Adaptive Start-up Control

Lenells, Matz

1982

Document Version:
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Lenells, M. (1982). *Adaptive Start-up Control*. [Doctoral Thesis (monograph), Department of Automatic Control]. Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

CODEN: LUTFD2/(TFRT–1023)/1–160/(1982)

Adaptive Start-up Control

Matz Lenells

LTH

Department of Automatic Control, Lund Institute of Technology

ADAPTIVE START-UP CONTROL

av

Matz Lenells
Civ ing, Sm

Akademisk avhandling som för avläggande av teknisk doktorsexamen vid Tekniska Fakulteten vid Universitetet i Lund kommer att offentligen försvaras i sal M:B, Lunds Tekniska Högskola, fredagen den 26 november 1982, kl 10.15.

Adaptive Start-up Control

Matz Lenells

Lund 1982

Printed in Sweden
Studentlitteratur

LUND INSTITUTE OF TECHNOLOGY
DEPARTMENT OF AUTOMATIC CONTROL
Box 725
S 220 07 Lund 7 Sweden

Document name
Doctoral Dissertation

Date of issue
November 1982

Document number
CODEN:LUTFD2/(TFRT-1023)/1-160/(1982)

Author(s)
Matz Lenells

Supervisor
Per Hagander, Karl Johan Åström

Sponsoring organization

Title and subtitle
Adaptive Start-up Control

Abstract
How to use a set of possible models for control purposes is investigated in the case of adaptive start-up control, i.e. when taking a process from one working point to another one without overshoot. The process properties are to some extent unknown. Concepts are introduced to describe optimality of a control law. An on-line control method is proposed based on the introduced concepts. The main idea is to reject the impossible models after each sampling and then to choose the input u according to the worst case. The rejection of models is based on a model error which is associated with each model. It may also depend on time and the input u. For a special class of model sets, called monotone sets, the optimal control law is found. Dynamical programming is used to prove the existence of an optimal control law in the general case.

Start-up control is discussed for different types of heat processes. Heating water on an ordinary hot-plate is chosen as an example. An implementation is described for this process of a controller based on the introduced concepts. The controller makes use of real-time programming using Pascal extended with a real-time kernel. A comparison is made with a PD-based controller, which is also analysed by means of the introduced concepts.

Key words
set of models, physical models, PD-controller, optimal control.

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title

ISBN

Language
English

Number of pages
160

Security classification

Recipient's notes

Distribution: The report may be ordered from the Department of Automatic Control or borrowed through the University Library 2, Box 1010, S-221 03 Lund, Sweden, Telex: 33248 lubbis lund.

CONTENTS

CHAPTER 1

INTRODUCTION

Servo problems are recognized as a special class of problems in automatic control. The purpose is to follow a given reference trajectory. A special type of servo problem is the start-up problem. The objective is then to take the process from one working point to another. Heating up a furnace in a metallurgical industry is an example of a start-up problem. Heating milk on a hot-plate without burning it is an example from every day life.

The start-up problem has been studied with great success using the theory of optimal control. Large computers have been of special significance in obtaining numerical solutions to many difficult problems. In spite of the progress in theory and numerics, the practical results are modest. The primary problem is the discrepancy between the model used and reality for which there are three main reasons. The first one is the necessity of making simplifying approximations of the physical laws describing the process. The second one is the lack of knowledge about the physical process. The third one is that there are changes in the physical process from time to time. In the problem of heating milk on an hot-plate the milk content can vary from time to time.

The problem of uncertainty in modelling a process is basically solved in two different ways. In robust control, the effect of the uncertainty is often minimized, by use of advanced forms of feedback control with high gain.

The other approach is to increase knowledge about the process on-line during the control, and use it to gradually improve the control. Such control is often called adaptive control (Åström, 1981) or self-organizing control (Saridis, 1970). A common type of adaptive control assumes that the process can be described by a model belonging to a general class of models, e.g. ARMA-models. The difference between process and model is assumed to be of statistical nature. The control algorithm assumes that the process is described by the model most closely related to the process in some sense. However, the process-model discrepancy is, in essence, due to modelling errors. Even if the control algorithm takes into consideration the uncertainty described by a statistical method, serious mistakes are made especially in a start-up problem, in assuming the applicability of a given model to the process.

An alternative approach is to let the control algorithm at each sampling instant consider a set of models as possible descriptions of the process and control with respect to the worst case. To develop, study, and discuss such a method in the case of start-up control is the topic of the present work. The work is guided by the idea that the method is to include as much a priori knowledge as the control engineer wants to use. Further, the information gained during control has to be absorbed in an effective way. This means that it is not the method itself, but the computer capacity, that limits the amount of usable information. This implies, further, that the proposed method should encourage the engineer to seek more information about the process which is to be presented in the common language of physics and not in a special language for control engineers. Such an approach would also stimulate others around the control engineer to use and increase knowledge about the process.

The proposed method of adaptive control is especially suitable for start-up problems. The time-sequence is limited which, among other things, implies that it is possible to store the whole course of events during the start-up. This permits computers to apply advanced control schemes without the use of recursive algorithms. Adaptive control used to start-up problems is herein called adaptive start-up control.

Synopsis

An introduction to adaptive start-up control is given in Chapter 2 as is also an outline of the thesis. In Chapter 3, concepts of the proposed method are defined. For a certain class of adaptive start-up problems the optimal control is given in Chapter 4. Here also dynamic programming is shown to be applicable in seeking an optimal control law to adaptive start-up problems. A discussion of the concepts defined in Chapter 3 is undertaken in Chapter 5. In Chapter 6, the PD-controller is analysed as an adaptive start-up controller. The results of Chapter 5 are used in Chapter 7 when designing an adaptive start-up controller for a real process. Finally, a discussion about the proposed controller and some suggestions are given in Chapter 8.

CHAPTER 2

PRELIMINARIES TO ADAPTIVE START-UP CONTROL

Three examples of start-up control problems will be given. The specific properties are discussed. Based on the examples, a basic start-up problem is defined. The thesis is outlined in Section 2.4 together with a proposal of a method of start-up control.

2.1 EXAMPLES OF START-UP PROBLEMS

Examples will be presented to illustrate the use of adaptive start-up control.

Furnaces in metallurgical industry

Metallurgical industry uses many kinds of furnaces which can be divided into two groups: the batch type and the continuous type. The batch type is used, for instance, when giving cutting alloys the necessary heat treatment. Cutting alloys are made out of small chunks of pressed powder of iron and alloying materials. The chunks are given special temperature treatment by being placed on stacked trays into a furnace. The control problem is to heat the furnace so that the chunks of pressed powder follow a prescribed temperature program. In this process two main problems arise: the time lag is too long for conventional PID-controllers and the weight of the batch may vary from time to time.

A continuous furnace, in theory, does not have more than one start-up. In practice, however, the situation is different. A reheating furnace in a hot-strip mill is a good example. The furnace heats up slabs to enable them to roll. The slabs are often large, approx. 1x2x0.3 m and are pushed into the furnace from one side and taken out from the other. The furnace itself can be very impressive: the walking beam furnace at SSAB in Oxelösund, Sweden, for instance, is 40 m long, 8 m wide and about 5 m high. When a stop occurs in the rolling mill, to change the rolls for example, an interesting control problem appears. The stream of heated slabs must be stopped and thus the power of the furnace is lowered. When the rolling mill is in function again the stream of heated slabs must go on. During such a pause, the time of restarting the rolling mill can often be estimated in advance with good precision. It is tempting to reduce the power as much as possible at the beginning of the pause, and at a suitable time to raise the temperature so that the stream of slabs can continue with minimal disruption. Such a control would save a lot of energy, and reduce the oxidation of iron. Oxidated iron must be descaled: the material lost of this can be one to two per cent of the total weight which indicates that the financial benefit of better temperature control would be considerable.

Heating houses

With the increasing cost of energy it has become more and more interesting to heat houses according to need. This means, for instance, a difference in day and night temperatures. Though this is in no way new, it is only recently that constant indoor temperature has become standard in Sweden. Before the use of oil and electricity, wood, coal, or coke were used. Thermostats were not in use and during the night with no one tending it, the fire died out only to be lighted again in the early morning. Nowadays the trend is to use a timer to obtain the desired temperature changes. One of the problems is thus to determine when to start the heating in the morning.

When heating assembly halls like offices, schools, gymnasia, churches etc. the control problem is more accentuated. Here, automatic control is often necessary because the consequences of choosing a wrong starting-time are more serious than with the case of ordinary houses. It is often much harder to obtain immediate change when it is discovered that the temperature is too low: the heating capacity of the heating plant is often small in relation to the size of the building. From this it is also easier to motivate investment in a more expensive heating system.

