dc model in Example 9.1. To begin with, we had four system constants in the model building. But they entered the model only as the two combinations τ and β. It is thus obvious that if we had a θ that consisted of these four parameters we would not be able to decide their individual values from only measurements of the system's inputs and outputs.

In other cases it can be considerably more difficult to decide such questions. We use the term *identifiability* to describe that a certain parameter can be uniquely identified from the input–output signal data. The key relationship is (9.53).s If two different parameter vectors θ_1 and θ_2 give rise to identical predictions they cannot be separated by identification methods. We shall say that a certain parameterization is *identifiable* at θ_*, if

$$\hat{y}(t|\theta_*) \equiv \hat{y}(t|\theta) \quad \text{implies} \quad \theta = \theta_* \tag{9.62}$$

There are two different reasons that (9.62) may not hold. One is that two different values of θ simply give identical input-output properties in the model. The dc motor was an example of this. The other reason is that we do get different models with different values of θ, but because of deficiencies in the input the predictions are still the same.

Example 9.6 Nonidentifiability
Consider the model

$$\hat{y}(t|\theta) = b_1 u(t-1) + b_2 u(t-2), \qquad \theta = (b_1\ b_2)^T$$

The input is chosen as a constant $u(t) \equiv u_0$. The actual prediction is then

$$\hat{y}(t|\theta) = (b_1 + b_2)u_0$$

and we see that all values of b_1 and b_2 whose sum is a given value give identical predictions. Equation (9.62) does not hold, and the parameters b_1 and b_2 are thus nonidentifiable. □

In Section 10.2 we will discuss choices of input that assure identifiability whenever possible.

9.5 Summary

To estimate parameters in models of dynamic systems gives a very powerful and wide range of possibilities for model construction. Here are the fundamental features in the method:

- The basic principle is to fit the parameters so that the model will predict the measured output as well as possible:

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta)$$

- The minimization work may demand heavy computations.
- Under the assumption that noise and disturbances that affect the process can be described as stochastic processes, the general expressions for the model properties and quality are as follows:

$$\hat{\theta}_N \to \theta^* = \arg\min E\varepsilon^2(t, \theta)$$

$$\text{Cov}\, \hat{\theta}_N \sim \frac{1}{N} \lambda [E\psi(t, \theta^*)\psi^T(t, \theta^*)]^{-1}$$

- To be able to estimate the magnitude of the variance error is very valuable: Each model has to be delivered to the user together with a quality declaration!

The parameter estimation has the following advantages:

- Generally applicable. Can be applied to detailed tailor-made models as well as to crude ready-made models.
- Only what is unknown needs to be estimated.

The disadvantages are the following:

- Certain insights in to the system's properties are needed to be able to test reasonable model structures.
- A comprehensive computer and program support is needed to compute and evaluate different kinds of models with reasonable effort.

9.6 Appendix

Computing Derivatives for Numerical Minimization

We gave equation (9.39) as the basic algorithm for numerical minimization. In order to use it, derivatives and second derivatives of the criterion $V_N(\theta)$ are needed.

The computation of these expressions depends on the model structure used. In this appendix we will first study a general case with a quadratic criterion function:

$$V_N(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}(y(t_i) - \hat{y}(t_i|\theta))^2 \tag{9.63}$$

Direct, term-wise differentiation of (9.63) gives

$$V_N'(\theta) = \frac{1}{N}\sum_{i=1}^{N} -\frac{d}{d\theta}\hat{y}(t_i|\theta)(y(t_i) - \hat{y}(t_i|\theta)) \tag{9.64}$$

and

$$\begin{aligned} V_N''(\theta) = &\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d}{d\theta}\hat{y}(t_i|\theta)\right)\left(\frac{d}{d\theta}\hat{y}(t_i|\theta)\right)^T \\ &+\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d^2}{d\theta^2}\hat{y}(t_i|\theta)\right)(y(t_i) - \hat{y}(t_i|\theta)) \end{aligned} \tag{9.65}$$

In general we approximate the second derivative $V_N''(\theta)$ by neglecting the second sum in (9.65). It is small anyway, close to the minimum point, and it is then unnecessary to compute the second derivative of $\hat{y}(t_i|\theta)$. With this term canceled in the expression for $V_N''(\theta)$, (9.39)

is usually called The *Gauss-Newton method.* Now it only remains to describe how the derivatives

$$\psi(t,\theta) = \frac{d}{d\theta}\hat{y}(t|\theta) \tag{9.66}$$

are computed. The computation of these depends on the model structure used. We will give these estimations for both the ARMAX model (9.19) and for the general structure (9.7).

Derivatives for the ARMAX Model

The ARMAX model is given by (9.19):

$$A(q)y(t) = B(q)u(t) + C(q)e(t)$$

The general prediction formula (9.25) gives in this case

$$\hat{y}(t|\theta) = [1 - \frac{A(q)}{C(q)}]y(t) + \frac{A(q)}{C(q)} \cdot \frac{B(q)}{A(q)}u(t)$$

or

$$C(q)\hat{y}(t|\theta) = [C(q) - A(q)]y(t) + B(q)u(t) \tag{9.67}$$

Now differentiate with respect to the coefficient a_k in the $A(q)$ polynomial:

$$C(q)\frac{d}{da_k}\hat{y}(t|\theta) = -y(t-k)$$

In the same way we have

$$C(q)\frac{d}{db_k}\hat{y}(t|\theta) = u(t-k)$$

and by differentiating $C(q)\hat{y}(t|\theta)$ as a product

$$\hat{y}(t-k|\theta) + C(q)\frac{d}{dc_k}\hat{y}(t|\theta) = y(t-k)$$

or

$$C(q)\frac{d}{dc_k}\hat{y}(t|\theta) = y(t-k) - \hat{y}(t-k|\theta) = \varepsilon(t-k,\theta)$$

These expressions can be summarized as

$$C(q)\psi(t,\theta) = \begin{bmatrix} -y(t-1) \\ \vdots \\ -y(t-na) \\ u(t-1) \\ \vdots \\ u(t-nb) \\ \varepsilon(t-1,\theta) \\ \vdots \\ \varepsilon(t-nc,\theta) \end{bmatrix} \tag{9.68}$$

The General Case (9.7)

Consider the general state-space form (9.7). By differentiating (9.7b) with respect to θ, we get by the chain rule

$$\begin{aligned} \psi^T(t,\theta) &= \frac{d}{d\theta}\hat{y}(t|\theta) = \frac{d}{d\theta}[h(x(t,\theta),u(t),\theta)] \\ &= H_1(x(t,\theta),u(t),\theta)z(t,\theta) + H_2(x(t,\theta),u(t),\theta) \end{aligned} \tag{9.69}$$

where

$$\begin{aligned} H_1(x,u,\theta) &= \frac{d}{dx}h(x,u,\theta) \qquad (\text{a } 1 \times n \text{ vector}) \\ z(t,\theta) &= \frac{d}{d\theta}x(t,\theta) \qquad (\text{an } n \times d \text{ matrix}) \\ H_2(x,u,\theta) &= \frac{d}{d\theta}h(x,u,\theta) \qquad (\text{a } 1 \times d \text{ vector}) \end{aligned}$$

In the same way we get from (9.7a)

$$\begin{aligned} \frac{d}{dt}z(t,\theta) &= \frac{d}{dt}\frac{d}{d\theta}x(t,\theta) = \frac{d}{d\theta}\frac{d}{dt}x(t,\theta) \\ &= \frac{d}{d\theta}[f(x(t,\theta),u(t),\theta)] = F_1(x(t,\theta),u(t),\theta)z(t,\theta) \\ &\quad + F_2(x(t,\theta),u(t),\theta) \end{aligned} \tag{9.70}$$

where

$$F_1(x,u,\theta) = \frac{d}{dx}f(x,u,\theta) \qquad (\text{an } n \times n \text{ matrix})$$

$$F_2(x,u,\theta) = \frac{d}{d\theta} f(x,u,\theta) \qquad (\text{an } n \times d \text{ matrix})$$

In the second expression we changed the order of differentiation. This is allowed if $x(t,\theta)$ is twice continuously differentiable.

Equations (9.7), (9.69), and (9.70) now make a system of $n + n \cdot r$ coupled differential equations, from which we can determine the derivative $\frac{d}{d\theta}\hat{y}(t_i|\theta)$.

The preceding expressions look complicated. But if a program package for solving differential equations numerically is available (see Chapter 11), the necessary programming work is still moderate.

It is evident from the expressions that the computational work to minimize (9.63) increases rapidly with n and d. (The necessary number of iterations typically grows faster than linearly with d.) There is therefore reason to try to estimate unknown system parameters using as small partial systems as possible. It is thus not necessary to let (9.1) be the *whole* system model if there is a smaller subsystem that also gives information of θ or some of θ's components.

Proof of (9.54) and (9.59)

According to (9.30), let

$$\hat{\theta}_N = \arg\min V_N(\theta)$$

We then have that the derivative of the criterion is zero

$$V_N'(\hat{\theta}_N) = 0$$

Assume that N is sufficiently large and that $\hat{\theta}_N$ lies close to θ_0. Then Taylor expansion gives

$$0 = V_N'(\hat{\theta}_N) \approx V_N'(\theta_0) + V_N''(\theta_0)(\hat{\theta}_N - \theta_0) \tag{9.71}$$

We thus have for large N

$$\hat{\theta}_N - \theta_0 \approx -[V_N''(\theta_0)]^{-1} V_N'(\theta_0) \tag{9.72}$$

For $V_N''(\theta_0)$ we have from (9.65)

$$V_N''(\theta_0) = \frac{1}{N}\sum_{t=1}^{N} \psi(t,\theta_0)\psi'(t,\theta_0) + \frac{1}{N}\sum_{t=1}^{N} \frac{d^2}{d\theta^2}\hat{y}(t|\theta_0)e(t)$$

For large N this expression will be close to its expected value:

$$V_N''(\theta_0) \approx E\psi(t,\theta_0)\psi^T(t,\theta_0) \triangleq \bar{R} \tag{9.73}$$

Here we make use of the fact that $e(t) = y(t) - \hat{y}(t|\theta_0)$ is white noise and thereby independent of $\hat{y}(t|\theta)$. For $-V_N'(\theta_0)$ we have from (9.64)

$$-V_N'(\theta_0) = \frac{1}{N}\sum_{t=1}^{N}\psi(t,\theta_0)e(t) \tag{9.74}$$

The expected value of this expression is zero. [$e(t)$ and $\psi(t,\theta_0)$ are independent and $e(t)$ has the expected value of zero.] If the terms were independent, the central limit theorem would also give that $-\sqrt{N}V_N'(\theta_0)$ converges to a normally distributed variable with zero mean and variance:

$$ENV_N'(\theta_0)[V_N'(\theta_0)]^T = \frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}E\psi(t,\theta_0)\psi(s,\theta_0)\cdot Ee(s)e(t) = \lambda\bar{R} \tag{9.75}$$

where we made use of the fact that $Ee(s)e(t) = \delta_{t-s}\lambda$.

The terms in (9.74) are indeed not independent, but the central limit theorem is still valid under general conditions. If we combine this with (9.72), (9.73), and (9.75), we have that $\sqrt{N}(\hat{\theta}_N - \theta_0)$ converges to a normal distribution with zero mean and covariance matrix $\bar{R}^{-1}[\lambda\bar{R}]\bar{R}^{-1} = \lambda\bar{R}^{-1}$, which gives (9.59) and (9.54).

Chapter 10

System Identification as a Tool for Model Building

The previous two chapters have shown the techniques and methods available for model building based on measurements. In this chapter we will discuss what the identification tool looks like in the hands of the user and how it is used for model building.

The most important aspect is that there are now many interactive computer program packages for identification in which the methods and theories are packaged in a user-friendly way. The focus has thereby moved from algorithms for identification to understanding of the possibilities and limitations of identification.

With given data the user's main task is to decide on a suitable model structure – a suitable parametrization of the model – and to evaluate the calculated models.

Another main task is to construct and carry out experiments that give data with good information contents for the subsequent identification step.

In this chapter we will first describe typical computer packages. After that, in Section 10.2, we will deal with construction of identification experiments and in Section 10.3 with posttreatment of data. The choice of model structure is discussed in Sections 10.4 and 10.5. An example and a discussion of the possibilities and limitations of system identification conclude the chapter.

10.1 Program Packages for Identification

The work to produce a model by identification is characterized by the following sequence:

1. Specify a model structure.
2. The computer delivers the best model in this structure.
3. Evaluate the properties of this model.
4. Test a new structure, go to step 1.

See Figure 10.1. The first thing that requires help is to compute the model and to evaluate its properties. There are now many commercially available program packages for identification that supply such help. They typically contain the following routines:

A *Handling of data, plotting, and the like*
Filtering of data, removal of drift, choice of data segments, and so on

B *Nonparametric identification methods*
Estimation of covariances, Fourier transforms, correlation and spectral analysis, and so on.

C *Parametric estimation methods*
Calculation of parametric estimates in different model structures

D *Presentation of models*
Simulation of models, estimation and plotting of poles and zeros, computation of frequency functions and plotting in Bode diagrams, and so on

E *Model validation*
Computation and analysis of residuals ($\varepsilon(t, \hat{\theta}_N)$); comparison between different models' properties, and the like

The existing program packages differ mainly by various user interfaces and by different options regarding the choice of model structure according to item C.

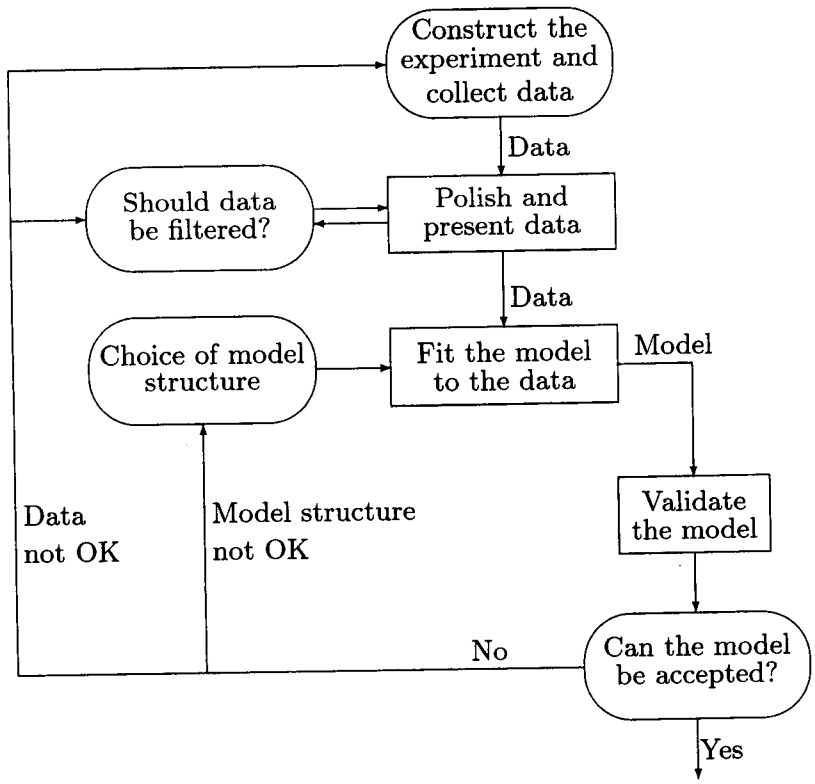


Figure 10.1: Identification cycle. Rectangles: the computer's main responsibility. Ovals: the user's main responsibility.

One of the most used packages is MathWorks SYSTEM IDENTIFICATION TOOLBOX (SITB), which is used together with MATLAB. The command structure is given by MATLAB's programming environment with the work-space concept and MACRO possibilities in the form of m-files. SITB gives the possibility to use all model structures of the type (9.17)–(9.20) with an arbitrary number of inputs. The user can also define arbitrary tailor-made linear state-space models in discrete and continuous time.

10.2 Design of Identification Experiments

A successful identification application demands that the collected measurement data contain significant information about the system. This in turn requires that the data acquisition be well planned. Several decisions regarding this have to be taken:

- Which signals in the process should be measured?
- How should the inputs be chosen?
- Which sampling interval should be used?
- How much data need be collected?

The following questions concerning posttreatment of data are also related to data quality:

- Do slower variations, drift, and the like, have to be removed?
- Do the data need filtering to remove disturbing frequencies?
- Are all data points reliable?
- Should the data sequence be decimated (that is, be resampled at a lower frequency)?

To deal with all these questions in depth demands a long discussion. In this section we will only indicate the most important aspects.

Guiding Principles

The basic result of the properties of the estimates, as described in Section 9.4, can be summarized in three points:

1. The estimate $\hat{\theta}_N$ converges to a value that gives the best approximation θ^* of the system's properties *under the conditions valid during the data acquisition.*
2. The limiting model θ^* gives a true description of the system if the model parameterization allows this, and if the experiment conditions are such that no two different models have identical behaviors under these conditions.

3. The covariance matrix for the estimation error in the parameters is given approximately by

$$\frac{\lambda}{N}[E\psi(t,\theta_0)\psi^T(t,\theta_0)]^{-1}$$

where $\psi(t,\theta_0)$ is the gradient of the prediction with respect to the parameters. N is the number of collected data points and λ is the noise variance.

These three points give guidance for the construction of identification experiments in the following way:

- Let the experiment be carried out under conditions that are as similar as possible to those under which the model is going to be used
- See to it that the inputs excite all interesting aspects of the system
- Choose the measurement $y(t)$ and the input $u(t)$ so that the prediction $\hat{y}(t|\theta)$ will be as sensitive as possible with respect to θ

Choice of Input

The input $u(t)$ should excite the system. One single pure sinusoid with the frequency ω, for example, only gives information of the value of the frequency function at ω. There are infinitely many systems that give the same output with this input. It is thus important that $u(t)$ contain enough frequencies. If the system allows the user to choose the input, a good choice is to let it shift randomly between two levels (a "telegraph" signal). Such signals contain all frequencies. The levels should be chosen so that they correspond to maximally allowed variations. If the system is nonlinear, an interval for the input that corresponds to a desired operation point should be chosen. When constructing a nonlinear model, it is usually also necessary to work with more than two input levels.

A simple way to generate binary random signals is to filter a white noise signal (from a normal random generator) through a suitably

chosen filter, take the sign of this filtered output and then adjust the level and interval to what is required by the application. The filter will determine the spectrum of the input. The expressions (9.48) and (9.61) can be used as a guide for the choice of spectrum. The input should have the major part of its energy in the frequencies that are important for the model fit.

In the time domain we could think as follows: First, make a step response to get a general feel for the time constants of the system. If we seek an input as a pulse train, consisting of pulses of different durations, it is of course not much use to have pulses so short that the response is hardly visible, that is, just covering a negligible part of the rise time of the step response. Moreover, it should be useful to have occasional pulses that are constant over such long periods that the step response more or less settles. There is, however, no need to have longer pulses than these. This gives some practical guidelines for the choice of pulse lengths.

An often used signal for identification is the PRBS (pseudo random binary signal). It has approximately the same properties as the telegraph signal, but is deterministic in nature (periodic with a rather long period) and can be realized by a shift register.

The major aspects on the choice of input can be summarized as follows:

- Binary signals are often suitable to identify linear systems.
- Choose the frequency range of the input so that it has most of its energy in the frequency bands that are important for the system, that is, where the Bode diagram's breaking points are.
- Alternatively expressed: Let the input contain pulses that occasionally allow the step response to more or less settle, and also clearly excite interesting fast modes in the system.
- It is often a good idea to first generate the input sequence in an off-line fashion and examine its properties before applying it to the system.

Feedback Experiments

It is often expensive or dangerous to experiment with the system, and then data have to be collected during normal operating conditions. Normally, this means that the process is controlled so that the input is partially determined by feedback from the output. That this can cause some problems is evident from the following example.

Example 10.1 Identification under Feedback
Consider the simple ARX model

$$y(t) + ay(t-1) = bu(t-1) + e(t) \tag{10.1}$$

Assume that the system is controlled by a proportional regulator during the data acquisition:

$$u(t) = -fy(t) \tag{10.2}$$

The predictor is thus

$$\hat{y}(t|\theta) = -ay(t-1) + bu(t-1) = (-bf - a)y(t-1)$$

All values $(\hat{a}, \hat{b})$ *of the model parameters such that* $\hat{b}f + \hat{a}$ *is a certain given number will therefore produce identical predictions under the feedback (10.2). This is a situation similar to Example 9.6. Note especially that it does not help to know the regulator constant* f. *There are thus no possibilities to determine* a *and* b *uniquely in (10.1) when (10.2) holds, regardless of what the true system is.*

If we change (10.2) so that a set point for the output $r(t)$ *is included, we have*

$$u(t) = f(r(t) - y(t)) \tag{10.3}$$

The predictor will then be

$$\hat{y}(t|\theta) = (-bf - a)y(t-1) + bfr(t-1)$$

If r *is not identically zero, the predictor will now distinguish between different values of* a *and* b.