Heating water on a hot-plate

In the introduction, the example of heating milk on a hot-plate was mentioned. A common but less dramatic start-up problem is heating water on a hot-plate. Ordinarily the heating of water is hardly considered a control problem since boiling water is often desired. Boiling water is easy to get and maintain if no attention is paid to energy consumption. There are, however, situations in which water should not boil but rather have a temperature of about 95°C, for example when something in the water is cooking. For instance, fish can be spoiled if cooked in boiling water. Many more reasons are given in Gyllensköld (1977).

2.2 CHARACTERISTIC FEATURES OF START-UP PROBLEMS

The above examples of start-up problems not only vary greatly, they are also of very different character because of physical and economical reasons.

A furnace used to heat up cutting alloys is very expensive which might motivate work towards quality models. Therefore, accurate and extensive measurements can be motivated. If it is assumed, for instance, that the weight of the load is known, it might be possible to describe the furnace by a fixed model. The problem of such a start-up control is therefore reduced to a problem of calculating a suitable control for a fixed model. Observe that the discrepancy between model and reality still has to be considered.

The example of the hot-plate shows a completely different picture. Exact modelling here is out of the question. The actual model depends on which kettle is used, if it is used with or without a lid, and on the amount of water in the kettle. To mount sensors in this case is not economical. The uncertainty of the model has to be accepted.

From a purely technical view point, the problem of heating houses is, in a way, like the furnace problem if it is disregarded that the controller cannot know how the weather will be. The way to solve this problem is thus to make a good temperature model of the house, install temperature sensors, humidity sensors, sun radiation meters, wind meters etc.. The disadvantage with such a solution is that it is, today, too expensive. Therefore a controller has to rely on very few sensors. Further, the models cannot be tailor-made for individual houses using standard methods. There are, thus, similarities with the hot-plate case; however, there is one significant difference. Every start-up of the hot-plate is a new start-up because the amount of water varies from time to time. With houses there are, however, many parameters which are constant from time to time but unknown from the beginning. In such a case the controller should remember the procedure of previous start-ups, or at least gather information about the unknown parameters. In a way, the controller could function as does an installation engineer. Presently the routine when installing a heat plant controller in a building (not small houses) is that an engineer has to trim the controller. This takes from a couple of hours to a couple of days. Even after two day's work it is a matter of chance partly because the weather is seasonal. A smart adaptive controller could thus cut the costs of installation.

2.3 THE BASIC START-UP PROBLEM

The start-up problems presented in Section 2.1 are of various forms. In order to make the presentation concrete both here and in the rest of the thesis, a basic start-up problem is formulated.

Problem formulation: "The basic start-up problem". Regard a dynamic system with input u and output y. The basic start-up problem is to raise the output y to a reference value y_{ref} in minimum time without a too big overshoot.

□

This problem may seem to be an oversimplification of the start-up problems presented in the examples. However, it is in fact complex enough to give inspiration to and illustration of new concepts and theory believed to be useful when doing adaptive start-up control.

2.4 OUTLINE OF THE THESIS

The main topic of the thesis is to solve the basic start-up problem. The concepts and definitions introduced in Chapter 3 are, however, such that they can be used for, or extended to, more general start-up problems. On the other hand, the results given by the theorems deal only with the basic start-up problem.

A laboratory set up of a kettle of water on a hot-plate (see Figure 2.1 and Appendix A), together with appropriate mathematical models (see Appendix B), is used to illustrate the thesis.

The idea of the proposed method is to use a set, Σ, of models of the process. At least one of the models is supposed to describe the process, but not necessarily exactly. It is assumed that the measured output differs from the output of the describing model with, at most, a certain

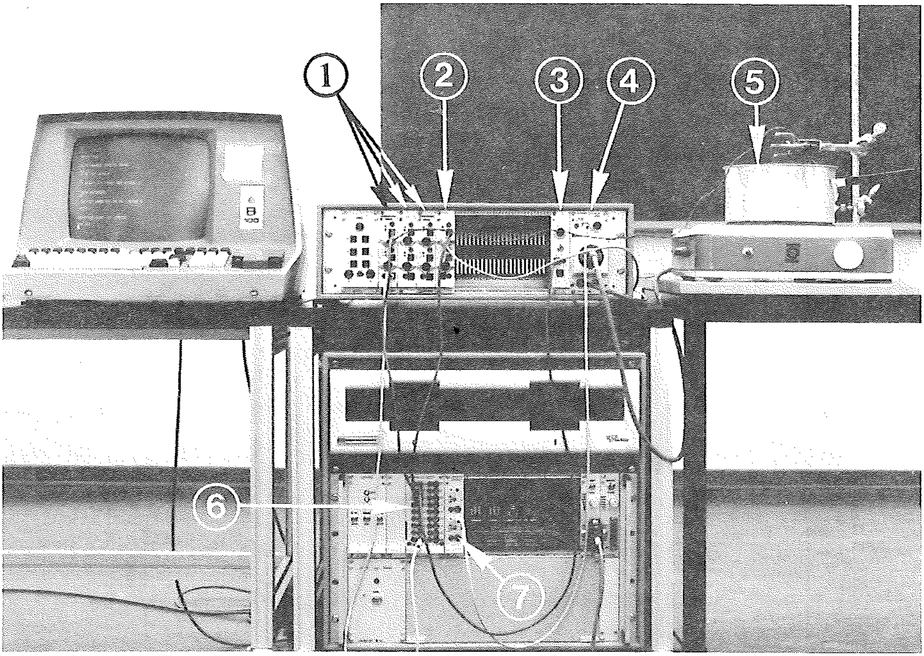

Fig. 2.1: The laboratory set-up of the hot-plate process. (1) Temperature bridges. (2) Constant current source used when measuring the plate temperature. (3) Voltage source used as a trigging signal to the computer. (4) Power control of the hot-plate. (5) 2 litre aluminum kettle. (6) A/D converters, 12 bits, ±10 V. (7) D/A converters, 12 bits, ±10 V. For more detailed information, see Appendix A.

quantity ε, which has to be given by the user. During control, all models which differ with more than the quantity ε from the measurements, are rejected as being impossible descriptions of the process. Figure 2.2 shows what is meant by a possible model and an impossible one.

The control u(t) has to be such that, independently of which possible model that describes the process, the overshoot will not be too large. Observe that the set of possible models will diminish with time.

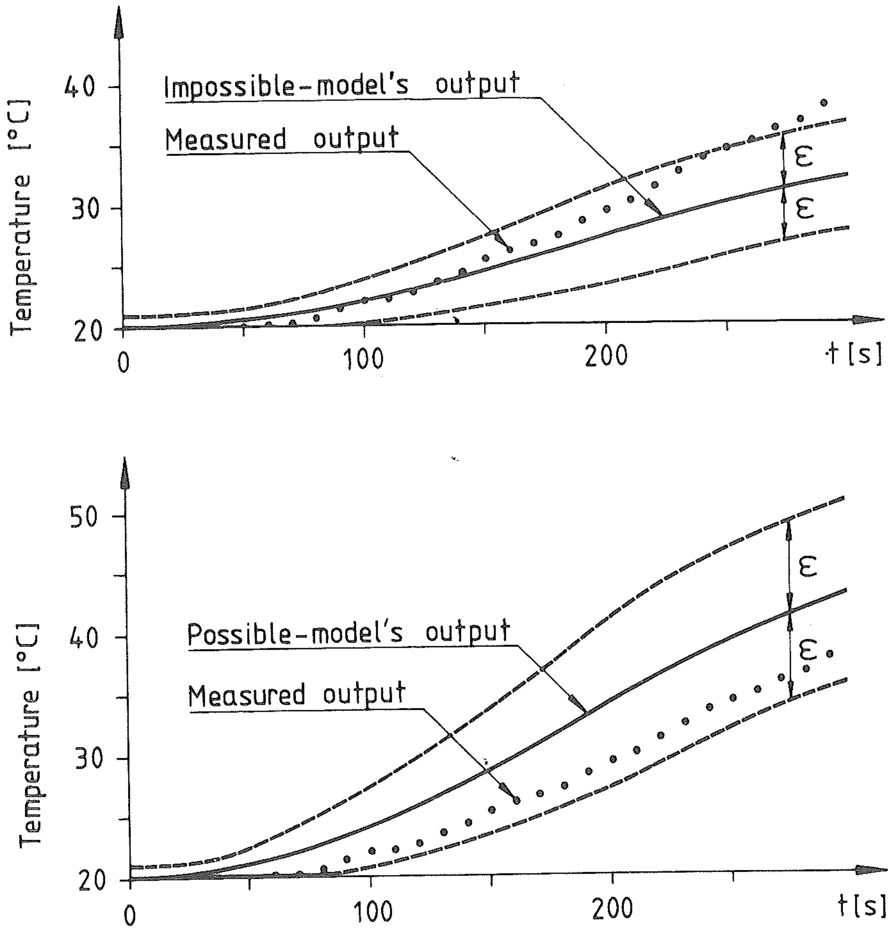

Fig. 2.2: The exemplification of a possible and an impossible model.