The system (10.1), (10.3) has been simulated for the values $a = -0.9$, $b = 0.5$, *and* $f = 1$ *in Figures 10.2 and 10.3.* $\{e(t)\}$ *was simulated as normally distributed white noise with variance 0.1 and* $r(t)$ *alternating between 0 and 1 according to the figure. An estimate*

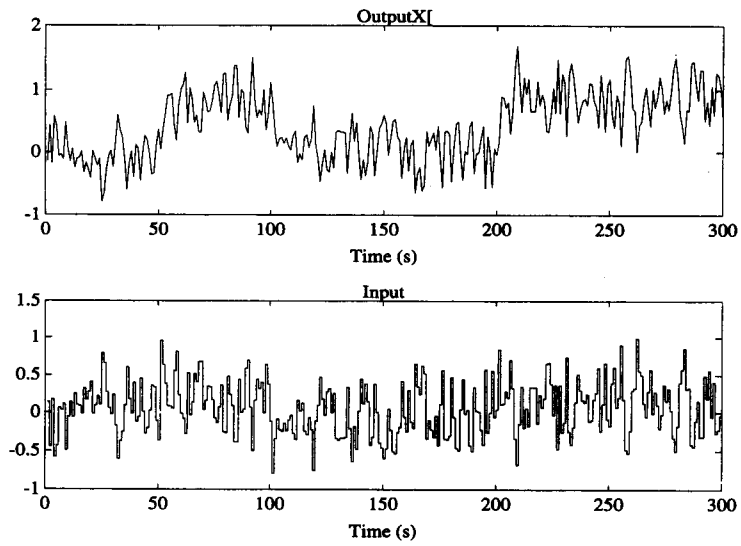


Figure 10.2: Input and output from the feedback system.

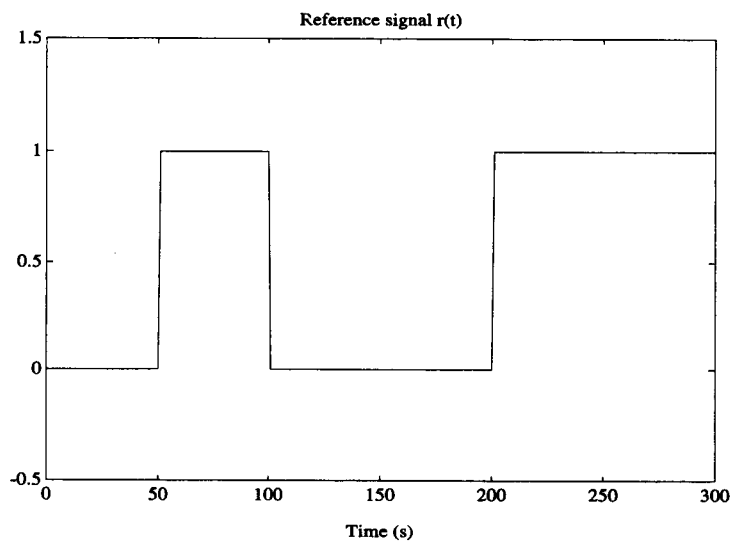


Figure 10.3: The reference signal $r(t)$ for the feedback system.

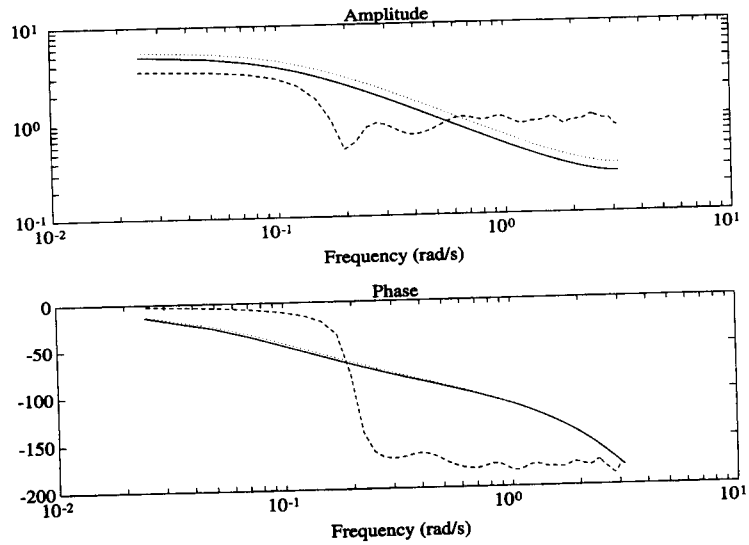


Figure 10.4: Spectral analysis estimation of the transfer function. Solid line: true system. Dashed line: spectral analysis of the transfer function. Dotted line: ARX model's transfer function.

of the parameters a and b in the ARX structure (10.1) based on these data gave the values

$$\hat{a}_{300} = -0.8902 \pm 0.0521, \qquad \hat{b}_{300} = 0.6100 \pm 0.0521$$

Spectral analysis according to the algorithm SPA (8.49) was also carried out for the data given in Figure 10.4. We see that the spectral analysis estimate is useless. This is because of the two relationships that exist between the signals u and y: partly due to the system dependence (10.1) and partly due to the regulator dependence (10.3). In spectral analysis these dependences are mixed. When we use the ARX model (10.1) we look explicitly for how $y(t-1)$ *and* $u(t-1)$ *affect* $y(t)$. *The regulator (10.3) does not influence this relation.* □

The example shows the importance of handling closed-loop systems with care. The advice can be summarized as follows:

- If possible, avoid simple regulators. Adjust the set points (or adjust the regulators) as much as allowed during the experiment.

- Conventional spectral analysis according to Section 8.6 does not work when applied to data from systems operating under feedback.
- Instead use parameter estimation with prediction error models of the type ARX, ARMAX, or BJ (or in tailor-made models). Apply the identification routine in a straightforward fashion to the actual input and the actual output from the process to be identified.

Choice of Sampling Interval

The choice of sampling interval is coupled to the time constants of the system. Sampling that is considerably faster than the system dynamics leads to data redundancy and relatively small information value in the new data points. Sampling that is considerably slower than the system dynamics leads to serious difficulties in determining the parameters that describe the dynamics. In general we can say that it is much worse to sample too slow than too fast. A rule of thumb is to choose a sampling frequency about 10 times the bandwidth of the system (or, rather, the bandwidth of interest for the modeling). This corresponds approximately to placing about 5–8 sampling points over the rise time of the interesting part of the system's step response. It is thus valuable to first obtain a step response of the system.

If we are uncertain of the choice of sampling interval and data collection is "cheap," it is wise to sample fast at the data collection. The decision of which sampling interval to use in model building is then deferred to the computer sessions. The data sequences can then be decimated; that is, every rth value of the original data is chosen before they are used in the identification routine.

In this context it is important to consider how the frequency contents in the signals are affected by sampling. According to Poisson's summation formula (C.12) (or the sampling theorem), the frequencies in the measured signal that are higher than the Nyquist frequency will be (mis)interpreted as lower ones; the *alias effect*. Such frequencies have thus to be filtered out before the sampling by help of a low-pass filter with cut-off frequency just at the Nyquist frequency. Such a filter is also called an *antialias filter*. Note that this also holds if an already sampled signal has to be decimated.

The most important aspects on the choice of sampling interval can be summarized as follows:

- Choose the sampling interval so that it corresponds to 5–8 sampling points over the rise time of the system's step response.
- It is better too sample too fast than too slow.
- Do not forget the antialias filter.

10.3 Posttreatment of Data

The first step in an identification application is to plot the data. One then often discovers that they have certain deficiencies. It might be that the signal levels drift away or that there are high-frequency disturbances above the frequency interval of interest for the system dynamics. There can also be obvious faulty values (*outliers*) among the data. There may also be reasons to enhance certain frequency bands in the data in order to get a better model fit. There thus has to be quite a bit of *posttreatment of data* before the real identification work can start. An important aspect of the posttreatment is also to choose a part from the data that looks good and that thereby is suitable for model fit and model validation.

Drift and High-frequency Disturbances

The linear models that are estimated from data are normally based on signals measured relative to a certain equilibrium. If we work with a tailor-made model with absolute signal levels built in, the signal should of course be kept at their physical units. Otherwise, the rule of thumb is always to subtract at least the mean level in each measured signal. The mean levels also often drift away during the experiment. This can, be eliminated by high-pass filtering.

High-frequency noise in measurements in "uninteresting" frequency bands is really a sign that the sampling interval and the antialias filter have not been adequately chosen. At posttreatment of data they can be removed by low-pass filtering, possibly followed by decimation.

Outliers

In a difficult measurement environment it often happens that some measurement values are obviously incorrect. This is usually most visible in plots of the residuals $y(t) - \hat{y}(t|\hat{\theta}_N)$. If these values were accepted uncritically, they would have a devastating effect on the estimated models. The reason is that quadratic criteria of the type (9.29) give an unreasonably large weight to data points that give large prediction errors. To protect from this, we should use criteria of the type (9.31), where $\ell(\varepsilon)$ behaves quadratically for small- and middle-sized ε, but linearly for large ε.

Consequently, the data have to be evaluated critically. Segments that contain inaccurate or doubtful measurement values should be avoided. If this is impossible, the inaccurate values could be smoothed by hand to interpolated or predicted values.

Prefiltering of Data

In equation (9.48) we saw that the estimate of a linear system can be interpreted as a best fit between the model's frequency function and the true frequency function in a weighted frequency norm (9.49). Assume now that both the input and output are filtered through the same filter before the estimation takes place:

$$y_F(t) = L(q)y(t), \qquad u_F(t) = L(q)y(t) \tag{10.4}$$

This will affect the weighted frequency norm. It changes from

$$\frac{\Phi_u(\omega)}{|H(e^{i\omega}, \theta)|^2} \qquad \text{to} \qquad |L(e^{i\omega})|^2 \cdot \frac{\Phi_u(\omega)}{|H(e^{i\omega}, \theta)|^2} \tag{10.5}$$

If L is chosen as a band-pass filter, the passband will be given priority for the model fit. This is a very valuable possibility when building models with complicated dynamics.

Example 10.2 A Hydraulic Crane
Loading cranes are usually controlled hydraulically. In this example we are going to study a hydraulic log loader. It is approximately 5 meters tall and is controlled by oil being pumped via a valve to a cylinder with a piston, which in turn is coupled to the crane arm. We are especially

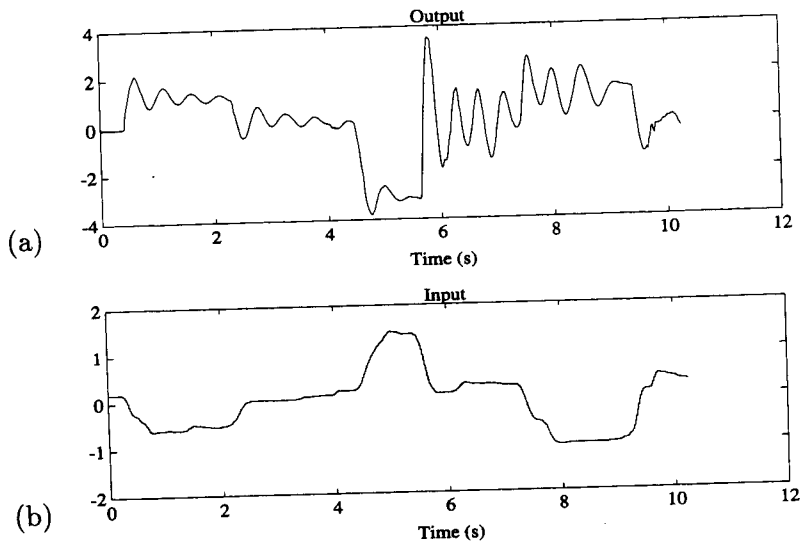


Figure 10.5: Collected data from a hydraulic crane. (a) The pressure in the hydraulic cylinder. (b) Valve position. Time scale: seconds. Sampling interval 0.02 s. Mean values have been subtracted.

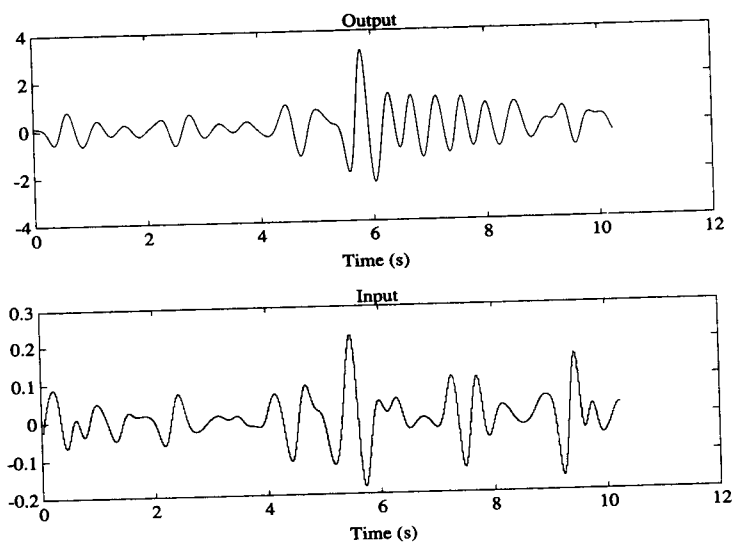


Figure 10.6: Data in Figure 10.5 filtered through a Butterworth filter of 10th order with a passband between 7.5 and 22.5 rad/s.

interested in the mechanical resonances in the crane arm, and we know that these lay somewhere in the interval 8 to 20 rad/s. The input, valve position, and the output, pressure in the hydraulic cylinder, were measured during 20 s., with a sampling intervalm of 0.02 s. The first half of these data are shown in Figure 10.5. Data were also filtered through a band pass filter (10th-order Butterworth) with a passband between 7.5 and 22.5 rad/s. Filtered data are shown in Figure 10.6. A second-order ARX model ($na = 2,\ nb = 1,\ nk = 1$) was estimated for both data sets. The model's Bode diagrams are shown in Figure 10.7. Now, which model is best? We tested both models on the data set that was not used for identification. The models are simulated with the input, and their output is compared with the measured output in Figure 10.8. We see that the model we obtained by help of filtered data is much better in describing the actual resonance oscillations. That the levels differ is, in this case, not so important. The models do not use the output in the simulation, so it can easily happen that the levels separate, especially since low frequencies have not been emphasized in the fit when using filtered data. □

Since the ARX model (9.20) definitely is the most used parametric model, it is important to note that it uses the noise model

$$H(q,\theta) = \frac{1}{A(q)}$$

The weighting function in (10.5) then contains $|A(e^{i\omega})|^2$ in the numerator, which typically gives a clear high-pass character. This implies that the ARX model is adjusted with an emphasis on high-frequency behavior (close to the Nyquist frequency). This is not desirable in general. The effect can be compensated for by prefiltering the input and output through a low-pass or band-pass filter. A simple and useful procedure is to first fit the ARX model to unfiltered data, which will give the model $\hat{B}_N(q)/\hat{A}_N(q)$. After that the input and output are filtered through

$$\frac{1}{\hat{A}_N(q)}$$

and a new ARX model is computed by help of filtered data.

We can summarize the discussion of prefiltering of data by saying that it is advisable to *filter the data through a band-pass filter that has*

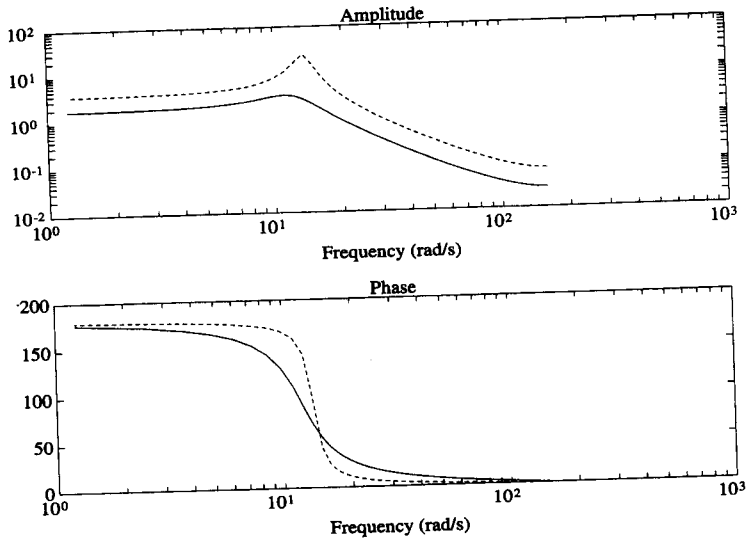


Figure 10.7: Bode diagram for the estimated models. Solid line: ARX model ($na = 2$, $nb = 1$, $nk = 1$) for unfiltered data. Dashed line: prefiltered data.

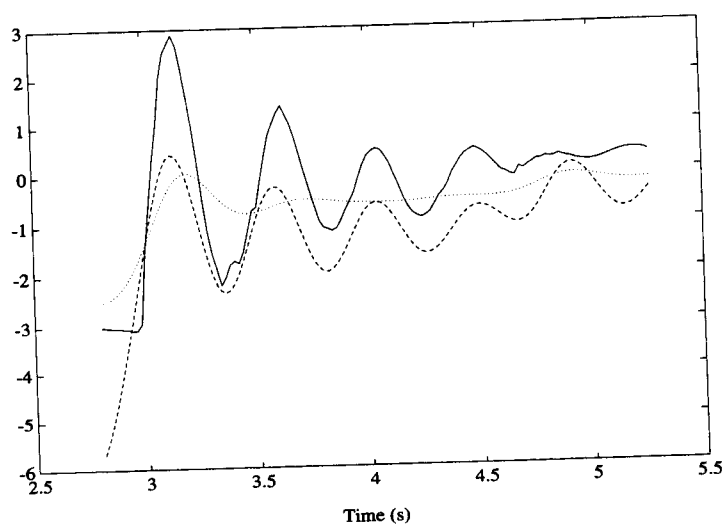


Figure 10.8: Solid line: measured pressure. Dashed line: simulated pressure, according to the model based on filtered data. Dash–dotted line: the same for models from unfiltered data.

a passband covering the interesting frequencies (breaking points in the system's Bode diagram). The effects of low-frequency disturbances, drift, and so on, are thereby reduced. The focus on the model fit is also automatically moved to the most important frequency bands.

10.4 Choice of Model Structure

To choose a model structure that is suitable for identification is perhaps the most difficult decision the user has to make. The choice has several aspects. First the decision has to be made whether to use a tailor-made or a ready-made model. If the latter choice is made, the next question will be whether to use ARX, OE, ARMAX, or BJ or some other model. Finally, we have to decide the orders for the ready-made model. In this section we will give a number of suggestions for making these decisions.

Tailor-made or Ready-made?

To tailor-make a model has one important advantage. The known physical relationships are built in. No parameters have to be wasted estimating what is already known.

The model will be parsimonious with its parameters, which often have a direct physical interpretation. This latter fact has the added advantage that it helps decide if the estimates are reasonable.

The most important disadvantages are that the modeling procedure can be time consuming and that the numerical minimization of the criterion (9.30) can be computationally demanding. The decision whether to tailor-make a model at all might in the end depend on whether the identification package that is used gives relevant support for such an exercise.

Example 10.3 The DC Motor
Data were collected from a dc servo of the type described in Example 9.1 (feedback MSI10). These data are shown in Figure 10.9. The data were adjusted to a ready-made model, ARX, of a second order [$na = nb = 2, nk = 1$ in (9.21)], which gives four parameters to estimate, as well as to the tailor-made model (9.6), where the static gain from voltage to angle velocity first has been determined to 4.51 from

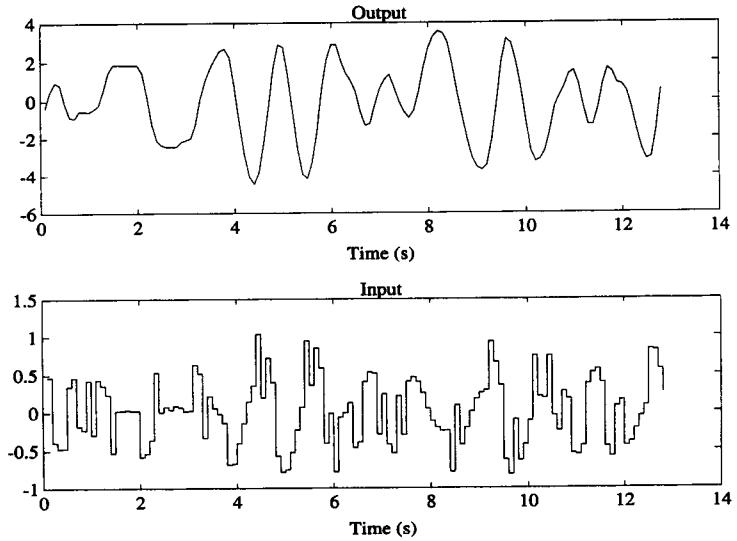


Figure 10.9: Measured data from a dc motor.

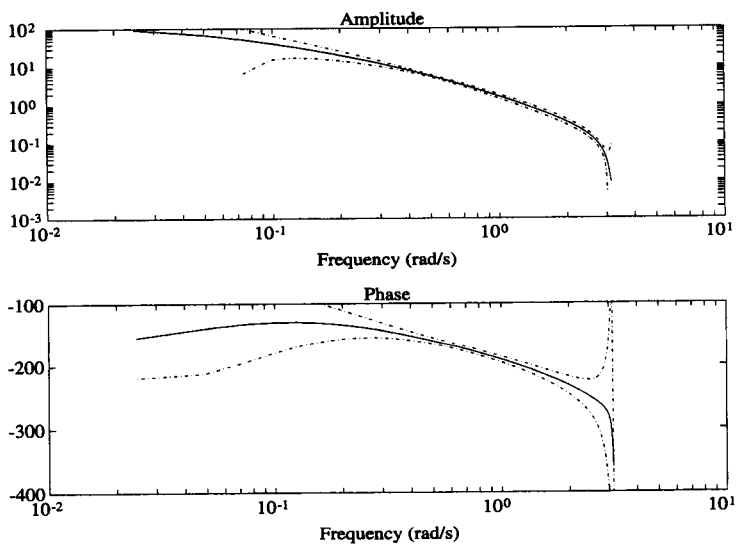


Figure 10.10: Estimated Bode diagram for the ARX model. The dash-dotted line shows a confidence interval corresponding to 3 standard deviations.

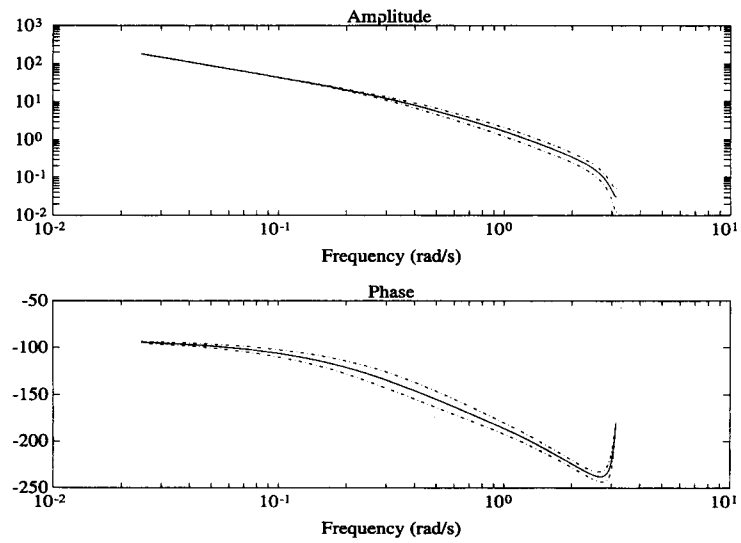


Figure 10.11: Estimated Bode diagram for the tailor-made model. The dash-dotted line shows a confidence interval corresponding to 3 standard deviations.

a step response experiment. The Bode diagrams, with uncertainty regions marked, are shown in the Figures 10.10 and 10.11 for the two models. The uncertainty regions correspond to a 99% confidence interval, computed for the model's parameters according to (9.57) – (9.59) and then translated to the frequency function of the model. The tailor-made model gives better accuracy as shown. This depends on the fact that fewer parameters (only one) have been estimated. It is, however, not known which model is best, since the true dynamics of the motor are unknown. Insight into this can be gained by simulating the two models for a new input and comparing the model outputs with the measured ones. Such curves are shown in Figure 10.12. The tailor-made model thus has an obvious advantage. □

The choice is finally a question of price and quality. A tailor-made model can give high quality, that is, small bias and variance inaccuracy, since it is fitted to the current system (→ small bias) with only a few parameters (→ small variance). But the price in the form of modeling, programming, and computational work can be high. Also, the tailor-made physical model may contain a large number of unknown

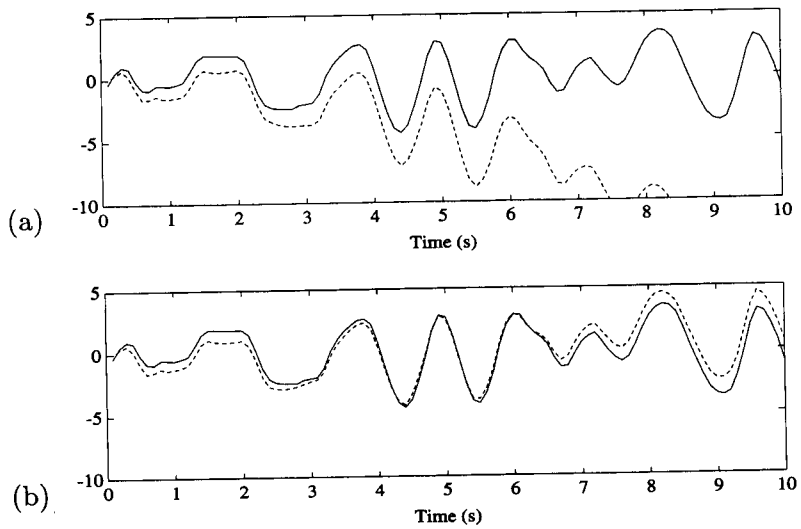


Figure 10.12: Measured output (solid line) and simulated model output (dashed). (a) ARX model. (b) Tailor-made model.

parameters whose individual values are not essential for the dynamics of the system. We saw an example of this in Example 10.1.will not be identifiable.

There are, as always, reasons to remember the principle TSTF (try simple things first). This principle suggests that we should test simple, cheap, ready-made models (for example ARX) first and then go on to more complex ones only if the simple models cannot solve the problem. There are, however, never any excuses for neglecting fundamental physical nonlinearities. However, these can be dealt with in a more simple fashion than by full-fledged physical modeling.

Semiphysical Modeling: Transformation of Raw Data

Even if we do not use a detailed tailor-made model structure, it is important to think through the system's function and basic physics before arbitrary using a ready-made model. This often shows that it is necessary to first do some nonlinear transformations of the original measured data. Such an example is given in Section 10.6. When new inputs and outputs have been constructed from the raw measurements, by nonlinear trasnformations they can be used in simple

model structures, such as the ARX structure. We term this combination of transformations based on physical insights and simple model structures *"semi-physical modeling"*.

Other reasons for nonlinear measurement data transformations can be to:

- Compensate for nonlinear sensors
- Compensate for nonlinear actuators

If, for example, a valve saturates so that it actuates $p(u(t))$ instead of the commanded input $u(t)$, then use $p(u(t))$ as input in the model.

Which Kind of Ready-made Model?

In Section 9.2 we described a number of different ready-made models. Which one should we choose? There is no general answer, instead the choice depends on data. In practice some different structures are tested and compared to the obtained models with the methods described later. However, some general points of view can be listed:

- The ARX model $A(q)y(t) = B(y)u(t) + e(t)$ is the easiest to estimate since the corresponding estimation problem is of a linear regression type (see Section 9.2). According to the TSTF principle, we should start by testing ARX models. The foremost disadvantage is that the disturbance model $H(q, \theta) = 1/A(q)$ comes along with the system's poles. It is therefore easy to get an incorrect estimate of the system dynamics because the A polynomial also has to describe the disturbance properties. Higher orders in A and B than necessary may be required. If the signal-to-noise ratio is good, this disadvantage is less important.
- The ARMAX model $A(q)y(t) = B(q)u(t) + C(q)e(t)$ gives extra flexibility to handle disturbance modeling because of the C polynomial. This is an often used model.
- The OE model $y(t) = \frac{B(q)}{F(q)}u(t) + e(t)$ has the advantage that the system dynamics can be described separately and that no parameters are wasted on a disturbance model. If the system operates without feedback during the data collecting, a correct

description of the transfer function $G(q) = B(q)/F(q)$ can be obtained regardless of the nature of the disturbance. However, minimization of the criterion function (9.29) can be more difficult than in the ARMAX case.

- The BJ model $y(t) = [B(q)/F(q)]u(t) + [C(q)/D(q)]e(t)$ is the "complete model" in which the disturbance properties are modeled separately from the system dynamics.

The models ARX and ARMAX have common dynamics (common poles) for the noise $e(t)$ and the input $u(t)$. This is suitable when the dominating disturbances enter "early" in the process, for example at the input. Correspondingly, the BJ model is preferable when modeling disturbances come in "late" to the process, for example, as measurement noise in the output.

Comparing Model Structures

When some models in different model structures have been calculated, the obvious question arises of how to compare them. Which one is best? The first thought is to compare the corresponding values of the criterion function: Which prediction error variance did they lead to? This question requires careful consideration. It is best to evaluate prediction error variances of the different models when they are confronted with *new* data sequences (that is, other than the ones used in the model estimation). In statistical terms this is called *cross validation.* This is a straightforward and natural method. If a model is able to predict a fresh set of data better than another, it should be treated as a better model.

The only disadvantage with cross validation is that a fresh amount of data has to be reserved for the model comparison, and therefore we cannot use all available information to estimate the model.

If the prediction error comparison has to be made on the basis of the data that already have been used at the model estimation (second hand data), a few problems emanate. A larger model will always give a lower value of the criterion function (9.29), since it has been obtained by minimizing over more parameters. If the values of the criterion function are plotted (the prediction error variances) as a function of the model order, a strictly decreasing function is obtained. To start

with, the value V_N decreases since the model includes more and more of the system's relevant properties. But even after a "correct" model order has been passed, the criterion function continues to decrease. The reason is that the extra — and unnecessary — parameters are used to fit the model to the specific disturbance signals in the present data set. This is called *overfit* and does not serve any purpose since the model will be used when other disturbances affect the system. On the contrary, the model will in fact be worse because of the overfit.

The goal is to find the transition from relevant model fit to overfit. A number of different methods for this have been suggested in the literature. In general, they are based on fundamental information theoretical principles. They all have the following characteristic:

$$\min_{d,\theta} f(d,N) \sum_{t=1}^{N} \varepsilon^2(t,\theta) \tag{10.6}$$

Here N is the number of data and d is θ's dimension (the number of estimated parameters). The function $f(d,N)$ increases with d and decreases with N. Minimizing (10.6) with respect to both d and θ will penalize a model that contains many parameters.

The model that finally is selected — the one that minimizes (10.6) — has to represent a balance between model fit and the number of parameters used.

The most well-known methods are the following:

1. *Akaike's information criterion: AIC*

$$\min_{d,\theta}(1+\frac{2d}{N}) \sum_{t=1}^{N} \varepsilon^2(t,\theta) \tag{10.7}$$

2. *Final prediction error: FPE*

$$\min_{d,\theta} \frac{1+d/N}{1-d/N} \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t,\theta) \tag{10.8}$$

This is a statistical estimate of the prediction error variance we would get for a fresh data set using the model $\hat{\theta}_N$.

3. *Rissanen's minimal description length*

$$\min_{d,\theta}(1 + \frac{2d}{N} \cdot \log N) \sum_{t=1}^{N} \varepsilon^2(t, \theta) \tag{10.9}$$

This criterion aims at minimizing the code needed to store data. A parameter-rich model is also penalized by the fact that the model's parameters have to be stored. Hence there is a higher penalty for model complexity.

Besides comparing models in terms of the prediction variances that they produce, the models can be simulated with a given input and then compared as to how well they describe the corresponding measured output.

For linear models it is also instructive to compare their Bode diagrams and pole–zero diagrams. A comparison between the frequency function that is obtained by spectral analysis and the functions produced by the different parametrized models is particularly useful.

Choice of Order of a Ready-made Model

To determine the order and delays of a ready-made model, the following procedure is quite useful.

1. First get a reasonable estimate of the delay by correlation analysis and/or by testing all reasonable ones in a, say, fourth-order ARX model. Pick the delay that gives the best model performance (sum of squared prediction errors on a validation set).

2. Then test many ARX models of different orders with this delay. Pick the orders that give the best model performance. This is done efficiently in several identification packages.

3. This model may be of unnecessarily high order to describe the system dynamics, since the poles of an ARX model also describe the noise properties. Thus plot the zeros and the poles of the resulting model (with uncertainty regions marked) and look for cancellations. The surviving numbers of poles and zeros give us indications of the necessary order for the dynamics from input to output. Then try ARMAX, OE, or BJ-models with this order for G and first- or second-order models for the noise characteristics.

Getting Started

To get a feel for how difficult it will be to find a good model structure for a certain data set, the following pragmatic route can be followed:

1. Get an indication of the time delay of the system, for example from an impulse response estimate using correlation analysis [see algorithm (8.11)].

2. Compute a fourth-order ARX model using this delay, and compare the measured output with the simulated output from the model using a separate validation data set. If the model/system is unstable or has some very long time constants, use the model's predicted output over a reasonably long prediction horizon. Now either of two things happens:

   - The comparison looks good
   - The comparison does not look good

   In the former case we can go to the procedure described in the previous subsection and fine-tune the choice of orders and structures. In the second case we have to go deeper into the physics of the application.

3. Test higher-order linear models. This might be necessary in particular for systems with mechanical resonances. If this does not help, go on to the next step.

4. Try to find out if there are further signals that affect the output. If they can be measured, include them as additional input signals. If this does not lead to a better agreement between the measured and the model output, go on to the next step.

5. Apply semiphysical modeling to come up with essential nonlinear transformations of the raw measurements. This is of course highly application dependent. An example will be given in Section 10.6.

10.5 Model Validation

To *validate* a model is to investigate if it can be accepted, given its intended use. This is closely related to the concept of model quality, and we shall in this section discuss some methods to study the quality of a model.

Model Quality

In Section 9.4 we discussed model quality. Model validation ties in closely to the aspects we mentioned there

The model's stability: Computing models from different measurement records and in different structures and then evaluating their input-output properties (for example in a Bode diagram or in simulation) is an important tool for gaining confidence in the model. If approximately the same model properties are obtained under such varied conditions, we should feel confident that the model found some significant features of the system.

It is especially informative to plot the model's frequency function in the same diagram as the one obtained by spectral analysis. They should show reasonable agreement. But remember that spectral analysis estimation is unreliable if the system operated under feedback when the data were collected.

The model's ability to reproduce the system's behavior: A natural test is to simulate the model from input only and then compare the simulated output with the measured one. This comparison is best made on a fresh data set. That is what we did in Example 10.3.

The corresponding comparison can also be made for k-step prediction of the output based on the model. The k-step-ahead predictor predicts $y(t)$ based on the model and on the information in $y(t-k), y(t-k-1), \ldots$ as well as in $u(t-1)$, $u(t-2), \ldots,$. It can thus be thought of as a simulation k time steps ahead. By picking the prediction horizon k larger than the time constants of the system, we can thus test if these have been correctly picked up by the model.

For $k = 1$, the model's short-term behavior (the high-frequency properties) is evaluated. A comparison plot between measured output and predicted output can look rather impressive even for the trivial model $\hat{y}(t|\theta) = y(t-1)$. One-step-ahead prediction is better evaluated

by analysis of the errors, the residuals.

Residual Analysis

Let us examine the *residuals* that the model leaves behind, that is, the prediction errors

$$\varepsilon(t) = \varepsilon(t, \hat{\theta}_N) = y(t) - \hat{y}(t|\hat{\theta}_N) \tag{10.10}$$

These should ideally be independent of the input. If this is not the case, then there are components in $\varepsilon(t)$ that stem from $\{u(t)\}$, which means that there are more system dynamics to describe than the model $\hat{y}(t|\hat{\theta}_N)$ picked up. Typically, we then form

$$\hat{R}_{\varepsilon u}(\tau) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t+\tau)u(t), \quad |\tau| \le M \tag{10.11}$$

and test whether these numbers are close enough to zero. It can be shown that if $\{\varepsilon(t)\}$ and $\{u(t)\}$ are really independent, then (10.11) is for large N approximately normally distributed, with mean zero and the variance

$$P_\tau = \frac{1}{N}\sum_{k=-\infty}^{\infty} R_\varepsilon(k)R_u(k) \tag{10.12}$$

Here R_ε and R_u are ε's and u's covariance functions. $\hat{R}_{\varepsilon u}(\tau)$ is usually drawn in a diagram together with the lines $\pm 3 \cdot \sqrt{P_\tau}$. That $R_{\varepsilon u}(\tau)$ goes outside these lines for any τ is an indication that $\varepsilon(t+\tau)$ and $u(t)$ probably are dependent for this value of τ.

When examining the correlation functions, note the following points:

- If there is a correlation for negative values of τ, that is, the influence from $\varepsilon(t)$ on later values of $u(s), s > t$, this is an indication that the data were collected during feedback, not that the model is incomplete.

- If an ARX model (9.20) and the expression (10.11) are computed for the same data, then $\hat{R}_{\varepsilon u}(\tau) = 0$ for $\tau = nk, \ldots, nk + nb - 1$. The least squares estimates are constructed in this way.

- If an ARX model is used and $\hat{R}_{\varepsilon u}(\tau)$ is significantly different from zero for a value τ_0, it is an indication that the term $u(t-\tau_0)$ should be included in the model. This can be a good tip for the choice of nk and nb.

If also a model of the disturbance signal's properties is sought, we should demand that the residuals be mutually independent. This can be tested analogously by computing and plotting

$$\hat{R}_\varepsilon(\tau) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t)\varepsilon(t+\tau) \tag{10.13}$$

The residual analysis will of course be more revealing (and more demanding for the model) if the residuals are computed for a set of data that have not been used in the model estimation.

Validation Decisions

In summary we can say that it is important to get as much advice as possible from available validation possibilities. The final validation criterion is, however, that the model is working well when it is used for its purpose: simulation, analysis, control design, or whatever.

10.6 An Example

Finally, we illustrate the choice of model structure with the following example in which both ready-made and simple tailor-made models are tested.

Example 10.4 Solar-heated House Dynamics
Consider the problem to identify the dynamics of a solar-heated house described in Example 3.9. The function of the system is to let the sun heat the air in the solar panel. The air is then pumped to the storage tank, from where the heat can be taken later for heating the house. Introduce the following variables:
$I(t)$: *Solar intensity at time t*
$x(t)$: *Temperature in the solar panel*
$y(t)$: *Temperature in the storage tank (at the inlet)*
$u(t)$: *Pump speed*

We want to build a model to show how the temperature of the storage tank y is affected by the pump speed and the solar intensity. For this purpose, measurement data were collected for a 3-day and night period. The sampling interval was chosen as 10 minutes. The data are shown in Figure 10.13.

We first test a model of the type (9.20):

$$y(t)+a_1y(t-1)+a_2y(t-2) = b_1u(t-1)+b_2u(t-2)+c_1I(t-1)+c_2I(t-2) \tag{10.14}$$

The least squares estimates of the coefficients a_i, b_i, and c_i based on 450 sampling points ($N = 450$) gave the values

$$\begin{array}{lll} a_1 = -0.772 & a_2 = -0.074 & b_1 = 1.030 \\ b_2 = -0.721 & c_1 = -0.066 & c_2 = 0.292 \end{array} \tag{10.15}$$

This is thus the best model of the type (10.14). Is it good enough? In order to test this, the model (10.14)–(10.15) is simulated with the real inputs u and I, and the simulated output y_M is compared to the measured one. The comparison is shown in Figure 10.14a. The model is obviously not very good. It is unable to follow any essential changes in the temperature.

So far we have not used any physical insight for the heating process, but assigned the linear ready-made model (10.14) without further consideration. But it soon becomes evident that a linear model is not very realistic. The effects of the solar intensity and the pump speed are obviously not additive. When the pump stands still, the solar radiation has no effect at all on the storage tank's temperature.

Let us think about what happens in the heating system. With some simplifications the developments can be described as follows in discrete time:

Energy balance means that the heating of the air in the solar panel $[x(t+1) - x(t)]$ is equal to the heat produced by the sun $[d_2\, I(t)]$ minus the heat loss to the environment $[d_3\, x(t)]$ minus the heat transferred to the storage tank $[d_0\, x(t)\, u(t)]$:

$$x(t+1) - x(t) = d_2I(t) - d_3x(t) - d_0x(t)u(t) \tag{10.16}$$

Similarly, the heating of the storage tank $[y(t+1) - y(t)]$ is equal to supplied heat $[d_0x(t)u(t)]$ minus heat losses $[d_1y(t)]$:

$$y(t+1) - y(t) = d_0x(t)u(t) - d_1y(t) \tag{10.17}$$

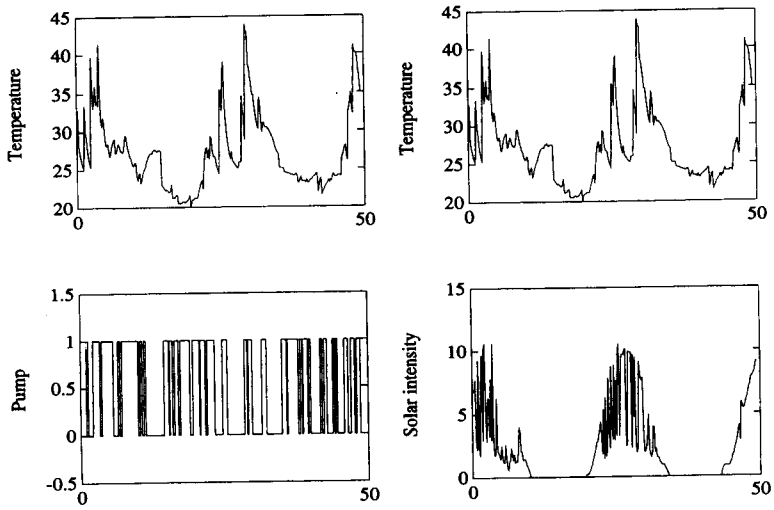


Figure 10.13: Storage temperature y, pump speed u, and solar intensity I for a 2-day period. Sampling interval: 10 min.