The ultimate goal of the control is to make the start-up in minimum time. The prohibition of too-large overshoots applies to every start-up. Statistical models are therefore not useful. Furthermore, the main part of the error is, in most cases, not of statistical nature. For instance, repeated experiments of the hot-plate process give nearly the same output provided the quantity of water is not changed. To find a model of reasonable complexity which accurately describes the process seems, however, to be impossible. Consequently, the difference between the outputs from the describing model and the process is nearly the same from experiment to experiment. The error will depend on the input but also on the model. This means that it is advantageous to calculate the quantity of water during control. This is discussed in Chapter 5.

The posed problem requires a way to compare different control laws and the establishment of a criterion of optimality. Due to the possibility of an error in the description of the process, the definition of optimality given in Section 3.2 is quite involved. In most cases it is very difficult to find an optimal control law. In Section 3.3, it is shown that all information of old outputs necessary for optimal control can be condensed to the set of models remaining after model rejection. For a special class of adaptive start-up problems it is possible to find an optimal control law. This is shown in Theorem 4.1. In Chapter 4, it is also shown that Theorem 4.1 can be applied for a special case of the basic start-up problem of the hot-plate process. The optimal control law is also found. In a general case, dynamic programming can be used at least in principle. This is discussed in the end of Chapter 4.

Even if the main purpose is to present a new way of control, it is of interest to examine the possibility of using PID-controllers. In Section 6.1, it is proved that a high-gain PD-controller with controller saturation is optimal in a special case. Furthermore, the hot-plate the high gain controller may work well. Even if the high gain PD-controller works well in a certain case it is argued in Section 6.3 that in a case of more practical interest several difficult problems would be experienced.

In Chapter 7 the proposed method is applied to the hot-plate process in the case for which it is known that no lid is used and that the plate temperature starts at room temperature. The only uncertainties are the quantity of water and the amount of losses. The computer used is so slow that the rejection of impossible models does not keep up with real time. However, by the use of concurrent programming, the control is done properly even if the lag in the rejection of impossible models makes the start-up time considerably longer than necessary. Furthermore, the short-comings in speed and memory of the computer under use prevents application to cases where the high gain PD-controller would have been unsuitable. For the present case the PD-controller is better.

Last, in Chapter 8, a discussion is undertaken about the proposed method.

C H A P T E R 3

T H E O R E T I C A L F R A M E W O R K F O R S T A R T - U P C O N T R O L

In this chapter the start-up problem is formulated and involved concepts are defined. By using these concepts optimality of a control law can be defined. This is done in two steps. In Section 3.1, accurate models are assumed, simplifying the definition a great deal. The concepts of Section 3.1 are extended in Section 3.2 to the case of inaccurate models.

The definition of optimality applied to the basic start-up problem implies that there are optimal control laws which in a certain way are insensitive to small perturbations in the output. It is also shown in Section 3.3 how this is used when a control law compress the output history into a set of possible models.

3.1 FORMULATION WITH ACCURATE MODELS

This section deals with the ideal case when an accurate model of the system is possible to formulate. The following concepts are defined: description map, learning, control objectives, and optimality.

Description of systems

The input-output function for a discrete time dynamical system will be described by map H. The time unit chosen is to be the sampling period if nothing else is stated.

Definition 3.1: A sequence $s_{[0,t]}$ is a sequence of real numbers

$$s_{[0,t]} = [s(0), s(1), \ldots\ldots , s(t)].$$

S_t denotes the set of sequences $s_{[0,t]}$. S denotes all sequences for all t. If $t<0$, $s_{[0,t]}$ is defined as the empty sequence. S_∞ denotes the set of sequences of infinite length.

□

Definition 3.2: A description map H (often called model) is a function from S to S, such that

$$\text{(i)}\quad H(u_{[0,t]}) \in S_{t+1}$$

$$\text{(ii)}\quad H(u_{[0,s]})(r) = H(u_{[0,t]})(r) \qquad 0 \le r \le \min(s,t)+1$$

where H may be restricted to a subset of S.

□

In order to illustrate the concept regard a process defined by

$$\left.\begin{aligned} x(t+1) &= a\cdot x(t) + b\cdot u(t) \\ y(t) &= c\cdot x(t) \\ x(0) &= x_0 \end{aligned}\right\} \qquad (3.1)$$

All variables and parameters are real numbers. When a, b, c and x_0 are given, the process (3.1) defines a description map. By varying a, b, c, and x_0 the process (3.1) defines a whole family of description maps. Notice that the initial condition is regarded as just another parameter.

Definition 3.3: A process with some unknown parameters defines a set of description maps, denoted by Σ^a (a = accurate).

□

The hot-plate described in Section 2.4 defines a process. A priori knowledge, such as knowing that water is used with the water content m_w satisfying $m_w \in [0.5, 2.0]$ kg limits the number of description maps defined by the process.

Description of learning

During the start-up of a process, additional information is gained about the process. This learning can be expressed by using the subset of description maps H that can describe the experienced input-output behaviour.

Definition 3.4: The subset $\Sigma^a_{t,u,y}$ of Σ^a is defined by:

$$\Sigma^a_{t,u,y} = \{H;\ H \in \Sigma^a \text{ and } H(u_{[0,t-1]})(s)=y(s) \text{ for } 0 \le s \le t\}.$$

The dependence of u and y may be suppressed in order to get a shorter notation.

□

Definition of loss function and control law

The basic start-up problem can be formulated in a more rigorous way than earlier in Section 2.3. Figure 3.1 illustrates the following definition:

Definition 3.5: The basic start-up problem is to find u which raises y in minimum time t_m, such that

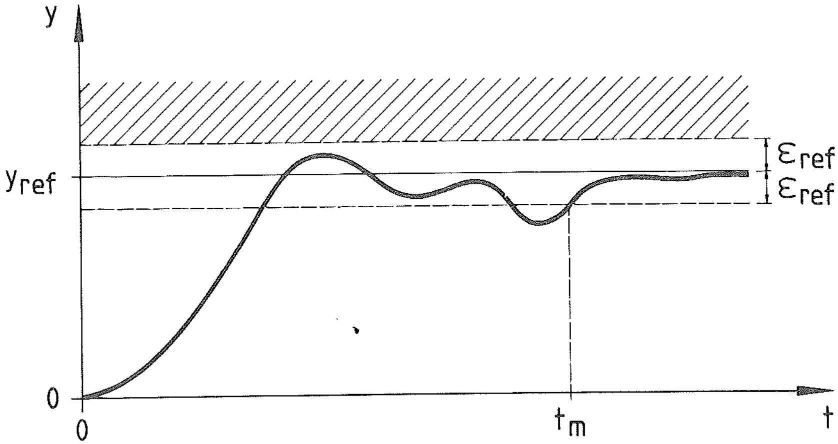

Fig. 3.1: Illustration of the definition of loss as minimum time t_m in the case of the basic start-up problem.

$$y(t) \geq y_{ref} - \varepsilon_{ref} \qquad t \geq t_m$$

subject to the constraints

$$\left.\begin{array}{l} y(t) \leq y_{ref} + \varepsilon_{ref} \\ u(t) \in [0, u_{max}] \end{array}\right\} \text{for all} \quad t \geq 0 \qquad (3.2)$$

where y is given by $y_{[0,t]} = H(u_{[0,t-1]})$

and H belongs to Σ^a.

□

In the basic start-up problem the loss is given by t_m. In general, for adaptive start-up problems, both the loss function and the constraints can be more complicated. In the case of furnaces for instance, the temperature gradients can be limited and the energy consumption can be included in the loss.

The control of a process is based on available knowledge, i.e. on old inputs and outputs and the set Σ^a of description maps H. A natural definition of a control law with a corresponding loss is the following:

<u>Definition 3.6:</u> A control law, denoted by F, is a sequence of functions

$$\{f_t\}_0^\infty, \quad \text{defined by}$$

$$S_{t-1} \times S_t \ni (u_{[0,t-1]}, y_{[0,t]}) \xmapsto{\;f_t\;} u(t) \in R \qquad t \geq 0.$$

The loss, i.e. the minimum time t_m in Definition 3.5, obtained when using a control law F on a system described by H is denoted by $J^a(F,H)$.

□

In adaptive start-up problems there are constraints on u and y. This means that admissible control laws F are of interest.

<u>Definition 3.7:</u> A control law F is admissible (with respect to Σ^a), if for every $H \in \Sigma^a$

$$u(t) = f_t(u_{[0,t-1]}, y_{[0,t]}) \quad \text{where} \quad y_{[0,t]} = H(u_{[0,t-1]})$$

gives u and y fulfilling constraints (3.2).

The set of admissible control laws is denoted by Φ^a. Subsets of Φ^a are defined by

$$\Phi^a_{t,u,y} = \{F\in\Phi^a: u(r) = f_r(u_{[0,r-1]}, y_{[0,r]}) \quad 0\leq r\leq t-1\}.$$

□

Defining optimal control laws

A control law F will be considered optimal if it gives the lowest loss $J^a(F,H)$ for the worst possible H at each time t. The precise meaning of optimality is given by the following definition:

Definition 3.8: An admissible control law F is called optimal if for all $t\geq 0$, $u\in S_{t-1}$, $y\in S_t$, and $G\in\Phi^a$,

$$F,G \in \Phi^a_{t,u,y}$$

implies that

$$\sup_{H\in\Xi^a_{t,u,y}} J^a(F,H) \leq \sup_{H\in\Xi^a_{t,u,y}} J^a(G,H).$$

If the inequality can be strengthened to

$$J^a(F,H) \leq J^a(G,H) \quad \text{for all } H\in\Xi^a$$

F is said to be completely optimal.

□

This is a worst case definition: However, some form of averaging technique is often used in it's stead, for instance,

$$\bar{J}(F) = \sum_{i=1}^{n} J(F,H_i)\ m(H_i), \qquad \Xi^a = \left\{H_1,\ldots,H_n\right\},$$

where m is a semi-positive functional and Ξ^a is supposed to be finite. By using measure theory it may be possible to extend this idea to infinite sets Ξ^a. The technique of using a functional to compare different control laws is common in stochastic control, where $m(H_i)$ is the probability density. $\bar{J}$ is often called Bayes cost (See eg. Sworder 1966). Such cost functions are, however, not appropriate for the start-up problems.

Discussion

Notice the necessity that some $H \in \Xi^a$ actually can describe an obtained output in exact terms. This means that it is very difficult to characterize the set Ξ^a. Trying to list all relevant aspects of a practical system is not realistic. For instance, in the hot-plate case it would mean that the salt concentration of the water should be included as well as the frequency and the voltage of the AC power-net. Even if it is assumed that Ξ^a is known, it would, in most practical cases be impossible to solve the problem. The models would be too complex.

Approximate models is the solution to such problems. They require, however, a modification of the concepts introduced in this section.

3.2 FORMULATION WITH INACCURATE MODELS

The concepts of learning, admissible control law, and optimality will be defined in the case of inaccurate models.

Description of processes

Let Ξ be defined as a set of description maps, $H:S \rightarrow S$, which approximates Ξ^a. Ξ is chosen by the control engineer, thus the functions $H \in \Xi$ are all known.

Example 3.1: The set Ξ^a of maps describing the hot-plate process can be approximated by Model (B.1) in which all parameters are known apart from m_w and T_{po}. Assume the values of m_w and T_{po} to be discretized so that the resulting set Ξ consists of all models of type (B.1) with

$$m_w = [0.5 \text{ kg}, 0.6 \text{ kg}, \ldots\ldots\ldots, 1.9 \text{ kg}, 2.0 \text{ kg}]$$

$$T_{po} = [20°C, 30°C, \ldots\ldots\ldots, 290°C, 300°C].$$

Assume also no losses (α=0 W/°C) and the sampling interval h=10 s. The application of the formulation of learning in Definition 3.4 will, however, not be relevant. The knowledge of y(1) and y(2) would normally be enough to determine m_w and T_{po}. Following this, ordinary optimal control could have been applied. This is, however, no solution to the real problem, not even a good approximation. Consider for instance the case of a small sample interval. In Figure 3.2, pulse responses are shown for the hot-plate process and the models. If an attempt to choose a model is made after two samples the natural choice is 8 m_w = 2.0 kg and T_{po} = 20°C instead of the correct m_w=1 kg and T_{po} = 20°C.

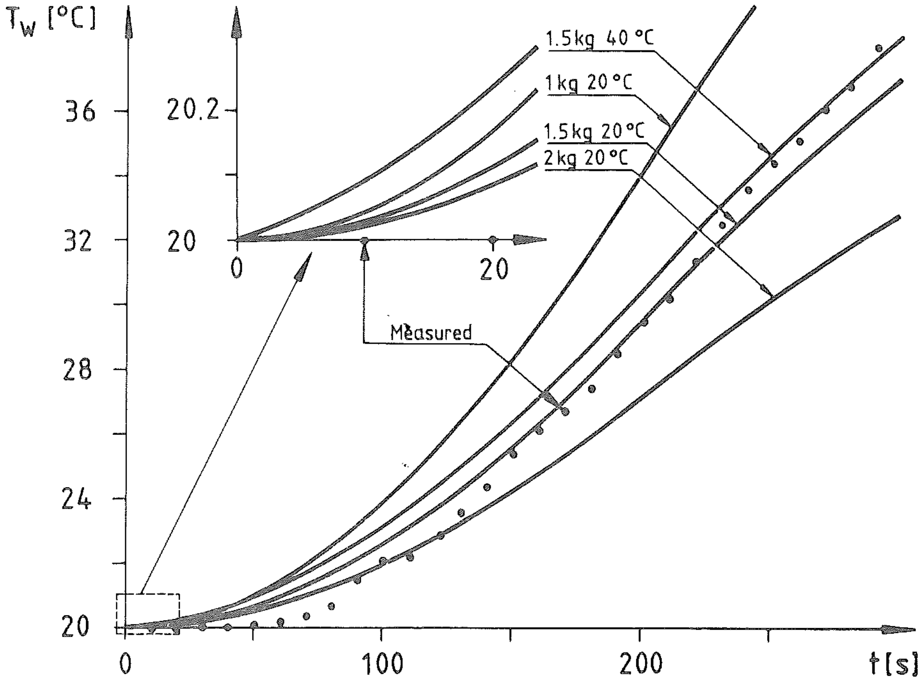

Fig. 3.2: Pulse responses for the hot-plate. The pulse was 200 seconds long and the amplitude was 1350 W. Dots are measured data with m_w = 1 kg and $T_p(0)$ = 20°C. Full lines are simulations of Model (B.1) for different values of m_w and T_{po}.

Description of learning

The drawback of the method of choosing model in Example 3.1 is that it restricts the model class too soon. Definition 3.4 suggests how to avoid the problem, the essential idea being not to decide on the best model as soon as possible but to reject the impossible ones as soon as possible. For instance, in the example above (see Figure 3.2) it can be stated with some confidence that after 300 seconds m_w is not greater than 1.5 kg if it is assumed that T_{po} = 20°C. One way is to reject models if the measured output is too far from the model output. Definition 3.4 can then be modified to describe learning.

Definition 3.9: $\Sigma_{t,u,y}$ is defined as a subset of Σ for which

$$\Sigma_{t,u,y} = \left\{H;\ H\in\Sigma \text{ and } \|H(u_{[0,r-1]}) - y_{[0,r]}\|_{S_r} \leq \varepsilon,\ 0\leq r\leq t\right\}$$

where $\varepsilon=\varepsilon(s,u_{[0,s-1]},H)$ is a sequence of positive real numbers. The subscript on u is sometimes suppressed. The dependence on the norm and on ε for $\Sigma_{t,u,y}$ is suppressed in the notation. The same is often done with u and y. The set $\Sigma_{\infty,u,y}$ is defined by

$$\Sigma_{\infty,u,y} \equiv \bigcap_{t=1}^{\infty} \Sigma_{t,u,y}.$$

□

Remark 3.1: In Section 2.4 it was mentioned that the model error will depend upon both the input and the approximating model. This is elaborated in Chapter 5.

□

Observe that Definition 3.9 is such that Ξ_t is always decreasing. This is in line with the common notion of learning. The norm is of course crucial for the set Ξ_t. For the basic start-up problem the seminorm

$$\|y - z\|_{S_t} = |y(t) - z(t)|$$

is often sufficient. In other cases it can for instance be of interest to consider also the derivative y'(t) or the difference y(t+h)-y(t).

In the sequel it is supposed that $\varepsilon(s;u;H)$ is chosen so that the following assumption holds:

<u>Assumption 3.1:</u> To every member $H^a \in \Xi^a$ there is a member $H \in \Xi$ such that

$$|H^a(u_{[0,t-1]})(t) - H(u_{[0,t-1]})(t)| \leq \varepsilon(t;u_{[0,t-1]};H) \quad (3.3)$$

for all $t \geq 0$ and for any member u of S_∞ such that u and y(t)=H(u)(t) fulfil constraints (3.2).