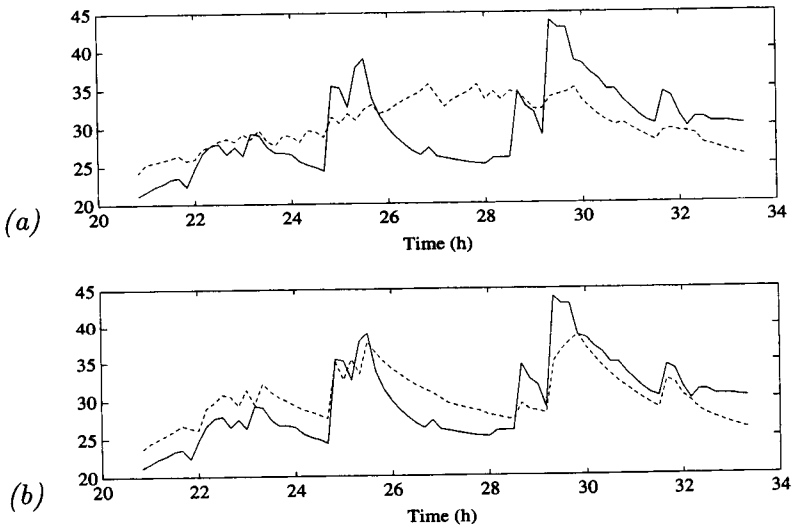


Figure 10.14: Measured temperature (solid line) and the output from the model (dashed line). (a) The linear model (10.14)–(10.15). (b) The nonlinear model (10.21)–(10.22).

In equations (10.16) and (10.17) d_0, d_1, d_2 *and,* d_3 *are unknown constants whose numerical values we want to determine. But the temperature* $x(t)$ *was not measured, so the equations cannot be used directly. We have to eliminate* $x(t)$*. From equation (10.17),* $x(t)$ *is determined to be*

$$x(t) = \frac{y(t+1) - y(t) + d_1 y(t)}{d_0 u(t)} \tag{10.18}$$

If we substitute this expression into (10.16), we have after some calculations

$$\begin{aligned} y(t) = (1-d_1)y(t-1) + (1-d_3)\frac{y(t-1)u(t-1)}{u(t-2)} \\ +(d_3-1)(1-d_1)\frac{y(t-2)u(t-1)}{u(t-2)} + d_0 d_2 u(t-1)I(t-2) \\ -d_0 u(t-1)y(t-1) + d_0(1-d_1)u(t-1)y(t-2) \end{aligned}$$

In this expression we do not have a simple linear relation between the coefficients d_i *and the output. We therefore introduce the reparameterization*

$$\begin{array}{ll} \theta_1 = (1-d_1) & \varphi_1(t) = y(t-1) \\ \theta_2 = (1-d_3) & \varphi_2(t) = (y(t-1)u(t-1))/u(t-2) \\ \theta_3 = (d_3-1)(1-d_1) & \varphi_3(t) = (y(t-2)u(t-1))/u(t-2) \\ \theta_4 = d_0 d_2 & \varphi_4(t) = u(t-1)I(t-2) \\ \theta_5 = -d_0 & \varphi_5(t) = y(t-1)u(t-1) \\ \theta_6 = d_0(1-d_1) & \varphi_6(t) = u(t-1)y(t-2) \end{array} \tag{10.19}$$

giving us a model that is linear in its parameters:

$$y(t) = \theta_1\varphi_1(t) + \cdots + \theta_6\varphi_6(t) \tag{10.20}$$

which is of the form (9.26). The price for the linear expression in θ *is that the knowledge of the algebraic relationships between* θ_i, $i = 1, \ldots, 6$, *according to (10.19) have been lost. The least squares estimation of* θ *does not give a* θ *value that directly corresponds to values of* d_i, $i = 0, 1, 2, 3, 4$.

If we adjust θ_i *in (10.20) to the same* u, I, *and* y *data as before, we obtain the result*

$$\begin{array}{lll} \hat{\theta}_1 = 0.887 & \hat{\theta}_2 = 0.04 & \hat{\theta}_3 = 0.056 \\ \hat{\theta}_4 = 0.756 & \hat{\theta}_5 = -0.243 & \hat{\theta}_6 = -0.002 \end{array} \tag{10.21}$$

(When the pump was off, the value of u was given a small postive number rather than zero.) The model (10.20)-(10.21) is then simulated with the real inputs u and I according to

$$\begin{aligned} y_M(t) = \hat{\theta}_1 y_M(t-1) + \hat{\theta}_2 \frac{y_M(t-1)u(t-1)}{u(t-2)} \\ +\hat{\theta}_3 \frac{y_M(t-2)u(t-1)}{u(t-2)} + \hat{\theta}_4 u(t-1)I(t-2) \\ +\hat{\theta}_5 u(t-1)y_M(t-1) + \hat{\theta}_6 u(t-1)y_M(t-2) \end{aligned} \quad (10.22)$$

The model's output y_M is compared to the measured input y in Figure 10.14. As shown, the model (10.22) gives a reasonable description of the system. If we know the solar radiation and pump speed for the next 24 hours, we can predict the storage tank's temperature with an accuracy of $3°C$ on the average. This is acceptable in order to test control strategies and to evaluate the heating capacity over longer time periods. □

The example shows the importance of using physical insights during the identification work. Regardless of how accurate the data we use and which ready-made models we test, it is impossible to get an acceptable description of the system until, in this case, the important nonlinearities are built into the model structure.

The example also shows that the physical insight does not have to be worked into the model structure during a laborious modeling process. Here we reached our goal by doing simple and rather obvious nonlinear transformations of raw data.

10.7 Conclusions: The Possibilities and Limitations of Identification

System identification has proved to be a convenient and useful tool for model building. Well over 10,000 successful applications are published in the literature. These cover widely varying areas from process industry, ship dynamics, signal processing, and seismology to biomedicine, ecology, and econometrics. Possibilities to handle completely unknown systems with the help of ready-made models, as well as carefully tailor-made model parameterizations, give a very wide area of applications.

There are two main limitations:

- We have to have informative data.
- We have to be able to find a suitable model structure.

In this chapter we have discussed how both these conditions can be met. But in some cases there are limitations in the application or in our knowledge that make it impossible to fulfill the conditions.

There are several reasons why the measured data can be unsatisfactory. A common reason is that we are unable to make experiments and to manipulate essential inputs. Another is that there are considerable, nonmeasurable disturbances that do not fit into the standard pattern of stationary processes. A third reason is measurement problems, and a fourth is that the system's properties may vary in time.

If the system, at the operating point in question, can be described by a ready-made model, the prospects are good for finding such a model. If the model has to be based on physical insights — Example 10.4 shows such a case — it is necessary to have so much understanding of the system that basic nonlinear elements can be built into the model structure.

Bibliography for Part III

The methodology of building models of dynamical systems with the help of system identification tools is described in detail in, for example,

L. Ljung: *System Identification – Theory for the User.* Prentice Hall, Englewood Cliffs, N.J., 1987.

T. Söderström and P. Stoica: *System Identification.* Prentice Hall International, London, 1989.

The spectral analysis method is treated from an engineering point of view in

J. S. Bendat and A. G. Piersol : *Engineering Applications of Correlation and Spectral Analysis.* Wiley, New York, 1980.

A more theoretical point of view of this area is presented in

D. R. Brillinger: *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco, 1981.

A classical book on model building with practical methods is

G. E. P. Box and G. M. Jenkins: *Time Series Analysis, Forecasting and Control*, 2nd Ed. Holden-Day, San Francisco, 1989.

The MATLAB-based software for system identification is described in

L. Ljung: *The System Identification Toolbox, User's Guide*, 3rd ed. The MathWorks, Inc., Natick, Mass., 1992.

Part IV

Simulation and Model Use

How to Use Models for Simulation

In Parts II and III we have presented methods for system modeling. These models can be used for a number of purposes, for example:

1. Increasing the understanding of the system itself
2. Predicting the future behavior of the system
3. Carrying our technical computations for control design
4. Optimizing constructions
5. Studying human-machine cooperation.

Regardless of the purpose, we almost always want to study the solutions to the model equations. We saw in Chapter 7 on computer support that in certain cases explicit solutions to differential equations can be generated with computer algebra methods. For most models, however, the complexity is such that some form of approximation is needed. The classical way of attacking the problem has been to solve the equations using analog computers. Today the equations are most often solved numerically in a computer. Regardless of the method, we usually call the solving of the equations, together with the presentation of the result, *simulation.* We present different simulation methods in Chapter 11. Modern technology gives a wide range of possibilities for presenting the results from simulations in a striking way so that the real system's behavior is illustrated. We describe them briefly, also in Chapter 11. This part is concluded by Chapter 12 with a discussion of model use and to what extent the model can be trusted.

Chapter 11

Simulation

11.1 Short Review

The earlier simulation tools were usually based on analog techniques. The different variables in the model corresponded to physical quantities like positions, angles, pressures, voltages, and so on. The user then tried to construct a device that gave the same relationship between the variables as the mathematical model. During the 1800s several mechanical tools of this kind were used. They were often specially constructed for a certain purpose, for example, Lord Kelvin's tide predictor from 1879. Gradually, more general devices, which could be used for a whole class of problems, were constructed. An example of this is a mechanical differential analyzer built by Vannevar Busch at MIT about 1930. The advances in electronics opened vast possibilities to build flexible analyzers. These were called analog computers. After World War II electronic analog computers became widely used, in the airplane industry, in particular. The transistorization of electronics in the 1960s improved the analog computers. At the same time they got competition from digital computers. Since the introduction of workstations and simulation languages suitable for interactive work, digital computers have become much more effective and flexible simulation tools.

11.2 Scaling

We have seen that mathematical models can be expressed in different ways. Transfer functions, weighting functions, a higher order differential equation, or a system of first order differential equations can be used. At the simulation it is often advantageous, and in some cases necessary, that the model be written in a state form. If that is not the case, the first step is to get a state-space description using one of the earlier stated methods.

The simulation work is made easier through *scaling* of the equations. Scaling means that in the equations

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \ldots, x_n), \qquad i = 1, \ldots, n \tag{11.1}$$

a variable change is made

$$\tilde{x}_i = \alpha_i x_i, \quad \tau = \beta t \tag{11.2}$$

corresponding to *amplitude* and *time scaling*, respectively. This means that the following new functions are considered

$$\tilde{x}_i(\tau) = \alpha_i x_i \left(\frac{\tau}{\beta}\right), \quad i = 1, \ldots, n \tag{11.3}$$

Differentiating this expression gives the differential equations

$$\frac{d\tilde{x}_i(\tau)}{d\tau} = \frac{\alpha_i}{\beta} f_i\left(\frac{\tilde{x}_1}{\alpha_1}, \frac{\tilde{x}_2}{\alpha_2}, \ldots, \frac{\tilde{x}_n}{\alpha_n}\right), \qquad i = 1, \ldots, n \tag{11.4}$$

The scaling can be used to get the same order of magnitude of the variables and thereby avoid numerical problems. Another use of scaling is to reduce the number of cases that we need to simulate.

Example 11.1 Simulation of a Tank
As an example we can study the tank in Section 2.3. We found there that the mathematical model is

$$\frac{dh}{dt} = \frac{u}{A} - \frac{a}{A}\sqrt{2gh}$$

Now assume that we want to investigate how an originally empty tank is filled with a constant input flow u for different parameter values in

the model. At first glance, this seems to be an extensive amount of work since there are three parameters to vary: a, A, and u. If each parameter has to assume N values, N^3 simulations are, in principle, needed. However, if the model's behavior under scaling is utilized, only one simulation is needed. Apply

$$\tilde{h} = \frac{h}{H}, \qquad \tau = \frac{t}{T}, \quad \text{i.e., } \tilde{h}(\tau) = \frac{h(T\tau)}{H}$$

where H and T are scaling factors. We then have

$$\frac{d\tilde{h}}{d\tau} = \frac{Tu}{AH} - \frac{Ta}{A} \cdot \sqrt{\frac{2g}{H}} \cdot \sqrt{\tilde{h}}$$

If we choose H and T so that $\frac{Tu}{AH} = 1$ and $\frac{Ta}{A} \cdot \sqrt{\frac{2g}{H}} = 1$, that is, we take

$$H = \frac{u^2}{2ga^2}, \qquad T = \frac{AH}{u} = \frac{Au}{2ga^2},$$

then we get the equation

$$\frac{d\tilde{h}}{d\tau} = 1 - \sqrt{\tilde{h}}$$

All we need to do now is to simulate this equation and draw the result as shown in Figure 11.1.

When we then want to utilize the curve for a special problem, we compute the values of H and T for the case in question and can then go from τ and $\tilde{h}$ to t and h. We can also establish that H physically means the equilibrium the fluid assumes in the case in question. It can thus be considered as a natural unit of length for the model. In the same way, T is the quotient between the equilibrium volume and the flow, which is a natural time unit. □

Example 11.2 Simulation of a Pendulum

Let us also consider the pendulum in Example 6.4. If we make the approximation $\sin\theta \approx \theta$ and eliminate ω, we obtain the differential equation

$$\frac{d^2\theta}{dt^2} + \frac{b}{m\ell^2}\frac{d\theta}{dt} + \frac{g}{\ell}\theta = 0$$

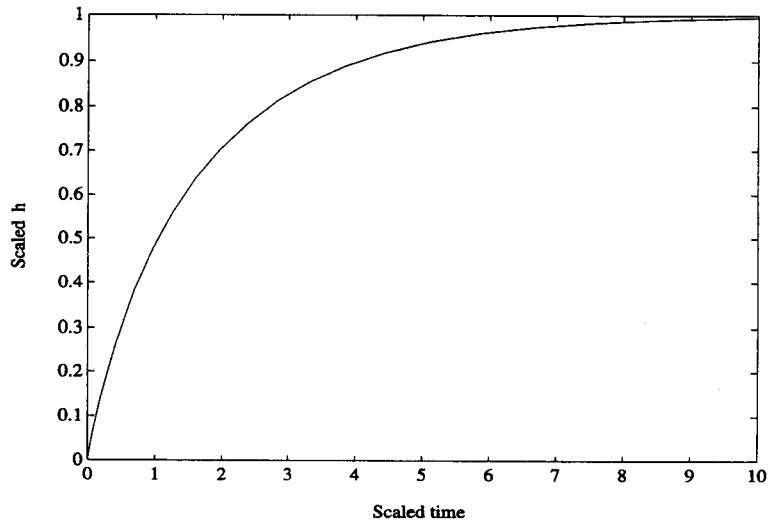


Figure 11.1: Step response of a tank.

Let us say that we want to simulate the case

$$\theta(0) = A, \quad \dot{\theta}(0) = 0$$

If we want to investigate the system's behavior for different values of the physical parameters, it is necessary to simulate all possible combinations of m, ℓ, b, and A. With the scaling $\tilde{\theta}(\tau) = \theta(t_0\tau)/\theta_0$ we have

$$\frac{d^2\tilde{\theta}}{d\tau^2} + \frac{t_0 b}{m\ell^2} \cdot \frac{d\tilde{\theta}}{d\tau} + \frac{t_0^2 g}{\ell}\tilde{\theta} = 0$$

$$\tilde{\theta}(0) = A/\theta_0, \quad \frac{d\tilde{\theta}}{d\tau}(0) = 0$$

The choice

$$t_0 = \sqrt{\frac{\ell}{g}}, \quad \theta_0 = A$$

gives

$$\frac{d^2\tilde{\theta}}{d\tau^2} + 2\zeta\frac{d\tilde{\theta}}{d\tau} + \tilde{\theta} = 0$$

$$\tilde{\theta}(0) = 1, \quad \frac{d\tilde{\theta}}{d\tau}(0) = 0$$

where

$$\zeta = \frac{b}{2m\ell^2}\sqrt{\frac{\ell}{g}}$$

It is obviously enough to simulate the scaled equation for different values of ζ. The parameter ζ, which decides the essential characteristic of the simulated curves, is called the relative damping. We also note that instead of using t_0, one could use

$$\omega_0 = \frac{1}{t_0} = \sqrt{\frac{g}{\ell}}$$

which becomes a scaling factor for the angular velocity. □

The two examples show the usefulness of both amplitude and time scaling as a way of making "standardized simulations". In addition the scaling can be physically necessary when the simulation takes place in terms of analog, physical quantities. Scaling can also give considerably better numerical properties in digital simulation.

11.3 Block Diagrams

In simulation it can sometimes be advantageous to think about a differential equation

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \ldots, x_n, u_1, \ldots, u_m) \\ &\vdots \\ \dot{x}_n &= f_n(x_1, \ldots, x_n, u_1, \ldots, u_m) \end{aligned} \tag{11.5}$$

as an integral equation

$$\begin{aligned} x_1(t) &= \textstyle\int_{t_0}^{t} f_1(x_1, \ldots, x_n, u_1, \ldots, u_m)d\tau \\ &\vdots \\ x_n(t) &= \textstyle\int_{t_0}^{t} f_n(x_1, \ldots, x_n, u_1, \ldots, u_m)d\tau \end{aligned} \tag{11.6}$$

This equation involves two types of operations: static nonlinear transformations and integrations (see Figure 11.2). With this viewpoint equations (11.5), (11.6) are represented as block diagrams, according to Figure 11.3.

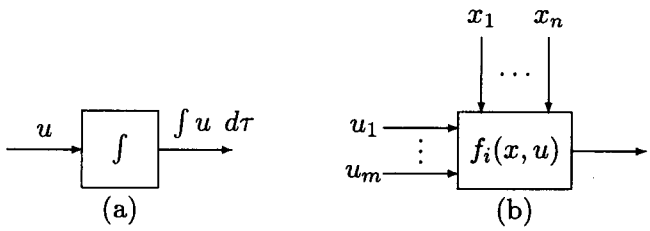


Figure 11.2: (a) Integrator and (b) static nonlinear element.

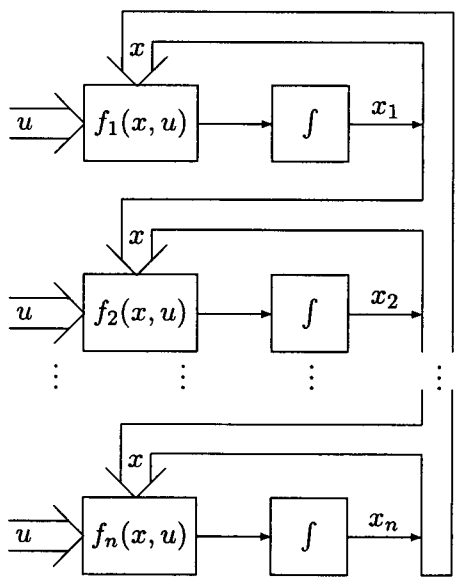


Figure 11.3: Block structure of a dynamic system.

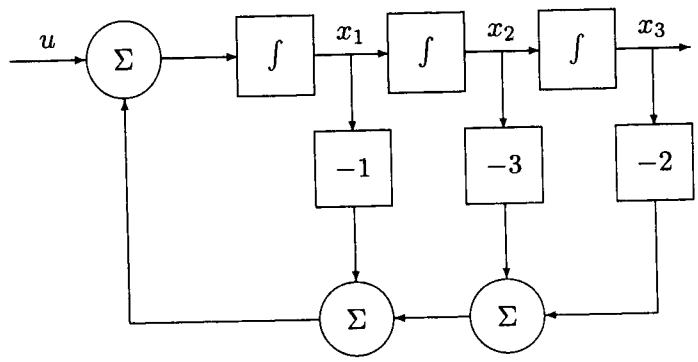


Figure 11.4: Block diagram for a linear system.

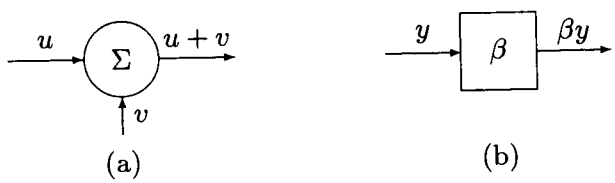


Figure 11.5: (a) Summation and (b) multiplication by a constant.

This figure also emphasizes the systems theoretically interesting fact that a general dynamic system can be constructed from simple elements, which are pure integrators, connected through static nonlinearities. The structure will be particularly simple for linear differential equations. The system

$$\begin{aligned} \dot{x}_1 &= -x_1 - 3x_2 - 2x_3 + u \\ \dot{x}_2 &= x_1 \\ \dot{x}_3 &= x_2 \end{aligned} \tag{11.7}$$

will thus have a block diagram representation according to Figure 11.4. Here we see that, aside from the integrators, we only need two other operations — summation and multiplication with a constant (see Figure 11.5).

For nonlinear systems where the righ side consists of a polynomial, we also need a multiplication element (see Figure 11.6). As an exam-

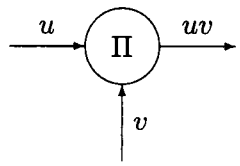


Figure 11.6: Multiplication element.

ple the system

$$\begin{aligned}\dot{x}_1 &= -x_2 + 2x_1^3\\ \dot{x}_2 &= x_1 + 3x_1x_2\end{aligned}$$

can be described by the block diagram in Figure 11.7.

Block diagrams of the types shown in Figures 11.4 and 11.7 can be used to directly simulate in *hardware*. Each block in the diagram will then have its counterpart in a physical component. In Lord Kelvin's tide predictor, the integrators for example were represented by mechanical devices with rotating discs and spheres. In electronic analog computers, summations and integrations are performed by operational amplifiers, while multiplication with constants are carried out by potentiometers.

Block diagrams can form the basis of an implementation in digital hardware as well, where integrations, summations, and multiplications correspond to suitably connected registers. An advantage with this approach is that an implementation is obtained, in which all operations take place in parallel. A system with 100 states thus takes no more time to simulate than one with one state (but requires more hardware).

A block diagram representation may also be interesting for software implementation of the simulation, since it can be a convenient method to enter the system description. The block diagram is then often presented in a more aggregated form. The linear parts are combined into units that are represented by their transfer functions. Let us consider a nonlinear variant of (11.7):

$$\begin{aligned}\dot{x}_1 &= -x_1 - 3x_2 - 2x_3 + (u - x_3)^3\\ \dot{x}_2 &= x_1\\ \dot{x}_3 &= x_2\end{aligned}$$

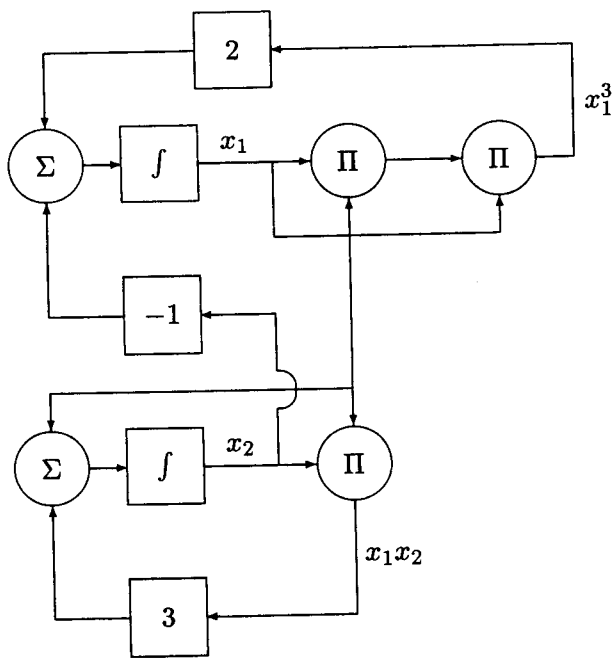


Figure 11.7: Block diagram for a polynomial, nonlinear system.

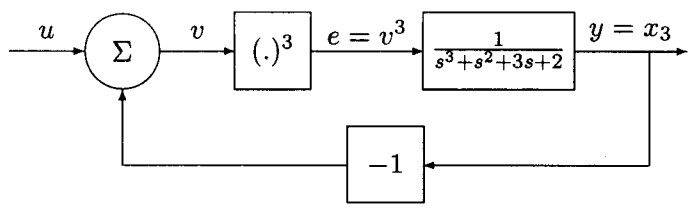


Figure 11.8: Block diagram for a linear system with nonlinear feedback.

This system can be regarded as the linear system

$$\begin{aligned}\dot{x}_1 &= -x_1 - 3x_2 - 2x_3 + e\\ \dot{x}_2 &= x_1\\ \dot{x}_3 &= x_2\\ y &= x_3\end{aligned}$$

with the transfer function

$$\frac{1}{s^3 + s^2 + 3s + 2}$$

from e to y, with a cubic feedback gain, according to Figure 11.8.

11.4 Connecting Subsystems

In Section 11.3 we saw that a block diagram description of a model could be quite useful. We can also recall from the discussion of modeling in Chapter 4 that a block diagram description can give insight into the structure of the system and the interdependence of different parts. There are, however, also some pitfalls in this view of systems. Throughout this book we have said that we consider the state space formulation

$$\begin{aligned}\dot{x} &= f(x, u)\\ y &= h(x, u)\end{aligned} \tag{11.8}$$

to be the natural way of describing a system. Suppose we have several subsystems described in this way and connect them. Will the resulting system then also have a state-space description? Will it be easy to calculate that description?

Connecting Inputs and Outputs of Subsystems

Consider the situation where we have two subsystems described in state-space form.

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x,u) \end{aligned} \tag{11.9}$$

$$\begin{aligned} \dot{z} &= F(z,v) \\ w &= H(z,v) \end{aligned} \tag{11.10}$$

Now suppose we make a series connection so that the output y becomes the input to the other subsystem, that is $v = y$. Then we get

$$\begin{aligned} \dot{x} &= f(x,u) \\ \dot{z} &= F(z,h(x,u)) \\ y &= h(x,u) \\ w &= H(z,h(x,u)) \end{aligned} \tag{11.11}$$

We see that we still have a state-space description; the state variables of the combined system are x, z. The series connection is thus fairly straightforward. For linear subsystems this type of connection is well known.

Example 11.3 Series Connection of Two Linear Systems
Consider two linear subsystems described by

$$\begin{aligned} \dot{x} &= -x + u \\ y &= x \end{aligned} \tag{11.12}$$

$$\begin{aligned} \dot{z} &= v \\ w &= 3z - v \end{aligned} \tag{11.13}$$

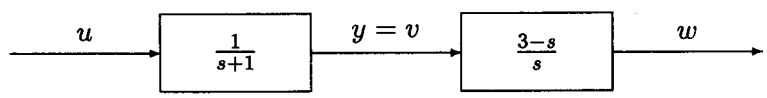
Suppose we make the series connection $v = y$. Then we have $v = x$ and the overall system becomes

$$\begin{aligned} \dot{x} &= -x + u \\ \dot{z} &= x \\ y &= x \\ w &= 3z - x \end{aligned} \tag{11.14}$$

Using (A.5) it is easy to see that the transfer functions of (11.12), (11.13) are

$$\frac{1}{s+1}, \quad \frac{3-s}{s}$$

respectively. We then get the following diagram.



The transfer function of (11.14) from u to w becomes

$$\frac{3-s}{s(s+1)}$$

which is the product of the transfer functions of the individual blocks. This is a well known fact from linear systems theory: a series connection corresponds to multiplication of the transfer functions. □

The situation becomes different if we decide to connect the subsystems in a *loop*.

Example 11.4 Two Linear Systems in a Loop (1)
Consider the same systems as in Example 11.3 but suppose that they are connected in a loop, that is $v = y$ and $u = w$. We then have the systems

$$\begin{aligned} \dot{x} &= -x + u \\ y &= x \end{aligned} \tag{11.15}$$

$$\begin{aligned} \dot{z} &= v \\ w &= 3z - v \end{aligned} \tag{11.16}$$

with the connecting equations

$$u = w = 3z - v, \quad v = y = x \tag{11.17}$$

Substituting from the second equation into the first gives

$$u = 3z - x, \quad v = x \tag{11.18}$$

The combined system is then

$$\begin{aligned} \dot{x} &= -2x + 3z \\ \dot{z} &= x \\ y &= x \\ w &= 3z - x \end{aligned} \tag{11.19}$$

□

This situation looks perfectly straightforward. However, let us change the example a little.

Example 11.5 Two Linear Systems in a Loop (2)
Suppose we change Example 11.4 by altering the output equation of the first subsystem so that we have

$$\begin{aligned} \dot{x} &= -x + u \\ y &= x + u \end{aligned} \tag{11.20}$$

$$\begin{aligned} \dot{z} &= v \\ w &= 3z - v \end{aligned} \tag{11.21}$$

Let the connection of subsystems be the same ($u = w$, $v = y$). The connecting equations are then

$$u = w = 3z - v, \quad v = y = x + u \tag{11.22}$$

This is in fact a linear system of equations in u and v:

$$\begin{aligned} u + v &= 3z \\ u - v &= -x \end{aligned} \tag{11.23}$$

The solution is

$$\begin{aligned} u &= -0.5x + 1.5z \\ v &= 0.5x + 1.5z \end{aligned} \tag{11.24}$$

The combined system is thus

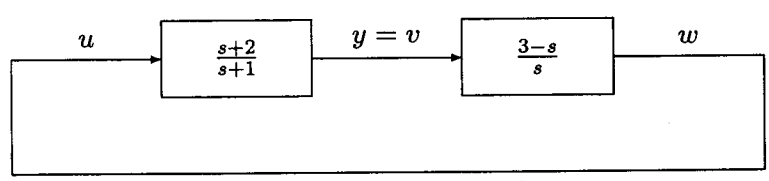
$$\begin{aligned} \dot{x} &= -1.5x + 1.5z \\ \dot{z} &= 0.5x + 1.5z \\ y &= 0.5x + 1.5z \\ w &= -0.5x + 1.5z \end{aligned} \tag{11.25}$$

□

Example 11.5 is more difficult to handle than Example 11.4 , because it involves the solution of a system of equations. This is because, for both subsystems, the output depends directly on the input (y depends directly on u and w depends directly on v). A system for which the output depends directly on the input is said to be a system with *direct feedthrough.* It is also useful to describe Example 11.5 in terms of transfer functions and block diagrams. Using (A.5), it is easy to see that the two subsystems have the transfer functions

$$\frac{s+2}{s+1}, \quad \frac{3-s}{s} \tag{11.26}$$

We can show that the transfer function of a linear system has the same degree in numerator and denominator precisely when there is a direct feedthrough from input to output. Thus we see clearly from the transfer functions that this is a case where both systems have direct feedthrough. The connected system is described by



The figure shows clearly why we get a system of equations. The signal u is equal to w, which depends directly on v, which is equal to y, which depends directly on u. The signals chase each other round the loop. This situation, in which systems with direct feedthrough are connected in a loop, is often referred to as an *algebraic loop.*

Let us now return to the general situation with two state-space systems (11.9) and (11.10). Let the systems be connected in a loop; that is, $u = w$ and $v = y$. Then we get the equations

$$\begin{aligned} v &= h(x, u) \\ u &= H(z, v) \end{aligned} \tag{11.27}$$

From these equations we have to solve for u and v in terms of x and z. If we are successful, we can substitute the solutions into (11.9) and

(11.10) and get a standard state-space description of the connected system. In the general case, (11.27) is a nonlinear system of equations and neither existence nor uniqueness of a solution is guaranteed. Even if a unique solution exists, it might be difficult to compute.

Suppose that one of the subsystems, for instance (11.9), has no direct feedthrough. Then $h(x, u)$ does not depend on u, so we have in fact

$$\begin{aligned} v &= h(x) \\ u &= H(z, v) \end{aligned}$$

which can be rewritten

$$\begin{aligned} v &= h(x) \\ u &= H(z, h(x)) \end{aligned}$$

The inputs can now be calculated from the state. When these expressions are substituted into (11.9), (11.10), we get a standard state-space description.

We can summarize our conclusions so far: If two dynamic subsystems that can be described by state-space equations are connected in a loop, there are two possibilities.

1. If at least one subsystem has no direct feedthrough, the resulting system has an easily computable state space description.

2. If both subsystems have direct feedthrough, we have an algebraic loop. The calculation of the resulting state space-description requires the solution of a system of equations. In the general case, these equations are nonlinear.

The discussion we have had so far generalizes in principle to many interconnected subsystems. If we have a number of interconnected subsystems, for which we can trace a loop, with direct feedthrough at every subsystem involved, then we have an algebraic loop. We then run into precisely the same kind of problems that we described for the case of two subsystems.

Connecting State Variables of Subsystems

In the previous section we studied the connection of inputs to outputs. The situation becomes more complex if subsystems are connected in

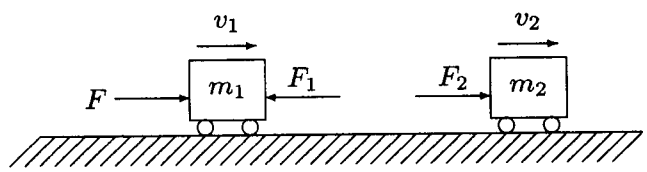


Figure 11.9: Mechanical system with two masses.

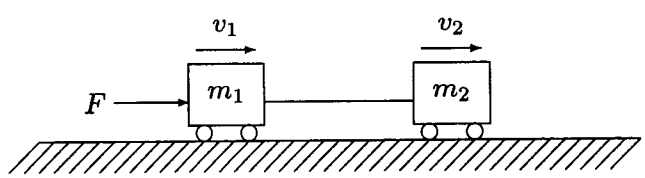


Figure 11.10: Mechanical system with connected masses.

such a way that certain relationships between the state variables are prescribed.

Example 11.6 Connecting Mechanical Subsystems
Consider the example of Figure 11.9. The left mass is described by

$$m_1\dot{v}_1 = F - F_1 \tag{11.28}$$

and the right mass by

$$m_2\dot{v}_2 = F_2 \tag{11.29}$$

Now suppose we connect the systems as described in Figure 11.10. We can regard this as a connection of the subsystems described by (11.28) and (11.29) through the equations

$$v_1 = v_2, \quad F_1 = F_2 \tag{11.30}$$

Note that this is a connection of the state variables v_1 *and* v_2. *By adding (11.28) and (11.29) and using (11.30), we get*

$$(m_1 + m_2)\dot{v}_1 = F \tag{11.31}$$

□

We see that a direct connection of state variables is something much more drastic than a connection of inputs and outputs. To get the resulting system we had to directly manipulate the state equations and eliminate some of the variables. In the process we reduced the system order from 2 to 1. Compare this example to the one of Figure 6.37, where we did a similar operation with bond graphs.

11.5 Simulation Languages

To make simulation possible, there must be a systematic way in which the user can enter a system description into the computer. We may call it a simulation language. Several approaches are possible. We list some of them.

1. The system is defined by subroutines written in a standard programming language like FORTRAN or C. Those routines define the rigth side of a state-space description like (11.8).

2. The right sides of (11.8) are entered in a specially designed command language.

3. The system description might be entered graphically in terms of block diagrams, as discussed in Section 11.3.

4. The system description is entered graphically in terms of bond graphs.

The first method is used in classical simulation languages like ACSL. The second method can be found for example, in SIMNON. The third method is found for instance in MATRIX-X and SIMULINK. Often a simulation language gives the user the choice between different methods. SIMNON, for instance, can handle methods 1 and 2, while SIMULINK gives the user a choice between C code, or m files, and block diagrams, that is, the three first methods.

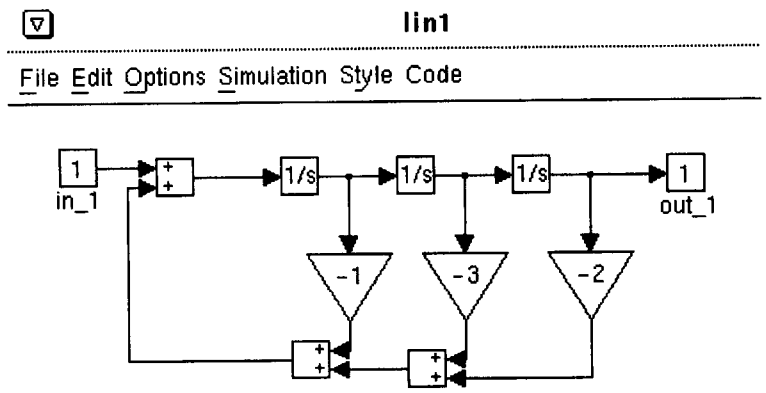


Figure 11.11: SIMULINK block diagram for linear system.

To give the reader some idea of what is possible in a modern simulation language, we show some examples of models in SIMULINK.

Consider the block diagram of Figure 11.4 in Section 11.3. In SIMULINK it can be represented as shown in Figure 11.11. We see that the block structure is copied exactly. Now we can group the blocks together and regard them as one block, as shown in Figure 11.12. The top block, `lin1`, is the system of Figure 11.11, while the other two blocks represent linear systems in transfer-function form and state-space form, respectively. The block in Figure 11.11 has the transfer function

$$\frac{1}{s^3 + s^2 + 3s + 2} \tag{11.32}$$

and a possible state-space description is

$$\dot{x} = Ax + Bu, \quad y = Cx + Du \tag{11.33}$$

with

$$A = \begin{pmatrix} -1 & -3 & -2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}, \quad D = 0$$

The descriptions (11.32) and (11.33) are also descriptions of the blocks `lin2` and `lin3`, so we have in Figure 11.12 three different blocks that

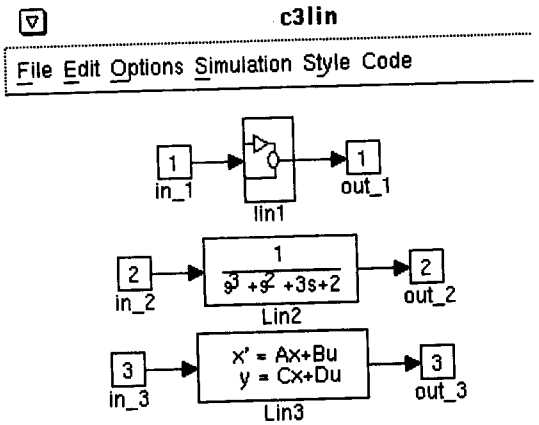


Figure 11.12: SIMULINK block diagram for three linear systems.

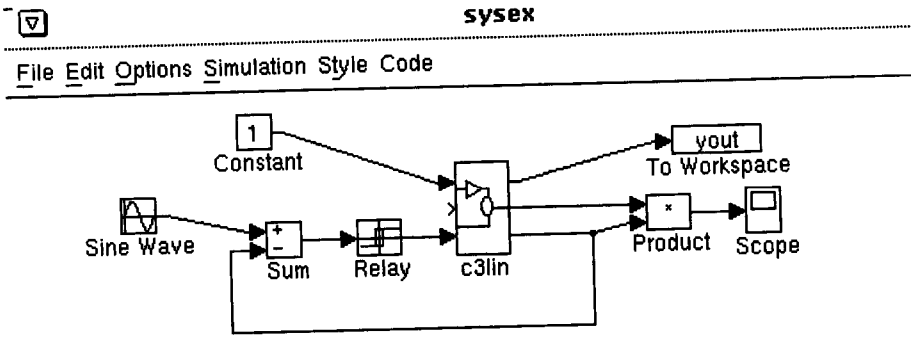


Figure 11.13: SIMULINK block diagram for nonlinear system.

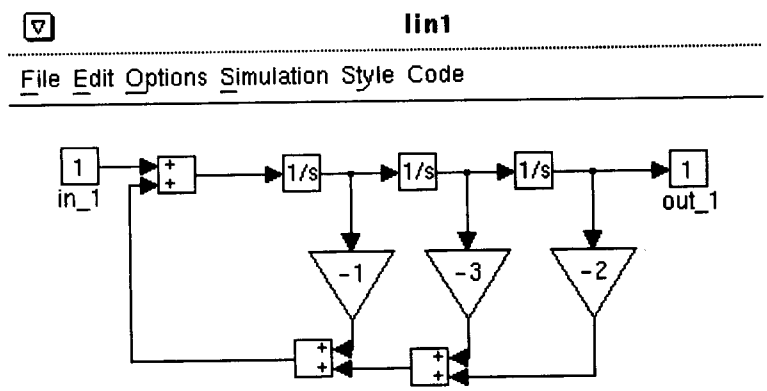


Figure 11.11: SIMULINK block diagram for linear system.

To give the reader some idea of what is possible in a modern simulation language, we show some examples of models in SIMULINK.

Consider the block diagram of Figure 11.4 in Section 11.3. In SIMULINK it can be represented as shown in Figure 11.11. We see that the block structure is copied exactly. Now we can group the blocks together and regard them as one block, as shown in Figure 11.12. The top block, `lin1`, is the system of Figure 11.11, while the other two blocks represent linear systems in transfer-function form and state-space form, respectively. The block in Figure 11.11 has the transfer function

$$\frac{1}{s^3 + s^2 + 3s + 2} \tag{11.32}$$

and a possible state-space description is

$$\dot{x} = Ax + Bu, \quad y = Cx + Du \tag{11.33}$$

with

$$A = \begin{pmatrix} -1 & -3 & -2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}, \quad D = 0$$

The descriptions (11.32) and (11.33) are also descriptions of the blocks `lin2` and `lin3`, so we have in Figure 11.12 three different blocks that

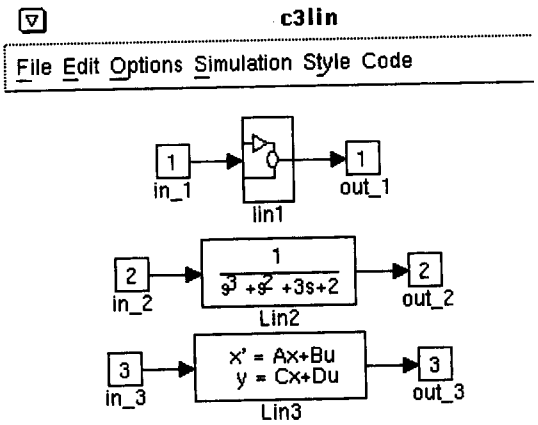


Figure 11.12: SIMULINK block diagram for three linear systems.