□

<u>Remark 3.2:</u> Observe that the requirement that u and y fulfil constraints (3.2) is natural. Without such a requirement it would be necessary to specify $\varepsilon(t;u;H)$ in cases having no interest with respect to the basic start-up problem.

□

Remark 3.3: In practice it may be necessary to substitute inequality (3.3) for

$$\varepsilon_-(t,u,H) \le H^a(u_{[0,t-1]})(t)-H(u_{[0,t-1]})(t) \le \varepsilon_+(t,u,H).(3.4)$$

Such a change of Assumption 3.1 does not change the theory of Chapter 3 in any essential way. Therefore the more convenient form (3.3) is used. ε_- and ε_+ are used in Chapters 5 and 7.

□

Admissible control laws

Definition 3.7 of admissible control laws has to be redefined in order to allow for the approximative description of the process.

Definition 3.10: A control law F given by $\{f_t\}_0^\infty$ is admissible if for every $H\in\Xi$ it happens that all sequences $\eta\in S_\infty$ fulfilling the condition

$$|\eta(t) - H(u_{[0,t-1]})(t)| \le \varepsilon(t,u_{[0,t-1]},H)$$

where $u\in S_\infty$ is given by

$$u(t) = f_t(u_{[0,t-1]},\eta_{[0,t]}),$$

give an input sequence u and an output sequence $\eta(=y)$ fulfilling constraints (3.2). The set of admissible control laws is denoted by Φ. Subsets of Φ are defined by

$$\Phi_{t,u,\eta} = \{F: F\in\Phi,\ u(r) = f_r(u_{[0,r-1]},\eta_{[0,r]}),\quad r\le t-1\}.$$

The set $\Phi_{\infty,u,y}$ is defined by

$$\Phi_{\infty,u,y} \equiv \bigcap_{t=1}^{\infty} \Phi_{t,u,y}.$$

□

Remark 3.4: η is used instead of y to emphasize that, for a given sequence $\{\eta(t)\}_0^\infty$, there does not have to be a $H^a \in \Xi^a$ and an input sequence u such that

$$\eta(t+1) = H^a(u)(t+1).$$

□

Remark 3.5: The domain of definition for a control law F can be restricted to all pairs of sequences u and $y \in S_\infty$ such that $F \in \Phi_{t,u,y}$ and $\Xi_{t,u,y}$ is non empty for all t.

□

Example 3.2 : The hot-plate process is studied with Ξ as in Example 3.1. Let y_{ref} = 40°C. Let F be a control law which for the first sample periods starts with u=1350 W. After two samples it determines which $H \in \Xi$ has generated the output, and it uses u=1350 W exactly the time needed for this model H to obtain the desired temperature. This control law is admissible according to Definition 3.7. However, in a practical case there are problems. First of all the measured output y(1) and y(2) may be such that there is no $H \in \Xi$, such that

$$y(1) = H(u_{[0,0]})(1) \quad \text{and} \quad y(2) = H(u_{[0,1]})(2).$$

If this is circumvented by using the closest H in some sense, there are still problems as seen from Figure 3.2. After 20 seconds (two sample intervals) the control law would choose $m_w = 2.0$ kg and $T_{po} = 20°C$ as the best model. This would give nearly twice the required amount of energy and result in a big overshoot violating constraints (3.2). F is therefore not admissible. Definition 3.10 takes into account the possibility of an error in the description.

□

It has easily been demonstrated that the control law in Example 3.2 is not admissible. In a general case, however, the matter is more difficult. Even if Ξ is finite and the time horizon is finite, the possibilities of different η can make it impossible to judge in finite time, except for special classes of feedback laws.

The basic view here is that all that is known about the process is given by old input and output values, the set of models Ξ and Assumption 3.1. Prediction at time t of the future output will only rely upon input u, $\Xi_{t,u,y}$ and Assumption 3.1. The definition of admissible control laws is in line with this view. The concept of inaccurate sets of models will also make it natural to study a special class of control laws.

<u>Definition 3.11:</u> A control law F is said to be reduced to Ξ if the functions f_t only depend on $u_{[0,t-1]}$ and $\Xi_{t,u,y}$, i.e.

$$f_t(u_{[0,t-1]}, y_{[0,t]}) = \varphi_t(u_{[0,t-1]}, \Xi_{t,u,y})$$

where φ_t is a sequence of functions.

□

In Section 3.3 the significance of Definition 3.11 will be made clearer.

Defining optimality

The meaning of optimality for a control law was discussed and defined in Definition 3.8. In the case of inaccurate descriptions the situation is more complicated. It is not possible to define the loss as in Definition 3.6 by simply changing Ξ^a to Ξ. The reason is the same as why the admissible control laws could not be defined by Definition 3.7. Again, the uncertainty in $y(t)$ must be considered. The definition is made in two steps.

Definition 3.12: The loss function $\jmath$ is defined by $\jmath(\eta,F;H)=t_m$ where t_m fulfils

$$\eta(t) \geq y_{ref} - \varepsilon_{ref} \qquad t \geq t_m$$
$$\eta(t_m-1) < y_{ref} - \varepsilon_{ref} \qquad t_m \leq 1$$

and $u,\eta \in S_\infty$, $H \in \Xi_{\infty,u,\eta}$, and $F \in \Phi_{\infty,u,\eta}$.

□

Remark 3.6: In the basic start-up problem the loss $\jmath$ can be expressed as a function of only η. In many other cases, the loss function requires the input or the control law as an argument. In order to give a characterization of possible η's through Assumption 3.1, an explicit dependence on $H \in \Xi$ is preferred in the notation $\jmath(\eta,F;H)$.

□

Definition 3.13: For $t \geq 0$, $u \in S_\infty$, $\eta \in S_t$, $H \in \Xi_{t,u,\eta}$, and $F \in \Phi_{t,u,\eta}$ there is associated a loss, $J(\eta_{[0,t]},F,H)$, defined by

$$J(\eta_{[0,t]},F,H) =$$

$$\sup_{\xi\in S_\infty} \{j(\xi,F,H): \xi_{[0,t]}=\eta_{[0,t]},\ H\in\Sigma_{\infty,u,\xi},\ F\in\Phi_{\infty,u,\xi}\}.$$

□

The analogy of Definition 3.8 is given in the next definition.

Definition 3.14: An admissible control law F is called optimal if for all $t\geq 0$, $u\in S_{t-1}$, $\eta\in S_t$ and $G\in\Phi$

$$F,G\in\Phi_{t,u,\eta}$$

implies that

$$\sup_{H\in\Sigma_{t,u,\eta}} J(\eta_{[0,t]},F,H) \leq \sup_{H\in\Sigma_{t,u,\eta}} J(\eta_{[0,t]},G,H).$$

If the inequality above can be strengthened to

$$J(\eta_{[0,t]},F,H) \leq J(\eta_{[0,t]},G,H),\quad H\in\Sigma_{t,u,\eta},$$

then F is said to be completely optimal.

□

Discussion

It is now possible to relate to the dual character of adaptive optimal control as introduced by Feldbaum (1965). The ability to control so that the output reaches the desired value requires a fairly good knowledge of the system in operation. In current terminology this means that Σ_t must be rather small before the goal is obtained. In this way it can be said that the controller tries to meet two objectives. The first one is making J small, the second one is making Ξ_t small. The two objectives are normally in conflict with each other, but not always, as will be seen in Section 4.1.

3.3 OPTIMALITY OF REDUCED CONTROL LAWS IN THE CASE OF THE BASIC START-UP PROBLEM.

In the case of the basic start-up problem, the output given as $y_{[0,t]}$ seems to be far more precise information than necessary if the objective is to use an optimal control law as defined by Definition 3.14. In Theorem 3.2, it will be proved that if F is an optimal control law and Ξ is a finite set, it is possible to change F to F^*, such that F^* is optimal and reduced to Ξ (see Definition 3.11). In the proof a successive modification of F is used. Reference outputs η are chosen one at a time and F is modified for all outputs, starting close to η. These reference outputs are chosen in special relation to possible models H. The modification for such a specific output η is defined in Definition 3.15 and characterized in Lemma 3.1 and Theorem 3.1. Theorem 3.1 underlines the meaning of Definition 3.11 that the subset Ξ_t is the essential concept. The result is interesting in its own, and not only as a preparation for the proof of Theorem 3.2.