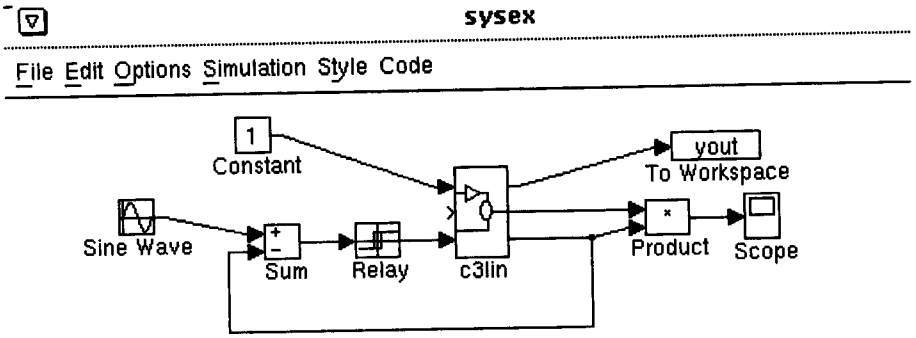


Figure 11.13: SIMULINK block diagram for nonlinear system.

describe the same linear system. In Figure 11.13 we have grouped the three systems together into one block with three inputs and three outputs. In order to show some other features of the language, we have connected some additional elements. These include two nonlinearities in the form of a relay with hysteresis and a multiplication of signals. Furthermore we have demonstrated two examples of how to generate input signals, one sine wave and one constant. We also see examples of the display of simulation results. The first is a scope, which gives on-line display on the computer screen, where as the second is an output to the program workspace. This output can then be used for later display or for processing of the data.

Several notable features are shown in this example:

- The hierarchial structure. The system **sysex** contains the block `c3lin`, which in turn contains several blocks, among them `lin1`, which in turn is made up of a number of blocks.
- The possibility of describing a given system in several ways, as was done in `c3lin`.
- Various possibilities of generating input signals.
- Various ways of storing and displaying results.

Many additional features are not displayed in this simple example. Some instances are the following:

- A block can represent a discrete time model. Discrete time and continuous time blocks can be connected.
- From a block model of a linear system it is possible to compute a state-space description and from that a transfer function. For a nonlinear system a linearized state-space model can be computed.
- It is possible to include blocks that represent a continuous time delay.

Current Trends

Many of the current developments of simulation languages can be described as making them more and more into modeling languages. In

our brief SIMULINK example we saw that it was possible to enter a model using one description and have it recomputed into another description. This means that the program is actually performing part of what we called phase 3 in our discussion of modeling principles in Chapter 4. Ideally a modeling and simulation language would automate all of phase 3 in the modeling process. The user would then have to supply the basic physical relationships of phase 2, but would not have to worry about putting them together into a model that can be simulated.

Another ongoing development concerns algebraic-differential equations,

$$\begin{aligned} \dot{z} &= f(z,u) \\ 0 &= g(z,u) \end{aligned} \tag{11.34}$$

that is, a mixture of differential and algebraic equations. There are several approaches to these. We could try to use the second set of equations to eliminate some variables from the first set and arrive at a standard state-space description. Alternatively, we can develop numerical methods that can handle algebraic-differential equations directly. Several simulation languages can handle equations arising from algebraic loops by solving them numerically on line. There are difficulties with this, of course, since there are no global numeric methods for nonlinear equations that are guaranteed to work in all cases. An alternative that is attractive in principle is to use the computer algebra methods we looked at in Chapter 7 to manipulate the equations. In Example 7.4 we actually saw how this can be done. There we started with a second-order differential equation and a static relationship and arrived at a state-space description. The methods used in the example generalize to arbitrary polynomial systems of differential equations, but the complexitiy is such that they cannot be considered the final answer to the problem.

The devopments in algebraic-differential equations and in modeling are actually tied together. We saw in Chapter 4 that phase 2 of the modeling resulted in a mixture of static and dynamic relationships, that is, a system of equations of type (11.34). Also we saw in Section 11.4 that the tying together of subsystems could result in algebraic loops. To resolve the algebraic loop, a system of equations, (11.27), has to be solved. We also saw that a modeling approach that is

sufficiently general to allow arbitrary connections between subsystems must be able to handle mixtures of differential and algebraic equations. These equations can be tricky in the sense that the static relationships actually reduce the order of the system. A modeling language that has advanced features for handling these types of problems is DYMOLA. It can handle models defined in many ways, including bond graphs. Subsystems can be tied together in general ways and algebraic relationships handled. The output of DYMOLA can be a model suitable for one of the standard simulation languages.

11.6 Numeric Methods

We will now look at the actual solving of the differential equations defining our model. We will assume that, no matter how the model was entered by the user, the simulation language/program has been able to convert it to a standard state-space form. In other words we will not consider numeric methods that handle algebraic-differential equations. We are going to present some of the most common algorithms and discuss the properties that are essential to consider in a simulation. For a more detailed discussion we refer to courses and textbooks in numerical analysis. (See the bibliography for Part IV.)

Basis of Numerical Methods

We consider a model in state-space form

$$\dot{x}(t) = f(x(t), u(t))$$

where x is a vector of dimension n. If we fix the input to a certain function $u(t) = \bar{u}(t)$, we can represent the influence of u as a time variation of f and write

$$\begin{aligned} \dot{x}(t) &= f(t, x(t)) \\ x(0) &= x_0 \end{aligned} \tag{11.35}$$

Assume that we have an initial state $x(0)$ and want an approximation of x at the points

$$0 < t_1 < t_2 < \cdots < t_f$$

Our algorithm will thus generate the values x_1, x_2, $x_3, \ldots$, which approximate

$$x(t_1), x(t_2), x(t_3), \ldots$$

The simplest algorithm is obtained by approximating $\dot{x}(t)$ with a difference ratio

$$\frac{x_{n+1} - x_n}{h} \approx \dot{x}(t_n) = f(t_n, x_n), \text{ where } h = t_{n+1} - t_n$$

This gives the equation

$$x_{n+1} = x_n + hf(t_n, x_n) \tag{11.36}$$

which is called *Euler's method.* It is often used in simple simulations, but as we will see below, is it not the most effective method.

An algorithm for differential equations can, more generally, be described by the equation

$$x_{n+1} = G(t, x_{n-k+1}, x_{n-k+2}, \ldots, x_n, x_{n+1}) \tag{11.37}$$

The integer k shows the number of previous steps that are utilized. We can thus speak of a (11.37) as a *k-step method.*

If x_{n+1} is not included in the expression for G, x_{n+1} is obtained directly by evaluating the right side. The method is then said to be *explicit.* In other cases the method is *implicit*, and an equation system has to be solved to get x_{n+1}. Euler's method is obviously an explicit one-step method.

It is of course important to know how accurate a method is in solving differential equations. It is natural to consider the *global error*

$$E_n = x(t_n) - x_n$$

which, however, turns out to be difficult to compute. Therefore, we also look at the one-step error, or the *local error*

$$e_n = x(t_n) - z_n,$$

where z_n satisfies

$$z_n = G(t, x(t_{n-k}), x(t_{n-k+1}), \ldots, z_n)$$

This is the error obtained when using the method for one step, provided all information from previous steps is exact. For Euler's method we have

$$\begin{aligned} e_{n+1} &= x(t_{n+1}) - x_{n+1} = x(t_{n+1}) - x(t_n) - hf(t_n, x(t_n)) \\ &= \frac{1}{2}h^2\ddot{x}(\xi), \quad \text{for} \quad t_n < \xi < t_{n+1} \end{aligned} \tag{11.38}$$

(In this case we have, for simplicity, looked at the scalar case.)

The local error is thus proportional to h^2. It can be shown that the global error will be proportional to h. (Intuitively, we can reason as follows: the number of steps will be proportional to h^{-1} so that the local error is multiplied by h^{-1}.) If the local error is of the form $\mathcal{O}(h^{k+1})$, k is usually called the *order of accuracy.* [Here we read $\mathcal{O}$ as "large ordo." It denotes a function that is of the same order of magnitude as its argument when the argument tends to zero. $\mathcal{O}(x)/x$ is thus limited when x tends to zero.] The global error is then $\mathcal{O}(h^k)$.

However, accuracy is not the only aspect that is important. In many simulations, stability is an essential qualitative property, which is studied. It can therefore be valuable to investigate the relationship between the stability of the numerical method and that of the underlying differential equation. To do this, we consider the simple scalar test equation

$$\begin{aligned} \dot{x} &= \lambda x, \qquad \lambda \text{ complex number} \\ x(0) &= 1 \end{aligned} \tag{11.39}$$

(We can envision that we start with a system of differential equations

$$\dot{z} = Az$$

and diagonalize A. Each component of the equation will then be of the preceding form. We note especially that it is natural to allow complex values of λ.)

If we use Euler's method we have

$$x_{n+1} = x_n + h\lambda x_n = (1 + h\lambda)x_n$$

The solution to this difference equation is

$$x_n = (1 + h\lambda)^n \tag{11.40}$$

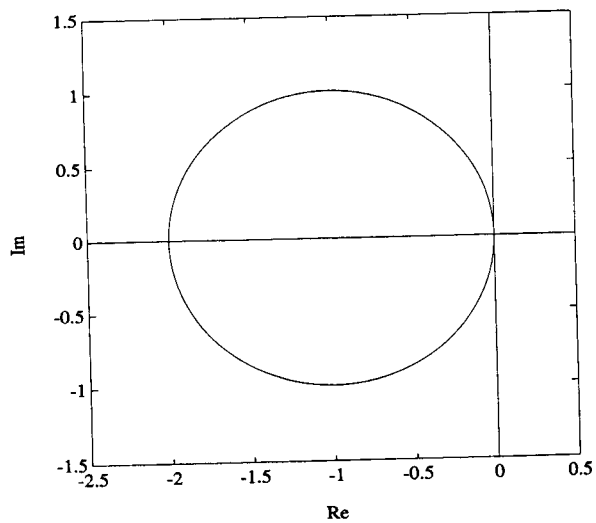


Figure 11.14: Values of λh that give a stable difference equation for Euler's method (stability inside the circle).

We see that

$$x_n \to 0 \text{ if } |1 + h\lambda| < 1$$

and

$$|x_n| \to \infty \text{ if } |1 + h\lambda| > 1$$

The stability area for the difference equation is thus a disk with radius 1 and center at -1 (see Figure 11.14). On the other hand, $x \to 0$ for the differential equation if $\text{Re}[\lambda] < 0$ and $|x| \to \infty$ if $\text{Re}[\lambda] > 0$. The fact that the region of stability of a differential equation does not necessarily coincide with the one valid for the numerical algorithm is something to remember when trying to determine the stability of a dynamic system by simulation. In this case we see, however, that for $\text{Re}[\lambda] < 0$, $h\lambda$ will be in the region of stability if h is small enough. If $x \to 0$, then also $x_n \to 0$ if the step-length is small enough.

As we mentioned earlier, there are more effective methods than Euler's method. We will describe some of them here.

The Runge-Kutta Methods

Again consider (11.35), which can also be written in integral form:

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} f(\tau, x(\tau))d\tau$$

If we approximate the value of the integral with the length of the interval times the value of the integrand in the center, we have

$$x_{n+1} = x_n + h \cdot f(t_n + \frac{h}{2}, x(t_n + \frac{h}{2})) \tag{11.41}$$

The problem is that $x(t_n + \frac{h}{2})$ is unknown. It can, however, be approximated with Euler's method:

$$x(t_n + \frac{h}{2}) \approx x_n + \frac{h}{2} f(t_n, x_n) \tag{11.42}$$

If we combine equations (11.41) and (11.42), we can write

$$\begin{aligned} k_1 &= f(t_n, x_n) \\ k_2 &= f(t_n + \tfrac{h}{2}, x_n + h\tfrac{k_1}{2}) \\ x_{n+1} &= x_n + hk_2 \end{aligned}$$

This is a simple example of an algorithm of the Runge-Kutta type. If we make a series expansion, we see, after some computations, that the local error is

$$x(t_{n+1}) - x_{n+1} = \mathcal{O}(h^3)$$

That is, the method is an order of magnitude more accurate than Euler's method.

In general, a Runge-Kutta method is described by the equations

$$\begin{aligned} k_1 &= f(t_n, x_n) \\ k_2 &= f(t_n + c_2 h, x_n + h a_{21} k_1) \\ k_3 &= f(t_n + c_3 h, x_n + h(a_{31} k_1 + a_{32} k_2)) \\ &\vdots \\ k_s &= f(t_n + c_s h, x_n + h(a_{s1} k_1 + \cdots + a_{s,s-1} k_{s-1})) \\ x_{n+1} &= x_n + h(b_1 k_1 + \cdots + b_s k_s) \end{aligned}$$

The number s and the coefficients c_i, b_i, and a_{ij} are chosen so that the method has the desired order of accuracy p, that is, such that

$$x(t_{n+1}) - x_{n+1} = \mathcal{O}(h^{p+1})$$

Depending on whether we want to make the calculation simple, to minimize the error term or some other criterion, we will get different values of the coefficients. There is thus a family of Runge-Kutta methods.

A classic method sets $s = p = 4$ and the coefficients are

$$c_2 = c_3 = \frac{1}{2}, \quad c_4 = 1, \quad a_{21} = \frac{1}{2}, \quad a_{32} = \frac{1}{2}, \quad a_{43} = 1,$$

$$b_1 = b_4 = \frac{1}{6}, \quad b_2 = b_3 = \frac{2}{6}$$

(The nonlisted coefficients have the value 0.)

Adams's Methods

The Adams's methods form a family of multistep methods that can be written in the form

$$x_n = x_{n-1} + \sum_{j=0}^{k} \beta_j f_{n-j}; \quad f_i = f(t_i, x_i)$$

The coefficients β_j are chosen so that the order of accuracy is as high as possible. If $\beta_0 = 0$ is chosen, the method will be explicit. This variant is often called *Adams-Bashforth,* while the implicit form ($\beta_0 \neq 0$) is called *Adams-Moulton.* The simplest explicit forms are

$$\begin{aligned}
k = 1: &\ x_n = x_{n-1} + f_{n-1}\, h \\
k = 2: &\ x_n = x_{n-1} + (3f_{n-1} - f_{n-2})\, \tfrac{h}{2} \\
k = 3: &\ x_n = x_{n-1} + (23f_{n-1} - 16f_{n-2} + 5f_{n-3})\, \tfrac{h}{12} \\
k = 4: &\ x_n = x_{n-1} + (55f_{n-1} - 59f_{n-2} + 37f_{n-3} - 9f_{n-4})\, \tfrac{h}{24}
\end{aligned}$$

while the implicit ones are given by

$$\begin{aligned}
k = 0: &\ x_n = x_{n-1} + f_n\, h \\
k = 1: &\ x_n = x_{n-1} + (f_n + f_{n-1})\, \tfrac{h}{2} \\
k = 2: &\ x_n = x_{n-1} + (5f_n + 8f_{n-1} - f_{n-2})\, \tfrac{h}{12} \\
k = 3: &\ x_n = x_{n-1} + (9f_n + 19f_{n-1} - 5f_{n-2} + f_{n-3})\, \tfrac{h}{24}
\end{aligned}$$

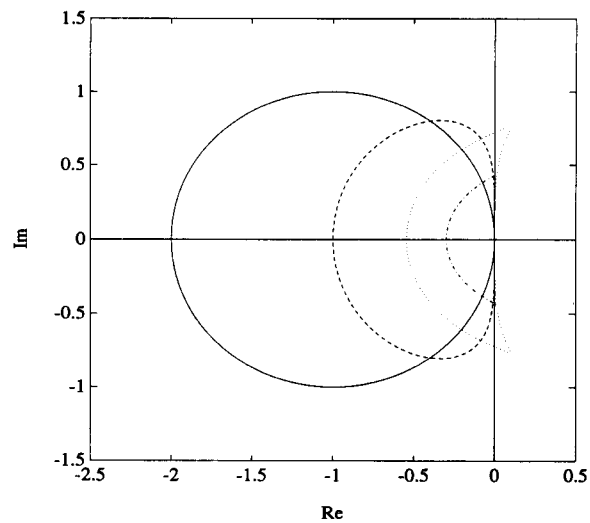


Figure 11.15: Stability regions for Adams-Bashforth for different k; $k = 1$: solid line, $k = 2$: dashed line, $k = 3$: dotted line, $k = 4$: dash-dotted line (stable inside the contours).

It can also be shown that for the explicit methods the order of accuracy is equal to k. For the implicit methods it is of order $k + 1$.

We see that for the implicit variant we have to solve the equation

$$x_n - \beta_0 f(t_n, x_n) = x_{n-1} + \sum_{j=1}^{k} \beta_j f_{n-j}$$

with respect to x_n. This can be done using an iterative procedure, where a point is generated from the explicit Adams-Bashforth method as an initial guess. What is the reason for using the more complicated implicit equations? One reason is shown in Figures 11.15 and 11.16. The implicit methods have, as we see, considerably larger stability areas than the explicit methods with the same order of accuracy. We also see that an increased order of accuracy might result in a smaller stability region.

Variable Step Length

It is often inefficient to use a constant step length when solving differential equations. A typical solution might contain segments with

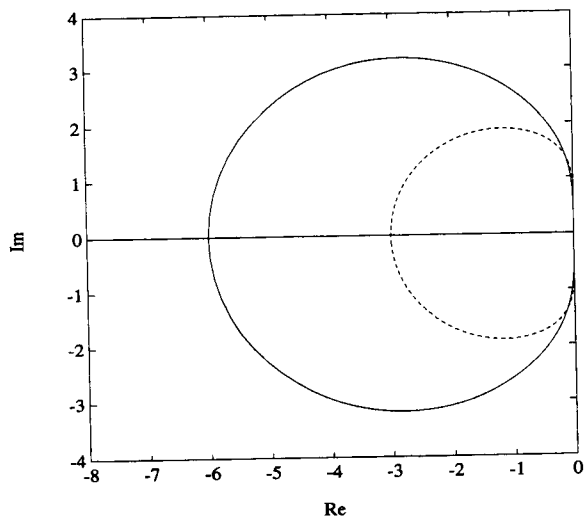


Figure 11.16: Area of stability for Adams-Moulton for different k; $k = 2$: solid line, $k = 3$: dashed line (stable inside contours).

rapid changes, where small steps are required, and segments where slow changes make large steps possible. Methods for automatic step adjustment are often based on an estimate of the local error. This can be done in the following way, for example. Assume that we have an algorithm in which the local error has the form

$$x(t_{n+1}) - x_{n+1} = Ch^{p+1} + \mathcal{O}(h^{p+2})$$

The coefficient C depends on the solution and is therefore normally unknown. If the algorithm takes two steps of length h we have

$$x(t_{n+2}) - x_{n+2} = 2Ch^{p+1} + \mathcal{O}(h^{p+2}) \tag{11.43}$$

since the errors in two steps are added (approximately). Let $\tilde{x}$ denote the value computed by the algorithm when taking a step of length $2h$ from t_n to t_{n+2}. Then we get

$$x(t_{n+2}) - \tilde{x} = C(2h)^{p+1} + \mathcal{O}(h^{p+2}) \tag{11.44}$$

By subtracting (11.43) from (11.44), we get

$$x_{n+2} - \tilde{x} = 2Ch^{p+1}(2^p - 1) + \mathcal{O}(h^{p+2})$$

If this is used to eliminate C in (11.43), the result will be

$$x(t_{n+2}) - x_{n+2} = \frac{x_{n+2} - \tilde{x}}{2^p - 1} + \mathcal{O}(h^{p+2})$$

If we assume that h is small enough for the $\mathcal{O}(h^{p+2})$ term to be negligible, then the right side only contains known numbers, and we thus have an estimate of the error.

This estimate can be used in different ways, but the general idea is the following. The algorithm has a tolerance level for the local error (for example, determined by the user). If the error estimate exceeds the tolerance, the step length h is decreased. If it is far below the tolerance, the step length is increased. Ideally, a given accuracy is obtained with a minimum amount of computational work.

Stiff Differential Equations

Stiff differential equations are characterized by the fact that their solutions contain both fast and slow components and that the difference between their time constants is large. An example is the differential equation

$$\dot{x} = \begin{pmatrix} -10001 & -10000 \\ 1 & 0 \end{pmatrix} x$$
$$x(0) = \begin{pmatrix} 2 \\ -1.0001 \end{pmatrix}$$

which has the solution

$$x_1 = e^{-t} + e^{-10000t}$$
$$x_2 = -e^{-t} - 0.0001e^{-10000t}$$

If we want to simulate this equation, we have to start with a very short step length in order to follow the fast term $e^{-10000t}$. This term is, however, soon close to zero, and the solution is only characterized by the term e^{-t}. Since it is much slower, we would now like to increase the step length. But then we have to take stability into account. The product $-10,000 \cdot h$ has to be within the stability region, which means that h has to be small if the stability region looks like Figures 11.14,

11.15, and 11.16. One way to avoid the stability problem would be to use methods that are always stable if $\mathrm{Re}[\lambda h] < 0$. Then it is possible to use an unlimited step length without instabilities if the differential equation itself is stable. Unfortunately, it is difficult to combine such a stability region with the demand for high accuracy. Methods for stiff differential equations often make a compromise such that almost the whole left half-plane is included in the stability region, while at the same time the accuracy order will be reasonably high. We saw earlier that implicit methods often have a wider stability area than the explicit ones. Methods for stiff differential equations are therefore implicit in general, and an equation system thus has to be solved at each step.

Comments about Choice of Methods

Several investigations of the efficiency of numerical algorithms for differential equations have been done. They indicate that Runge-Kutta methods are most effective when the complexity is relatively low, while Adams's methods are preferable at high complexities. By complexity, we mean the computational work in evaluating the right side of the differential equation. It also seems that methods for stiff problems are usually ineffective for nonstiff problems. These methods should thus be reserved for the problems where they really are needed.