As an introduction to Lemma 3.1 and the theorems, the following discussion is undertaken. Let $u_{[0,s-1]}$ and $\eta_{[0,s]}$ be such that $F \in \Phi_{s,u,\eta}$, and $\Xi_{s,u,\eta}$ is non-empty. Let ζ be such that

$$\Xi_{r,u,\zeta} = \Xi_{r,u,\eta} \qquad 0 \leq r \leq q. \tag{3.5}$$

Equation (3.5) means that for any time instant $r \in [0,q]$, $\zeta(r)$ could have been obtained instead of $\eta(r)$ when F is used on η. Nothing is said about the possibility of obtaining ζ when using F on ζ. It *may* very well happen that F is not defined on the whole segment $\zeta_{[0,q]}$.

If the control law F is disregarded for the moment, it follows that the predictions of further outputs $y(t)$, $t>s$, given that $(u_{[0,s-1]},\eta_{[0,s]})$ were known, are the same as if $(u_{[0,s-1]},\zeta_{[0,s]})$ were known. Therefore it should be possible to alter F so that the control signal is the same on $(u_{[0,s-1]},\zeta_{[0,s]})$ as on $(u_{[0,s-1]},\eta_{[0,s]})$, and that the modified F is optimal if F is optimal. The modified F will be called F'.

When F is changed to F' and a different input is generated, totally different outputs may be obtained. Therefore F' has to be defined for all such outputs ζ and not only for those close to the reference output η.

Definition 3.15: Regard a pair $u,\eta \in S_\infty$, such that $\Xi_{\infty,u,\eta}$ is non-empty, and a control law F, such that

$$F \in \Phi_{\infty,u,\eta}.$$

Define F' by

$$u'(r)=f'_r(u'_{[0,r-1]},\zeta_{[0,r]})=f_r(u'_{[0,r-1]},\eta'_{[0,r]}) \qquad r \geq 0 \tag{3.6}$$

for all possible outputs ζ, i.e. $\Sigma_{\infty,u',\zeta}$ is non-empty. The sequence η' is obtained by concatenation

$$\eta' = (\eta_{[0,q]}, \zeta_{[q+1,\infty)}) \tag{3.7}$$

using q fulfilling

$$\left.\begin{array}{ll} \Sigma_{r,u,\zeta} = \Sigma_{r,u,\eta} & , \quad r \leq q \\ \Sigma_{q+1,u,\zeta} \neq \Sigma_{q+1,u,\eta} & \end{array}\right\} \tag{3.8}$$

For ζ not starting close to η, i.e. $q=-1$, F' is equal to F.

□

<u>Remark 3.7:</u> In the definition it is assumed that η' is a possible output when using F, so that the right hand side of (3.6) is defined. In Lemma 3.1 it is proved that this is the case. Note that $u'_{[0,q]} = u_{[0,q]}$.

□

<u>Lemma 3.1:</u> The following assertions are true for an arbitrary η and ζ such that $\Sigma_{\infty,u',\zeta}$ is non-empty:

$$\left.\begin{array}{lll} \text{(i)} & \Sigma_{r,u',\zeta} = \Sigma_{r,u',\eta'} & r \geq 0 \\ \text{(ii)} & f'_r \text{ is defined} & r \geq 0 \end{array}\right\} \tag{3.9}$$

(iii) $F' \in \Phi$

(iv) $J(\zeta_{[0,n]}, F', H) \leq \max\ [n+1, J(\eta'_{[0,n]}, F, H)]$

$H \in \Sigma_{n,u',\eta'}$

where η', u' and F' are defined in Definition 3.15.

<u>Proof:</u>

(i) and (ii): For $r \leq q$ (3.9) follows directly from (3.6) - (3.8). Assume that (3.9) is true for $r \leq t$. But $\zeta(t+1) = \eta'(t+1)$ together with the crucial property:

$$H \in \Xi_{t+1,u',\zeta} \quad \Leftrightarrow$$

$$\left\{ H \in \Xi_{t,u',\zeta} \;\; \& \;\; |\zeta(t+1) - H(u'_{[0,t]})(t+1)| \leq \varepsilon(t+1, u'_{[0,t]}, H) \right\}$$

imply that

$$\Sigma_{t+1,u',\eta} = \Sigma_{t+1,u',\zeta}$$

which is non-empty. Therefore $\eta'_{[0,t+1]}$ is a possible output and $F \in \Phi$ proves that $u'(t+1)$ and f'_{t+1} are well defined. Induction in t proves (3.9).

(iii): But an output may take off from η to $\zeta(t+1)$ also earlier than after $t > q$, (3.8). Therefore $F \in \Phi$ implies that u' and ζ fulfil constraints (3.2) for all $t = 0$, and thus $F' \in \Phi$, since ζ is arbitrary.

(iv): Now regard

$$p = J(\zeta_{[0,n]}, F', H) = \tag{3.10}$$

$$= \sup_{\xi_1 \in S_\infty} \left\{ J(\xi_1, F'; H) : \xi_{1[0,n]} = \zeta_{[0,n]},\; H \in \Xi_{\infty,u',\xi_1},\; F' \in \Phi_{\infty,u',\xi_1} \right\}$$

If $p > n+1$, choose a ξ_1, such that

$$\xi_1(p-1) < y_{ref} - \varepsilon_{ref},$$

for instance a supremal ξ_1 in (3.10), and consider this as ζ in Definition 3.15. Compare with

$$J(\eta'_{[0,n]},F,H) =$$

$$\sup_{\xi_2 \in S_\infty} \left\{ J(\xi_2,F,H) : \xi_{2[0,n]} = \eta'_{[0,n]},\ H \in \Xi_{\infty,u',\xi_2},\ F \in \Phi_{\infty,u',\xi_2} \right\}.$$

Therefore, choose a probably not supremal ξ_2, such that

$$\xi_2 = (\eta'_{[0,n]},\ \xi_{1[n+1,\infty]})$$

when $p-1>q$. In case $p-1 \leq q$ choose an output ξ_2 fulfilling

$$\xi_2(t) = \begin{cases} \eta(t) & t < p-1 \\ \xi_1(t) & t = p-1 \end{cases}$$

but unspecified for $t>p-1$. This is possible because $p>n+1$. Therefore, because $p>n+1$,

$$\xi_2(p-1) = \xi_1(p-1) < y_{ref} - \varepsilon_{ref} \text{ and } J(\xi_2,F,H) \geq p$$

so that (iv) is proven.

□

To prove optimality of F', it is first of all necessary to compare F and F' when a deviation $\zeta(t+1)$ from $\eta_{[0,t]}$ is experienced. Unfortunately it is not sufficient to compare only F and F'. The output $\zeta(t+2)$, obtained when using F', may be an impossible output when using F; therefore, F' has to be compared with all possible control laws G as in Definition 3.14 of optimality.

<u>Theorem 3.1:</u> If F is optimal then F' is also optimal.

<u>Proof:</u> Regard any $G \in \Phi$, $\zeta \in S_n$ and $u' \in S_{n-1}$ such that $\Xi_{n,u',\zeta}$ is non-empty and $F', G \in \Phi_{n,u',\zeta}$. It will be proved that

$$p = \sup_{H \in \Xi_{n,u',\zeta}} J(\zeta_{[0,n]}, F', H) \leq \sup_{H \in \Xi_{n,u',\zeta}} J(\zeta_{[0,n]}, G, H). \quad (3.11)$$

Define G" as F' was defined in Definition 3.15 by replacing F and η with G and ζ. The control signal u' is actually obtained when using G" on η' because Lemma 3.1 gave

$$\Xi_{r,u',\eta'} = \Xi_{r,u',\zeta} \qquad r \leq n$$

and G" is then defined to give what G gives on ζ. It also means that $\eta'_{[0,n]}$ may be obtained when using G".

Assume $p > n+1$. It then follows from Lemma 3.1 that

$$\sup_{H \in \Xi_{n,u',\zeta}} J(\zeta_{[0,n]}, F', H) \leq \sup_{H \in \Xi_{n,u',\eta}} J(\eta'_{[0,n]}, F, H).$$

F optimal gives, since G" can give $\eta'_{[0,n]}$, that

$$\sup_{H \in \Xi_{n,u',\eta'}} J(\eta'_{[0,n]}, F, H) \leq \sup_{H \in \Xi_{n,u'\eta'}} J(\eta'_{[0,n]}, G", H).$$

Further Lemma 3.1 can be modified to characterize G", so that

$$\sup_{H \in \Xi_{n,u',\eta'}} J(\eta'_{[0,n]}, G", H) \leq \sup_{H \in \Xi_{n,u',\zeta}} J(\zeta_{[0,n]}, G, H).$$

Thus if $p > n+1$ then (3.11) is proved. If $p \leq n+1$ then $\zeta(p-1)(y_{ref} - \varepsilon_{ref})$. Thus (3.11) is true also in this case.