11.7 Simulators

So far we have viewed the simulation only as solving differential equations. The solution can then be presented in different ways, for example, as curves on paper or on a screen. The presentation can, however, be taken much further. If, for example, an industrial robot is simulated, a three-dimensional picture can be shown on the screen showing the movements of the robot in three dimensions. In a car simulator a test person is placed in the driver seat, like the one in a regular car. The maneuvering tools like the steering wheel, accelerator, gear shift, and so on are tied to a sensor whose values are transferred to a computer. Mathematical models of a car are programmed into the computer, and this model is simulated in real time. With the guidance of the simulation results, the speedometer and other instruments are

controlled so that they show the same values as in a real car under corresponding conditions. The windshield is replaced by some kind of display, which shows how the road looks as a result of different maneuvers. In more advanced simulators, the whole driver's seat is put in motion so that the test person experiences the centrifugal force in curves, and so on.

If the simulation is combined with enough powerful presentation tools, we talk of a *simulator*. Examples of this are airplane simulators, car simulators, nuclear power plant simulators, process control simulators and so on.

Different Purposes of Simulators

There are a number of different uses for simulators.

Means of Assistance for the Designer

By simulating the dynamic course of events, a designer can find out what demands are put on different components. Many industrial processes have traditionally been designed from the demands valid for operation in a stationary state. Experiences show, however, that it is necessary to account also for transient phenomena. It can therefore become necessary to make extensive simulations to test all the different dynamic behaviors that can occur, both in normal operation and under exceptional conditions.

Operating Procedures

Simulations can also form the basis for certain operating procedures. As an example, different ways of starting an industrial process after a stop can be simulated. It is then possible to see the influence of different sequences and actions. The simulations can then be used when instructions for operators are drawn up.

Instruction

Simulators can be valuable for instruction. People learning how an industrial process works can vary different physical quantities and observe how the process behaves. Aside from simulations often being

cheaper, there are a number of advantages working with a simulation instead of the real process. All internal variables (which are represented in a simulation) can be illustrated. Some of these can be hard or impossible to measure in the physical process. It can also be difficult to illustrate certain events in reality, because the process then has to work under conditions that are uneconomical or dangerous.

Training

In training the emphasis is shifted from giving understanding of principles to showing, often in detail, what really happens. The classical example is training in airplane simulators. Here we want to show how the airplane behaves as an effect of different rudder commands, but also what all the instruments will show. Another example is training simulators for nuclear power plants or complicated processes in steel factories, refineries, and the paper industry. We then want to show how all instruments, screens, monitors, and so on, behave during different situations. An advantage of simulators is that it is possible to train for an operator's handling of abnormal situations, pure emergencies, and accidents. These situations are of course difficult to train for otherwise. Another advantage is that the training of process operators can start before the process is completed. The start-up procedure will then often be smoother since the staff already has a certain amount of experience.

Consequences of Decisions

All the earlier applications imply off-line simulation. If it is possible to simulate faster than the natural time scale, then the simulation can be used to help the operator on-line when controlling the real process. The operator can then "ask" the simulator, "What happens if I open this valve?" "How long will it take before the temperature rises if the fan is turned off?" and so on. Decisions about suitable actions can be taken with the guidance of the simulations.

Types of Simulators

Since simulation is used for so many different purposes, different types of simulators with different demands on hardware and software are

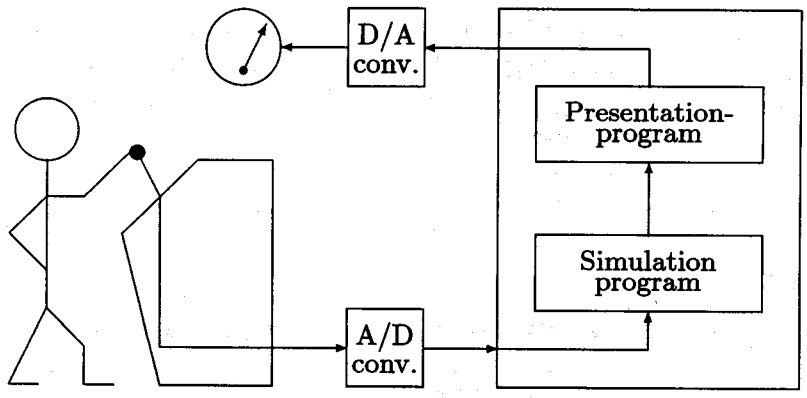


Figure 11.17: Basic structure of a simulator.

available.

Full-scale Simulators

In full-scale simulators we try, as far as possible, to imitate the real system seen from the operator's point of view. An airplane simulator, for example, is often an exact copy of the cockpit, with the same controls and instruments as in the real airplane. Pictures can, for example, be presented on a screen, which depict what the pilot sees. The cockpit can be turned and moved in different directions in order to simulate accelerations the pilot is exposed to in reality.

A complete control room, with the same type of commercial instruments and regulators as in a real process, can sometimes be built in industrial full-scale simulators. The construction shown in Figure 11.17 is then in principle obtained.

Completely computerized control systems are now used in many industries. For such systems a simulator can be obtained by running a simulation program through the system. A computer-based control system, used in a heating plant, is shown in Figure 11.18. The simulator is obtained by running the system in the configuration shown in Figure 11.19. The real processes are replaced by simulation programs, but the part of the system the operator sees at the console is the same.

From what we have said it is evident that a full-scale simulator is used mainly in the training of operators and similar tasks.

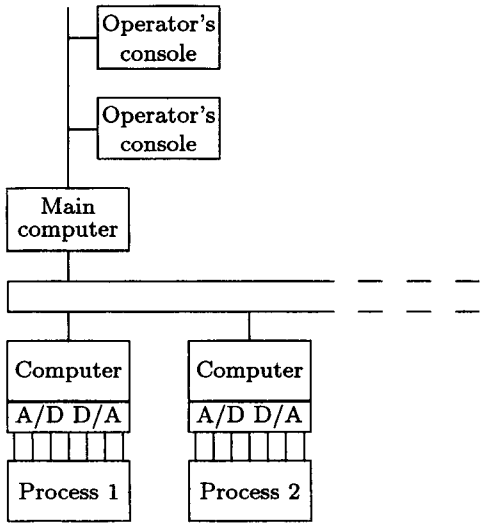


Figure 11.18: Basic drawing of a certain computerized control system.

Function and Compact Simulators

If the main interest with a simulation is to illustrate a principle, for example, for instruction purposes, it is often unnecessary to build a full-scale model. Simpler instrumentation or the computer's built-in units, for example the screens, can be satisfactory. We may, on the other hand, want to present variables and relationships in the simulation that are invisible in the real process (and in general also in the full-scale simulator).

Hardware and Software Demands

The demand on full-scale simulators that everything the operator sees and comes in contact with has to look like real life puts high demands on the hardware. In cases where advanced visual presentation is used, advanced software and perhaps special processors are also needed.

Simulation for construction support in general sets no such demands. Instead, high accuracy in the simulation model is often needed, which can lead to demands of large computational capacity in the computer.

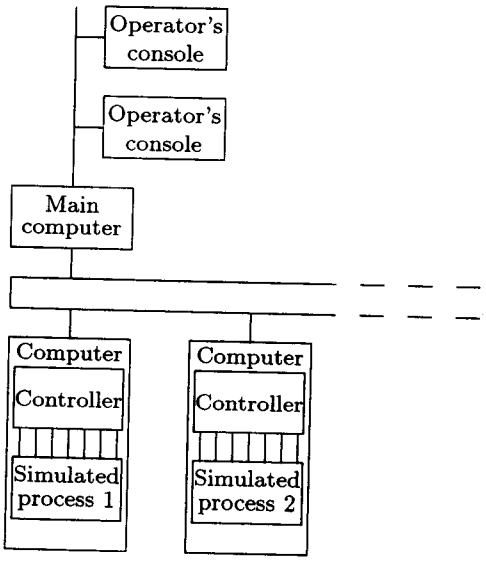


Figure 11.19: Simulation of process control.

When simulating for instructional purposes, we can in certain cases manage with lower accuracy in the model if the main interest is in showing the principal behavior of a process.

The demand for accuracy in the simulation part of a full-scale simulator is mainly that the operator has to "feel at home." This can lead to high accuracy demands on certain variables, while for others it may be enough that the qualitative behavior be correct.

11.8 Summary

In summary we see that modern computer and programming tools turn simulation into a relatively straightforward procedure compared to the modeling work. Difficulties can arise with models that are not in state-space form and that have very stiff systems. Complicated models can demand very large computer resources. In some cases the problem may need to be attended to on the modeling level, for example, by model reduction.

Chapter 12

Model Validation and Model Use

A model is not useful until its validity has been tested and established. Using simulation results from an untested model complacently can be worse than to guess on the basis of common sense. The formal "scientific method" can give a false impression of authority. In this chapter we are going to discuss how to achieve confidence in models and how to remain soundly critical of them.

12.1 Model Validation

A model is never the true description of a system. We thus do not demand that a model be "exact," "true," or "correct." Models have instead been developed to help in solving certain problems. We call a model that is useful in this way *valid* with regard to the purpose in mind.

We can as an example mention Ptolemy's model of the solar system from about 150 A.D. This model places the earth in the center and the sun, the moon, and the planets in complicated systems of circles and epicycles. See Figure 12.1. The purpose of the model was to be able to compute future planetary movements and solar eclipses. This could be done with impressive accuracy. Ptolemy's model of the solar system is thus valid with regard to its purpose, even though we view it as "incorrect."

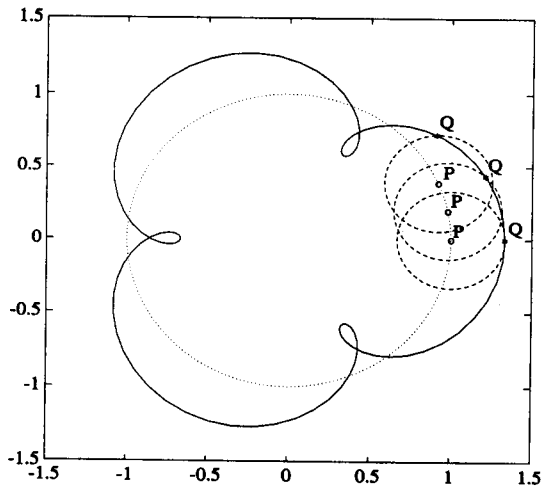


Figure 12.1: Ptolemy's model of the planetary movements. The planet Q moves around P in a circle (dashed curve), while P moves around the dotted circle. This results in a movement according to the solid curve (an epicycle).

Deciding if a certain model is valid is called *model validation.* The procedure implies that the steps in the model building have to be critically scrutinized and that some parts of the model have to be dismissed or improved. Model validation is thus intimately tied to modeling. See Figure 12.2. Let us discuss the lower block in this figure. How is the validity of a certain model tested? Since the validity refers to a certain purpose for the model, the test is problem dependent. A common characteristic is, however, that the output from the model and system are compared when they are run with the same input. See Figure 12.3. The difference has to be small, and what is meant by "small" depends on both the purpose of the model and the disturbances that influence the output. We discussed this in Section 10.5.

When the model has been compiled into a total system description, we also have the possibility of evaluating the influence of different model approximations and model parameters. The level of approximation can then be changed in a subsystem, and we can test how much the output of the total model changes. We should thus work more with

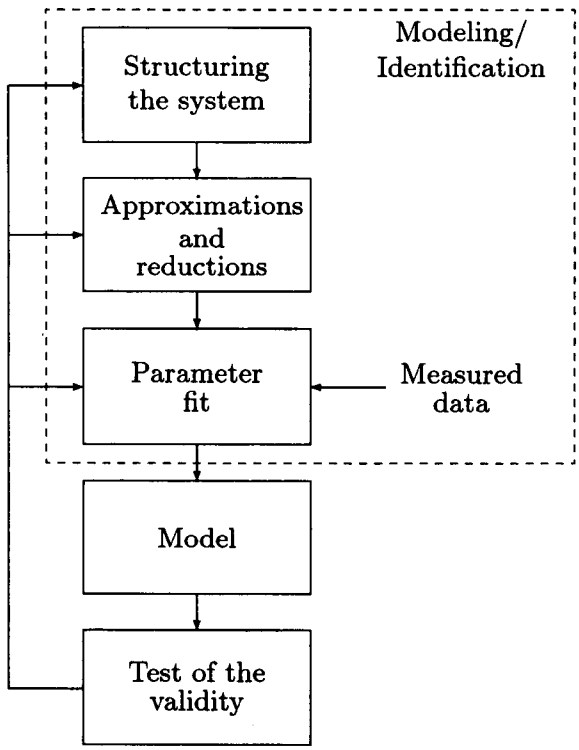


Figure 12.2: Model validation procedure.

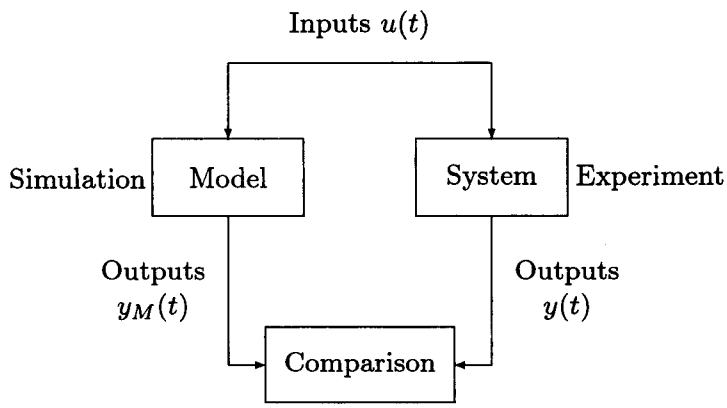


Figure 12.3: Test of the validity.

the subsystem in question if the change is significant. Also, if a certain parameter value is found to seriously influence the model properties, further attempts have to be made to estimate the value carefully. This iterative procedure is described symbolically in Figure 12.2.

12.2 Domain of Validity of the Model

All models have a limited domain of validity. They might relate to the system's properties at a certain operating point or have certain accuracy limitations. It is always risky to use a model outside the area for which it has been validated. A model can, on the other hand, be validated only within the area that the system itself is allowed to work in. A model of a nuclear reactor cannot be validated for use during catastrophe conditions. The purpose of the model in such cases is often precisely to use it to test dangerous situations.

We are thus sometimes forced to use a model outside its established area of validity. This places high demands on the model so that we can have intuitive confidence in it. We can thus speak of the model's *credibility* along with its *validity*.

Building models that are not only valid but also credible is a problem that lies close to the core of the philosophy of science, and we will

here refrain from further discussion.

We can, however, point to the solar system. Ptolemy's as well as Kopernicus-Kepler-Newton's model have a high degree of validity. The latter is however more credible, since it has a broad area of validity (from apples to planets), it contains few numeric parameters (the gravitational constant, the masses of the planets), compared with Ptolemy's many circular radii for each planet, and it has not been necessary to adjust it to be able to describe the movements of new planets (it has shown itself to be valid outside earlier established areas of validity).

12.3 Remaining Critical of the Model

Remember that a model is never true or correct. At its best it is valid and possibly credible. This means that we always must remain critical of the model.

In particular one should think of the following fallacies:

- *The Pygmalion effect:*[1] "Don't fall in love with your model!". Even if a considerable amount of work has been done to develop the model, we have to remember that it is the system that is most important — not the model.

- *The Procrustes effect:*[2]. Forcing reality to fit the model is not possible (although sometimes tempting). We must always be ready to develop and modify a model to include new facts and observations. We cannot disregard phenomena that conflict with the model. Many important scientific discoveries have their basis in facts that conflict with accepted models.

- *Be aware of the model's (lack of) accuracy!* It is necessary to keep in mind the degree of accuracy in the model and the level

[1] *Pygmalion:* fairy tale king of Cyprus, famous sculptor, who fell in love with one of his works, a sculpture of a young maid, and asked the Gods to make it alive, which then happened.

[2] *Procrustes:* in Greek mythology a robber at Eleusis, known for the bed (Procruste's bed), where he tormented the travelers who fell into his hands; if the victim was too short, he stretched its limbs until it fit the bed, if the victim was to tall, he cut off its legs and head.

of approximation when the simulation results are interpreted. This is particularly valid when the model contains estimated parameters. Forrester's world model has, for example, been criticized because it has completely different qualitative properties when certain parameters vary within a reasonable interval (see Example 4.3). It is obviously hard to base decisions on such a model.

12.4 Use of Several Models

Since a model has a limited domain of validity, it is interesting to work with several models for the same system. There can be a sequence of models for different operating conditions, such as airplane models for different flight conditions (altitude and speed). It is also possible to have one model for fast responses and one for slow responses.

Another possibility is to work with a hierarchy of models with different levels of accuracy and complexity to solve different problems. Finally, it is important to have a flexible attitude toward the model concept and its possibilities.

Bibliography for Part IV

Numerical methods for differential equations are described in many books. We mention

E. Hairer, S P. Nørsett, and G. Wanner: *Solving Ordinary Differential Equations.* Springer-Verlag, New York, 1987.

C. W. Gear: *Numerical Initial Value Problems in Ordinary Differential Equations.* Prentice Hall, Englewood Cliffs, N.J., 1971.

Both modeling and simulation aspects are described in the journal

Simulation. Technical journal of the Society for Computer Simulation, La Jolla, California.

We also refer to the article "Simulation" in

Encyclopedia of Science and Technology. Academic Press, New York, 1987, volume 12, pp. 659-698.

Details of practical simulation work are described in manuals for the different simulation languages. See, for example

Simulink, User's Guide. The MathWorks, Inc., Natick, Mass., 1990.

Matrix X Manuals, Integrated Systems, Inc., Santa Clara, Calif.

SIMNON, User's Guide, SSPA Systems, Göteborg, Sweden.

DYMOLA Manuals, DynaSim AB, Lund, Sweden.

Appendix A

Linear Systems. Description and Properties

A.1 Time Continuous Systems

A linear system in state-space form is described by

$$\begin{aligned} \tfrac{d}{dt}x &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{A.1}$$

If x is an n vector, u an m vector and y a p vector, A, B, C, and D will be matrices of dimensions $n \times n$, $n \times m$, $p \times n$ and $p \times m$, respectively. x is called the state, u the input, and y the output. n is called the the system order. Often $D = 0$.

If u, y are scalars ($m = p = 1$), B is then a column vector and C a row vector.

The solution to (A.1) is given by

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \tag{A.2}$$

where the matrix exponential is defined according to

$$e^{At} = I + t \cdot A + \frac{t^2}{2}A^2 + \cdots + \frac{t^k}{k!}A^k + \cdots \tag{A.3}$$

If u and y have the Laplace transforms $U(s)$ and $Y(s)$ respectively, these will be related by

$$Y(s) = G(s)U(s) \tag{A.4}$$

where G is a $p \times m$ matrix called the transfer function. (Here the initial conditions are supposed to be zero.) The system (A.1) corresponds to a transfer function $G(s)$

$$G(s) = C(sI - A)^{-1}B + D \tag{A.5}$$

If u and y are scalars ($p = m = 1$), $G(s)$ is a rational function:

$$G(s) = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n} \tag{A.6}$$

The values of s for which $G(s) = 0$ are called *zeros*, while values where $G(s) = \infty$ are called *poles* (these will be zeros to the denominator of G).

Normally the poles to G are identical to the eigenvalues of the matrix A in (A.1). Some eigenvalues may, however, correspond to dynamics that cannot be excited or observed from input–output behavior. Such eigenvalues are not included among the poles.

Since differentiation in the time domain corresponds to multiplication by s of the Laplace transform, we can formally rewrite (A.4) as

$$y(t) = G(p)u(t), \quad p = \frac{d}{dt} \tag{A.7}$$

If G is of the form (A.5), this will correspond to the linear, higher-order differential equation

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_0 \frac{d^n u}{dt^n} + b_1 \frac{d^{n-1} u}{dt^{n-1}} + \cdots + b_n u \tag{A.8}$$

If (A.4) is transformed to the time domain we can also write

$$y(t) = \int_0^t h(\tau) u(t - \tau) d\tau \tag{A.9}$$

where h is called the *impulse response*. $G(s)$ and $h(t)$ are related by G being the Laplace transform of h:

$$G(s) = \int_0^\infty h(t) e^{-st} \, dt \tag{A.10}$$

If the input to a system is chosen as $u = \sin \omega t$ and all its poles have strictly negative real parts, the output is given, when all transients have died out, by the expression

$$y(t) = |G(i\omega)| \sin(\omega t + \arg G(i\omega)) \tag{A.11}$$

The function

$$G(i\omega)$$

which thus can be interpreted as the system's response to the angular frequency ω, is the *frequency function* or *frequency response.*

A.2 Time Discrete Models

A time discrete linear model is given analogously to (A.1) by

$$\begin{aligned} x((k+1)T) &= Ax(kT) + Bu(kT) \\ y(kT) &= Cx(kT) + Du(kT) \end{aligned} \tag{A.12}$$

Here we assume that the signals are measured at equidistant times $(0, T, 2T, \ldots)$, separated by the *sampling interval* T.

If u and y have the z transforms $U(z)$ and $Y(z)$, respectively, they will correspond to

$$Y(z) = G(z)U(z) \tag{A.13}$$

where G is a $p \times m$ matrix called the (time discrete) transfer function.

For (A.12), $G(z)$ can be computed from

$$G(z) = C(zI - A)^{-1}B + D \tag{A.14}$$

If u and y are scalars ($p = m = 1$), then $G(z)$ is a rational function.

$$G(z) = \frac{b_0 z^n + b_1 z^{n-1} + \cdots + b_n}{z^n + a_1 z^{n-1} + \cdots + a_n} \tag{A.15}$$

The z values that satisfy $G(z) = 0$ are called *zeros*, while those for which $G(z) = \infty$ are called *poles* (these will be zeros to the denominator in G).

Normally, the poles of G are identical to the eigenvalues of the matrix A in (A.12).

The preceding equations can also be written in terms of the shift operator q_T, with the properties

$$q_T y(t) = y(t+T), \quad q_T^{-1} y(t) = y(t-T)$$

Since the time shift corresponds to multiplication by z in the z transform, (A.13) can also be written as

$$y(t) = G(q_T)u(t) \tag{A.16}$$

or with G according to (A.14),

$$y(t+nT) + a_1 y(t+(n-1)T) + \cdots + a_n y(t) = b_0 u(t+nT) + \cdots + b_n u(t) \tag{A.17}$$

or, equivalently,

$$\begin{aligned} y(t) = &-a_1 y(t-T) - \cdots - a_n y(t-nT) \\ &+ b_0 u(t) + b_1 u(t-T) + \cdots + b_n u(t-nT) \end{aligned} \tag{A.18}$$

The sampling interval (T) is often implicit, then we write $q = q_T$ and

$$y(t) = G(q)u(t)$$

If (A.13) is transformed to the time domain, we also have the representation

$$y(kT) = \sum_{r=0}^{\infty} g(rT)u((k-r)T) \tag{A.19}$$

where g is the impulse response. $G(z)$ and $g(t)$ are related by G being a z transform of g:

$$G(z) = \sum_{k=0}^{\infty} g(kT)z^{-k} \tag{A.20}$$

The function of ω that we get when we replace z by $e^{i\omega T}$ in G is called the system's *frequency function*:

$$G(e^{i\omega T}) \tag{A.21}$$

It has the same interpretation as in the time continuous case: If the signal $u(kT) = \sin \omega kT$ is subject to a system, with all poles inside

the unit circle, the output is given, when all transients have died out, by the expression

$$y(kT) = \left|G\left(e^{i\omega T}\right)\right| \sin(\omega kT + \phi) \tag{A.22}$$

where

$$\phi = \arg G\left(e^{i\omega T}\right)$$

A.3 Connections between Time Continuous and Time Discrete Models

If the input u is piecewise constant according to

$$u(t) = u(kT), \quad kT \leq t < (k+1)T$$

then (A.1) corresponds to

$$\begin{aligned} x((k+1)T) &= Fx(kT) + Gu(kT) \\ y(kT) &= Cx(kT) + Du(kT) \end{aligned} \tag{A.23}$$

where the matrices F and G are given by

$$F = e^{AT}, \quad G = \int_0^T e^{A\tau} B \, d\tau \tag{A.24}$$

We thus have an exact representation of the type (A.12).

Appendix B

Linearization

B.1 Continuous Time Models

Consider the system

$$\dot{x} = f(x, u) \tag{B.1a}$$

$$y = h(x, u) \tag{B.1b}$$

where

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$$

$$f(x, u) = \begin{pmatrix} f_1(x_1, \ldots, x_n, u_1, \ldots, u_m) \\ f_2(x_1, \ldots, x_n, u_1, \ldots, u_m) \\ \vdots \\ f_n(x_1, \ldots, x_n, u_1, \ldots, u_m) \end{pmatrix}$$

$$h(x, u) = \begin{pmatrix} h_1(x_1, \ldots, x_n, u_1, \ldots, u_m) \\ h_2(x_1, \ldots, x_n, u_1, \ldots, u_m) \\ \vdots \\ h_p(x_1, \ldots, x_n, u_1, \ldots, u_m) \end{pmatrix}$$

Assume that $x = a$ is a stationary solution to (B.1) corresponding to the constant input $u = b$. Then

$$f(a, b) = 0$$

If the function f has continuous partial derivatives in a neighborhood of the point $x = a$, $u = b$, we have for $k = 1, \ldots, n$

$$f_k(x, u) = f_k(a, b) + \frac{\partial f_k}{\partial x_1}(a, b)(x_1 - a_1) + \cdots + \frac{\partial f_k}{\partial x_n}(a, b)(x_n - a_n)$$

$$+\frac{\partial f_k}{\partial u_1}(a, b)(u_1 - b_1) + \cdots + \frac{\partial f_k}{\partial u_m}(a, b)(u_m - b_m) + r_k(x - a, u - b)$$

where the remainder term r_k is small. More precisely,

$$r_k(x - a, u - b) = o(|x - a| + |u - b|)$$

where $|\cdot|$ is some vector norm. With the notations

$$A = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{pmatrix} \tag{B.2}$$

where the partial derivatives have been evaluated at $x = a$ and $u = b$, we can write

$$\begin{aligned} f(x, u) &= f(a, b) + A(x - a) + B(u - b) + r(x - a, u - b) \\ &= A(x - a) + B(u - b) + r(x - a, u - b). \end{aligned}$$

With the notation

$$z = x - a, \quad v = u - b$$

(B.1a) can be written as

$$\dot{z} = Az + Bv + r(z, v) \tag{B.3}$$

In an analogous way, (B.1b) can be written as

$$w = Cz + Dv + \tilde{r}(z, v), \quad \tilde{r}(z, v) = o(|z| + |v|) \tag{B.4}$$

with $w = y - h(a, b)$ and

$$C = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_p}{\partial x_1} & \cdots & \frac{\partial h_p}{\partial x_n} \end{pmatrix}, \quad D = \begin{pmatrix} \frac{\partial h_1}{\partial u_1} & \cdots & \frac{\partial h_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial h_p}{\partial u_1} & \cdots & \frac{\partial h_p}{\partial u_m} \end{pmatrix} \tag{B.5}$$

where the partial derivatives have been evaluated at $x = a$ and $u = b$. If the remainder terms r and $\tilde{r}$ are neglected, we have a linear model of the type (A.1).

B.2 Discrete Time Models

The system

$$x(t_{k+1}) = f(x(t_k), u(t_k)) \qquad k = 0, 1, 2, ... \tag{B.6a}$$

$$y(t_k) = h(x(t_k), u(t_k)) \tag{B.6b}$$

can be linearized around $x = a$, $u = b$ in an analogous manner to the time continuous case:

$$x(t_{k+1}) - a = A(x(t_k) - a) + B(u(t_k) - b) + r \tag{B.7a}$$

$$y(t_k) - h(a, b) = C(x(t_k) - a) + D(u(t_k) - b) + \tilde{r} \tag{B.7b}$$

where the matrices A, B, C, and D are given by

$$A = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_m} \end{pmatrix} \tag{B.8a}$$

$$C = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_p}{\partial x_1} & \cdots & \frac{\partial h_p}{\partial x_n} \end{pmatrix}, \quad D = \begin{pmatrix} \frac{\partial h_1}{\partial u_1} & \cdots & \frac{\partial h_1}{\partial u_m} \\ \vdots & & \vdots \\ \frac{\partial h_p}{\partial u_1} & \cdots & \frac{\partial h_p}{\partial u_m} \end{pmatrix} \tag{B.8b}$$

with all partial derivatives evaluated at $x = a$, $u = b$. The remainder terms r and $\tilde{r}$ fulfill the same conditions as in the time continuous case if f and h are continuously differentiable functions.

Appendix C

Signal Spectra

C.1 Time Continuous Deterministic Signal with Finite Energy

We have the signal

$$w(t), \quad -\infty < t < \infty$$

with the property

$$\int_{-\infty}^{\infty} |w(t)|\, dt < \infty \tag{C.1}$$

We define the Fourier transform

$$W(\omega) = \int_{-\infty}^{\infty} w(t) e^{-i\omega t} dt \tag{C.2}$$

and the inverse Fourier transform

$$w(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\omega) e^{i\omega t} d\omega \tag{C.3}$$

The spectrum of $w(t)$ is defined as

$$\Phi_w(\omega) = |W(\omega)|^2 \tag{C.4}$$

Parseval's equation gives

$$\int_{-\infty}^{\infty} w^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_w(\omega) d\omega \tag{C.5}$$

C.2 Sampled Deterministic Signals with Finite Energy

We have a signal sampled with the sampling frequency T:

$$w[k] = w(kT), \qquad k = 0, \pm 1, \pm 2, \ldots \tag{C.6}$$

with the property

$$\sum_{k=-\infty}^{\infty} |w(kT)| < \infty \tag{C.7}$$

We define the discrete Fourier transform

$$W^{(T)}(\omega) = T \cdot \sum_{k=-\infty}^{\infty} w(kT) e^{-i\omega kT} \tag{C.8}$$

and the inverse transform

$$w(kT) = \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} W^{(T)}(\omega) e^{i\omega kT} d\omega \tag{C.9}$$

The spectrum is defined as

$$\Phi_w^{(T)}(\omega) = |W^{(T)}(\omega)|^2 \tag{C.10}$$

Parseval's equation gives

$$T \cdot \sum_{k=-\infty}^{\infty} w^2(kT) = \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} \Phi_w^{(T)}(\omega) d\omega \tag{C.11}$$

C.3 Connections between the Continuous and the Sampled Signal

The following links exist between $W(\omega)$ and $W^{(T)}(\omega)$ (Poisson's summation formula):

$$W^{(T)}(\omega) = \sum_{r=-\infty}^{\infty} W(\omega + r\omega_s) \tag{C.12}$$

($\omega_s = 2\pi/T$). The equation is obtained from (C.3):

$$\begin{aligned}
2\pi w(kT) &= \int_{-\infty}^{\infty} W(\omega)e^{i\omega kT}\,d\omega = \sum_{r=-\infty}^{\infty}\int_{-\omega_N+2\pi r/T}^{\omega_N+2\pi r/T} W(\omega)e^{i\omega kT}\,d\omega \\
&= \sum_{r=-\infty}^{\infty}\int_{-\omega_N}^{\omega_N} W(\omega+\frac{2\pi r}{T})e^{\omega kT+i2\pi rk}\,d\omega \\
&= \int_{-\omega_N}^{\omega_N}\left(\sum_{r=-\infty}^{\infty} W(\omega+\frac{2\pi r}{T})\right)e^{i\omega kT}\,d\omega
\end{aligned}$$

where ω_N is the Nyquist frequency ($\omega_N = \omega_s/2$).

C.4 Signals with Infinite Energy

If (C.1) does not hold, we define a truncated signal

$$w_S(t) = \begin{cases} w(t) & |t| < S \\ 0 & |t| \geq S \end{cases} \tag{C.13}$$

This signal has finite energy and its spectrum $\Phi_{w_S}(\omega)$ is defined as in Section C.1. The *power spectrum* for the original signal is defined as

$$\Phi_w(\omega) = \lim_{S\to\infty}\frac{1}{S}\Phi_{w_S}(\omega) \tag{C.14}$$

Through normalization with the length of the time interval the unit for $\Phi_w(\omega)$ will in this case be "power" per frequency.

Remark: For many signals $w(t)$ the limit (C.14) does not exist for each ω. The limit then has to be interpreted in a "weak" sense, that is, $\Phi_w(\omega)$ is a function such that

$$\lim_{S\to\infty}\int_{-\infty}^{\infty}\frac{1}{S}\Phi_{w_S}(\xi)f(\xi-\omega)d\xi = \int_{-\infty}^{\infty}\Phi_w(\xi)f(\xi-\omega)d\xi \tag{C.15}$$

for any "nice" function $f(\xi)$. Consider, for example, $f(\xi)$ as an approximation of Dirac's delta function. The delta function itself would give (C.14) from (C.15).

For the sampled signal the corresponding operations can be carried out. If $w(kT) = 0$, $k \leq 0$ we have

$$\Phi_w^{(T)}(\omega) = \lim_{S\to\infty}\frac{1}{S}\Phi_{w_S}(\omega)$$

$$= \lim_{S\to\infty} \frac{1}{S} \left| w_S^{(T)}(\omega) \right|^2$$

$$= T \cdot \lim_{N\to\infty} \frac{1}{N} \left| \sum_{k=1}^{N} w(kT) e^{-ikT\omega} \right|^2 \tag{C.16}$$

where $S = N \cdot T$. The comment about weak convergence is valid also here.

C.5 Stochastic Processes

Consider a stationary process $\{w(kT)\}$ with the sampling interval T, with mean 0 and the covariance function

$$R_w(kT) = Ew(t + kT)w(t) \tag{C.17}$$

(E denotes the expected value). Its spectrum (spectral density) is defined as

$$\Phi_w(\omega) = T \sum_{k=-\infty}^{\infty} R_w(kT) e^{-i\omega kT} \tag{C.18}$$

If the term "spectrum" for (C.18) is going to be reasonable, the $\Phi_w(\omega)$ has of course to describe the "typical frequency contents" in realizations of $\{w(t)\}$. That this is the case we can see from the following. Define

$$W_N(\omega) = T \sum_{k=1}^{N} w(kT) e^{-i\omega kT} \tag{C.19}$$

Then

$$\frac{1}{T} \lim_{N\to\infty} E \frac{1}{N} \mid W_N(\omega) \mid^2 = \Phi_w(\omega) \tag{C.20}$$

with $\Phi_w(\omega)$ defined in (C.18). Compare to (C.16)! This result follows from the following calculations:

$$E \frac{1}{TN} \mid W_N(\omega) \mid^2 = \frac{T}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} e^{-i\omega kT} e^{i\omega \ell T} Ew(kT)w(\ell T)$$

$$= \frac{T}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} R_w(\ell T - kT) e^{-i\omega T(\ell - k)}$$

$$= [s = \ell - k] = T \sum_{s=-N}^{N} (1 - \frac{\mid s \mid}{N}) R_w(sT)^{-iwsT}$$

When $N \to \infty$, the right side converges to (C.18) under the condition that

$$\sum_{-\infty}^{\infty} \mid R_w(sT) \mid < \infty$$

If we compare to (C.16) we find that the spectrum (C.18) for a stochastic process is the expected value of the effect spectrum for realizations of $\{w(t)\}$. $\Phi_w(\omega)$ thus describes the *average frequency contents* of $\{w(t)\}$ and has then the same physical interpretation as the spectra we discussed earlier. The corresponding is true for time continuous stochastic processes.

Parseval's equation takes on the following expression for stochastic processes:

$$Ew^2(t) = \int_{-\pi/T}^{\pi/T} \Phi_w(\omega) d\omega \tag{C.21}$$

Cross Spectra

The cross spectrum of two stationary processes with zero mean value and sampling interval T is defined as

$$\Phi_{yu}(\omega) = T \cdot \sum_{k=\infty}^{\infty} R_{yu}(kT) e^{-i\omega kT} \tag{C.22}$$

where

$$R_{yu}(kT) = Ey(t + kT)u(t) \tag{C.23}$$

For time continuous processes the definition is analogous. As in (C.20) we can show that

$$\Phi_{yu}(\omega) = \frac{1}{T} \lim_{N \to \infty} E \frac{1}{N} Y_N(\omega) \overline{U_N(\omega)} \tag{C.24}$$

where $Y_N(\omega)$ and $U_N(\omega)$ are defined analogously to (C.19).

For deterministic signals we can also work with cross spectra defined as $Y(\omega)\overline{U(\omega)}$ or (weak) limits of such expressions normalized, analogously to (C.4), (C.14) and correspondingly for sampled signals.

A good feeling for what a cross spectrum says about the relations between the two signals is obtained from the following: If y and u are subject to

$$y(t) = G(q_T)u(t) \tag{C.25a}$$

then

$$\Phi_{yu}(\omega) = G(e^{i\omega T})\Phi_u(\omega) \tag{C.25b}$$

This is obtained from (C.24) and (A.13).

Index

A

Adams-Bashforth, 323
Adams-Moulton, 323
aggregation, 101
AIC, 280
air drag, 111
Akaike's information criterion, 280
algebraic loop, 310
algorithm for equation sorting, 181
alias effect, 268
amplitude scaling, 298
amplitude spectra, 66
analog computers, 297
analogies, 121
analytical solution, 172
angular velocity, 113
antialias filter, 268
approximate, 72
approximation, 98
AR process, 63
area of validity, 17
ARMA process, 63
ARMAX model, 233
ARX model, 233
asymptotic stability, 48

B

basic equations, 91
bias error, 245
biological system, 170
BJ-model, 232
Blackman–Tukey approach, 214
block diagram models, 33
Box-Jenkins model, 232

C

capacitance, 108
causal stroke, 143
causality, 143
causality algorithm, 147
change-oriented models, 21
choice of input, 154
computer algebra, 170
conservation law, 92
constant, 36
constitutive relationships, 92
control signal, 38
controlled element, 154
correlation analysis, 196, 197
covariance function, 64
CRA, 196, 197
credibility, 336
cross spectra, 355
cross spectrum, 67
cross validation, 279
current, 108

D
d'Arcy's law, 117
dc motor, 138, 181, 228
deterministic model, 19
differential analyzer, 297
direct feedthrough, 310
discrete event systems, 21
discrete time model, 20
displacement operator, 60
distributed parameter model, 21
disturbance signal, 38, 53
dry friction, 111
dynamic model, 19
dynamic system, 40

E
ecological system, 23
economic system, 29
effort, 123
effort source, 126
effort storage, 125
effort variable, 120
eigenvalue, 342
empirical transfer function estimate, 207
energy storage, 109, 111
equilibrium, 47
estimate, 206
Euler's method, 44, 319
experimentation, 13
explicit method, 319
external signal, 37

F
final prediction error, 280
flow, 115, 123
flow resistance, 117
flow source, 126
flow storage, 124
flow systems, 114
flow variable, 120
fluid capacitance, 115
force, 110
Forrester's world model, 104
Fourier transform, 351
FPE, 280
frequency analysis, 200
frequency function, 343, 344
frequency resolution, 212
frequency response, 343
friction, 111

G
Gauss-Newton method, 254
global error, 319
globally asymptotically stable, 48
gyrator, 135

H
hardware, 304
head box, 85
heat conduction, 102
heat flow rate, 120

I
identifiability, 251
identifiable, 251
identification, 17
ill-posed problems, 152
implicit method, 319
impulse response, 342, 344
incompressible, 114
inductance, 108
inertance, 115

input, 38
integral, 173
internal variable, 39

J

Jacobian, 51

K

k-step method, 319
Kirchhoff's laws, 109

L

Laplace transform, 342
linear system, 46
linear systems, 341
linearization, 347
local error, 319
lumped, 21

M

MA process, 63
mass, 110
mathematical expectation, 64
mathematical model, 15
maximum likelihood, 237
MDL, 281
mechanical rotation, 112
mechanical translation, 110
mechanics-hydraulics, 139
mental model, 14
merging, 131
ML estimation, 237
model, 14
model order, 44
model validation, 334
modulated, 157
moment of inertia, 113

N

Newton's force law, 110
nonparametric identification methods, 260
Nyquist frequency, 60, 76, 353

O

OE model, 232
order of accuracy, 320
output, 37
output-error model, 232
overfit, 280

P

p junction, 130
paper machine, 85
parallel junction, 130
parameter, 37
parametric estimation methods, 260
Parseval's equation, 351, 355
parsimonious, 274
partial differential equations, 101
pendulum, 160, 173
periodogram, 210
phases of modeling, 83
physical model, 15
physical model building, 16
Poisson's summation equation, 352
Poisson's summation formula, 76
pole, 342, 343
posttreatment of data, 269
power spectrum, 66, 353
pressure, 115
Procrustes, 337

R
realization, 63
regression vector, 236
regressors, 236
residuals, 284
resistance, 108, 125
resistive, 125
Rissanen's minimal description length, 281

S
s junction, 127
sampled model, 20
sampled signal, 71
sampling frequency, 60
sampling interval, 60, 343
sampling theorem, 77, 268
scaling, 298
semiphysical modeling, 277
separation of time constants, 99
series junction, 127
shift operator, 344
signal, 36
signal to noise ratio, 196
simplification, 97, 131
simulation, 15, 295
simulation models, 36
simulator, 328
singular point, 47
SITB, 261, 291
source, 126
spectral factorization, 69
spectrum, 65, 351
spring constant, 111
state, 43, 150, 174, 178
state equations, 178
state-space model, 95
static gain, 50
static model, 19
static relationship, 49, 92
static system, 40
stationary point, 47
stationary solution, 47, 348
stiff differential equation, 100
stochastic model, 19
stochastic process, 63, 354
structuring, 86
system, 13

T
tank system, 27
temperature, 120
thermal capacity, 120
thermal systems, 119
time constant, 99
time continuous model, 20
time invariant system, 46
time scaling, 298
torque, 112
torsional stiffness, 113
trajectory, 46
transfer function, 59, 342, 343
transformer, 109, 135

U
uncorrelated, 67

V
validate, 17
validation, 260
validity, 336
variable, 36
variance errors, 245
velocity, 110
verbal model, 14, 36
viscous friction, 111

voltage, 107

W

Welsh's method, 213
white noise, 62
whitening filter, 196

Z

zero, 342, 343