Remark 3.8: F' is obtained by redefining F for all ζ fulfilling (3.8) for $q \geq 0$. When working with control laws, it can be of interest to not change F for so many outputs ζ. For instance, assume that it is only of interest to change F up to and including $t=s$, s fixed. In such a case, it is possible to change Definition 3.15 by replacing ∞ with s. Observe that this also means that $u, \eta \in S$ are replaced by $u_{[0,s-1]} \in S_{s-1}$ and $\eta_{[0,s]} \in S_s$. Further (3.8) is replaced with

$$\left.\begin{array}{ll} \Sigma_{r,u,\zeta} = \Sigma_{r,u,\eta} & r \leq q \leq s \\ q < s \Rightarrow \Sigma_{q+1,u,\zeta} \neq \Sigma_{q+1,u,\eta}. & \end{array}\right\} \tag{3.12}$$

Changing the definition of F' will not alter Lemma 3.1, Theorem 3.1 or their proofs. Observe that it is not possible to change F on an arbitrary subset of all ζ fulfilling (3.8) with $q \geq 0$. For instance, redefining F only for those ζ that have q in (3.12) equal to a s for a fixed s, will in general be impossible because at time $t<s$ it is impossible to know if $q=s$ even if $\Sigma_{r,u,\zeta} = \Sigma_{r,u,\eta}$ for $r \leq t$.

□

Theorem 3.2: If F is optimal and Σ is finite, it is then possible to change F to F^* such that F^* is optimal and reduced to Σ.

Proof: Assume that F is changed to F^p which is optimal and such that there is a φ_r such that

$$f^p_r(u_{[0,r-1]}, \eta_{[0,r]}) = \varphi_r(u_{[0,r-1]}, \Sigma_{r,u,\eta}) \quad 0 \leq r \leq p. \tag{3.13}$$

Let $v_{[0,p]} \in S_p$ be such that $F^p \in \Phi_{p,v,\eta}$ for some η. It follows from (3.13) that there are finitely many such $v_{[0,p]}$. Further, let $\{M_i\}_{i=1}^m$ be an enumeration of the power set of Ξ, where m is finite because Ξ is a finite set. For each M_i an $\eta_i \in S_{p+1}$ is chosen, if possible, such that

$$\Xi_{p+1,v,\eta_j} \neq \Xi_{p+1,v,\eta_i} = M_i \qquad 1 \leq j \leq i-1.$$

Define $F^{p,1}$ as F' in the variation of Definition 3.15 given in Remark 3.8 when $\eta_{[0,p+1]} = \eta_1$, $u_{[0,p]} = v_{[0,p]}$, and $F = F^p$. (η_1 is supposed to exist. Otherwise use the least k such that there is an η_k such that $\Xi_{r,u,\eta k}$ is non-empty.)

$F^{p,1}$ is optimal according to Theorem 3.1. Repeating this for all possible η_i gives a control law F_v^{p+1} which is optimal. Repeating this procedure for all possible $v_{[0,p]}$ gives a control law F^{p+1} which is also optimal and such that (3.13) is true when p is changed to p+1. Observe that the construction is such that F^p and F^{p+1} are the same up to and including t=p, i.e

$$f_r^p(u_{[0,r-1]},\eta_{[0,r]}) = f_r^{p+1}(u_{[0,r-1]},\eta_{[0,r]}) \quad 0 \leq r \leq p.$$

Let F^* be defined by

$$f_r^*(u_{[0,r-1]},\eta_{[0,r]}) = f_r^p(u_{[0,r-1]},\eta_{[0,r]}) \qquad 0 \leq r \leq p.$$

From $F^p = F$, p=-1, it can be concluded by induction that F^* is optimal and reduced to Ξ.

□

CHAPTER 4

THEORETICAL ANALYSIS OF SPECIAL START-UP PROBLEMS

The basic start-up problem similar to that of the hot-plate process is solved. It is proved that the control law, taking u(t) as high as possible at each time t, is completely optimal. However, in Sections 4.1 and 4.2 ε is assumed to be constant, which is not possible to assume in a practical case.

In Section 4.3 dynamic programming is used to show the existence of an optimal control law in the general case of the basic start-up problem. However, the use of dynamic programming to solve practical problems is, in general, not possible due to the amount of calculations required.

4.1 MONOTONE SETS OF SYSTEMS

It is normally difficult to derive an optimal solution for the basic start-up problem. For a class of problems the dual objectives, obtaining maximal decrease of Ξ_t and minimal loss J, coincide. With respect to the hot-plate problem, it is natural that the higher the power input the faster the temperature rise. At the same time it seems natural that it is easier to estimate m_w when the signals are large. In this and the next section it will be proved that under special assumptions, intuition is actually in agreement with Definition 3.14 of optimality. The method of proving this is not a straightforward approach, but rather starts off in a slightly different direction.

Definition 4.1: Σ is called a monotone set if

$$
\left.
\begin{array}{ll}
\text{(i)} & \Sigma \text{ can be parametrized with one parameter} \\
 & R \supset I \ni \alpha \xrightarrow{\varphi} H_\alpha \in \Sigma \\
 & \text{(I is an index set of real numbers)} \\
\text{(ii)} & \text{all members } H_\alpha \in \Sigma \text{ can be described as} \\
 & y(0) = 0 \\
 & y(t) = \sum_{i=0}^{t-1} h_\alpha(t-i)\, u(i) \qquad t \geq 1 \\
\text{(iii)} & h_\alpha(1) \geq 0 \\
\text{(iv)} & h_\alpha(t) \text{ is increasing in } t \\
\text{(v)} & h_\alpha(t) \leq h_\alpha(\infty) < \infty \qquad t \geq 1 \\
\text{(vi)} & \alpha_2 > \alpha_1 \text{ implies that } h_{\alpha 2}(t) - h_{\alpha 1}(t) \geq 0 \\
 & \text{and increasing in } t \text{ for all } t \geq 1.
\end{array}
\right\} \quad (4.1)
$$

□

Remark 4.1: The conditions of a monotone set imply that the models are stable but contain an integrator. In the case of the hot-plate no losses can be included.

□

Example 4.1: Consider Model (B.1) of the hot plate in the case of no losses, i.e. $\alpha=0$, and u replaced by $\gamma \cdot u$. The gain γ is the only variable assumed not to be known. Further, it is assumed that

$$T_w(0) = T_p(0) = T_{room} = 0°C.$$

The input-output relation is given by

$$T_w(t) = \int_0^t \frac{1}{c_w + c_p} (1 - e^{-a(t-\tau)}) \gamma u(\tau) d\tau$$

where $a = \Lambda(\frac{1}{c_w} + \frac{1}{c_p})$, $c_w = m_w C_w$, $c_p = m_p C_p$.

Assume $u(\tau)$ is constant during sampling intervals of length Δ (the notion h for the sampling interval is for the moment not used). The impulse response $h(k)$ is thus given by

$$h(k) = \int_0^\Delta \frac{1}{c_w + c_p} (1 - e^{-a(k\Delta-\tau)}) \gamma d\tau \qquad k \geq 1. \tag{4.2}$$

It is easy to see that conditions (4.1.i - vi) are fulfilled, so Ξ is a monotone set.

□

For monotone sets the basic start-up problem (see Definition 3.5) has an easy solution when ε is constant. This will be proved in Theorem 4.1 below. The following two lemmas are used in the proof.

<u>Lemma 4.1:</u> If

(i) $\sum_{i=0}^{j} w(i) \geq 0$ for all $j \leq p$

(ii) $h(i) \geq 0$ $\quad i \geq 0$

(iii) h is increasing

then $$\sum_{i=0}^{j} h(j-i)w(i) \geq 0 \text{ for all } j \leq p.$$

Proof: The result follows from

$$\sum_{i=0}^{j} h(j-i)w(i) = \sum_{k=0}^{j-1} [h(j-k)-h(j-k-1)] \sum_{i=0}^{k} w(i)$$

$$+ h(0) \sum_{i=0}^{j} w(i) \geq 0.$$

□

Lemma 4.2: For a monotone set Ξ, the conditions

(i) $\alpha_2 > \alpha_1$

(ii) $\sum_{i=0}^{t} v(i) \leq \sum_{i=0}^{t} u(i)$ all $t \leq p$

imply that

$$H_{\alpha 2}(u_{[0,p]})(p+1) - H_{\alpha 1}(u_{[0,p]})(p+1) \geq$$

$$\geq H_{\alpha 2}(v_{[0,p]})(p+1) - H_{\alpha 1}(v_{[0,p]})(p+1).$$

Proof:

$$H_{\alpha 2}(u_{[0,p]})(p+1) - H_{\alpha 1}(u_{[0,p]})(p+1) -$$

$$- \left\{H_{\alpha 2}(v_{[0,p]})(p+1) - H_{\alpha 1}(v_{[0,p]})(p+1)\right\} =$$

$$= \sum_{i=0}^{p} [h_{\alpha 2}(p+1-i)-h_{\alpha 1}(p+1-i)][u(i)-v(i)] \geq 0.$$

The last inequality follows from (4.1.vi) by Lemma 4.1.

□

<u>Theorem 4.1:</u> For a monotone set Ξ with ε constant there exists a completely optimal control law F for the basic start-up problem. F is defined recursively for y such that $\Xi_{t,u,y}$ is non-empty. Let $u(t)=f_t(u_{[0,t-1]}, y_{[0,t]})$, be the supremum of all

$u \in [0, u_{max}]$ such that $\bar{u}^t \in S$ defined by

$$\left.\begin{array}{l} \text{(i)} \quad \bar{u}^t(s) = \begin{cases} u_{[0,t-1]}(s) & s < t \\ u & s = t \\ 0 & s > t \end{cases} \\ \\ \text{fulfils} \\ \\ \text{(ii)} \quad H(\bar{u}^t)(s) + \varepsilon \leq y_{ref} + \varepsilon_{ref}, \; H \in \Xi_t, \; s \geq 0. \end{array}\right\} \qquad (4.3)$$

<u>Proof:</u> First of all it will be proved that F is admissible. According to the assumptions about y, there is an $H \in \Xi$ such that

$$y(t+1) \leq |y(t+1) - H(u_{[0,t]})(t+1)| + H(u_{[0,t]})(t+1) \leq$$
$$\leq \varepsilon + H(u_{[0,t]})(t+1) \leq y_{ref} + \varepsilon_{ref}$$

and it is true for any such $H \in \Xi_{t+1} \subset \Xi_t$. The last inequality follows from the equality:

$$u_{[0,t]} = \bar{u}^{t}_{[0,t]}$$

Constraints (3.2) are thus fulfilled for any t, and by Definition 3.10 it follows that F is admissible.

Let F be the proposed optimal control law and G be another control law. Let u and v represent the input, and y and z represent the output given by the control laws F and G, i.e. $F \in \Phi_{t,u,y}$ and $G \in \Phi_{t,v,z}$ for all t. Let r be arbitrarily chosen and assume that $G \in \Phi_{r,u,y}$ and $H_{\alpha 0} \in \Sigma_{r,u,y}$. If there is a z such that

$$z_{[0,r]} = y_{[0,r]}$$

$$H_{\alpha 0} \in \Sigma_{t,v,z} \qquad t \geq 0 \tag{4.4}$$

$$J(y,F;H_{\alpha 0}) \leq J(z,G;H_{\alpha 0}) \tag{4.5}$$

then F is completely optimal according to Definition 3.14.

It will be shown that

$$\left.\begin{aligned} & z_{[0,r]} = y_{[0,r]} \\ & z(t) = H_{\alpha 0}(v_{[0,t-1]})(t) + \\ & \qquad + y(t) - H_{\alpha 0}(u_{[0,t-1]})(t) \qquad t > r \end{aligned}\right\} \tag{4.6}$$

is such a z. If $y(t) \geq z(t)$, the time to reach the final limit in the basic start-up problem (Definition 3.5) would be shorter for y than for z, i.e. (4.5). This would follow from condition

$$\sum_{i=0}^{t-1} v(i) \leq \sum_{i=0}^{t-1} u(i) \qquad \text{for all } t \geq 1 \tag{4.7}$$

using Lemma 4.1 and

$$H_{\alpha o}(u_{[0,s]})(t) - H_{\alpha o}(v_{[0,s]})(t)$$

$$= \sum_{i=0}^{t-1} h(t-i)(u(i)-v(i)) \geq 0 \qquad t \leq s+1.$$

Condition (4.7) will be proved by induction in time together with condition (4.8),

$$\Xi_{t,u,y} \subset \Xi_{t,v,z} \qquad \text{for all } t \geq 0. \tag{4.8}$$

Assume (4.7) and (4.8) are true for $t \leq p$, $p \geq \max(0,r)$. When $p=\max(0,r)$ it follows immediately that (4.7) and (4.8) are true. First it is observed that $H(\bar{u}^{p})(s)$, $H \in \Xi$, is increasing in s and bounded. This means that restriction (4.3.ii) can be simplified to

$$\left.\begin{aligned} &\lim_{s\to\infty} \sum_{i=0}^{p} h(s-i)u(i) = h(\infty) \sum_{i=0}^{p} u(i) \leq y_{ref} + \varepsilon_{ref} - \varepsilon \\ &H \in \Xi_{p,u,y}. \end{aligned}\right\} \tag{4.9}$$

Similarly G, being an admissible control law, implies that

$$\left.\begin{array}{l} h(\infty) \sum_{i=0}^{p} v(i) \leq y_{ref} + \varepsilon_{ref} - \varepsilon \\ \\ H \in \Sigma_{p,v,z}. \end{array}\right\} \qquad (4.10)$$

If the inequality (4.9) is not the bounding factor for $u(p)$, then $u(p) = u_{max}$. This, $v(p) \leq u_{max}$, and

$$\sum_{i=0}^{p-1} v(i) \leq \sum_{i=0}^{p-1} u(i)$$

imply that (4.7) is true for $t=p+1$ in the case $u(p) = u_{max}$.

If the inequality in (4.9) is the bounding factor on $u(p)$, then it follows from (4.9), (4.10), and the inclusion

$$\Sigma_{p,u,y} \subset \Sigma_{p,v,z}$$

that

$$\sum_{i=0}^{p} v(i) \leq \sum_{i=0}^{p} u(i)$$

because $u(p)$ is chosen as large as possibly over the smaller set of restricting $h(\infty)$-values.

Now turn to (4.8) for $t=p+1$. Regard an arbitrary H such that

$$H \in \Sigma_{p+1,u,y} \qquad (4.11)$$

or equivalently

$$\left.\begin{array}{ll}\text{(i)} & H \in \Xi_{p,u,y} \\ \text{(ii)} & |H(u_{[0,p]})(p+1) - y_{[0,p+1]}(p+1)| \leq \varepsilon.\end{array}\right\} \qquad (4.12)$$

To prove (4.8) is equivalent to prove that

$$\left.\begin{array}{ll}\text{(i)} & H \in \Xi_{p,v,z} \\ \text{(ii)} & |H(v_{[0,p]})(p+1) - z_{[0,p+1]}(p+1)| \leq \varepsilon.\end{array}\right\} \qquad (4.13)$$

The condition (4.13.i) follows from (4.8) for $t=p$. To prove (4.13.ii), introduce α as the parameter of the H chosen in (4.11). Further, introduce

$$A = H_{\alpha}(v_{[0,p]})(p+1) - H_{\alpha 0}(v_{[0,p]})(p+1)$$
$$B = H_{\alpha 0}(u_{[0,p]})(p+1) - y(p+1)$$
$$C = H_{\alpha}(u_{[0,p]})(p+1) - y(p+1).$$

First, observe that the left hand side of (4.13.ii) can be rewritten using the definition (4.6) of z, as

$$H_{\alpha}(v_{[0,p]})(p+1) - z_{[0,p+1]}(p+1) = A + B,$$

and from $H_{\alpha 0} \in \Xi_{p+1,u,y}$ and $H_{\alpha} \in \Xi_{p+1,u,y}$ it follows that

$$|B| \leq \varepsilon \quad \text{and} \quad |C| \leq \varepsilon.$$

Now assume that $\alpha > \alpha 0$. It then follows from the monotony of Ξ that $A \geq 0$. Further, (4.7) for $t=p+1$ gives that $A+B-C \leq 0$ using Lemma 4.2. Hence,

$$A + B \geq -\varepsilon \quad \text{and} \quad A + B \leq C \leq \varepsilon$$

so that (4.13.ii) follows. The case $\alpha<\alpha o$ is proved in the same way. This completes the proof of (4.7) and (4.8) for $t\leq p+1$, and induction gives the theorem.

□

<u>Remark 4.2:</u> If Σ is a monotone set with

$$h_{\alpha}(t) = \alpha \qquad t > 0 \qquad \alpha \geq 0,$$

then the completely optimal control law gives $u(t)=0$ for all t. The controller does not dare to use a positive u because of the risk of an overshoot already in the very first step.

□

In the proof of Theorem 4.1 it is taken only that condition (4.1.vi) in Definition 4.1 is satisfied up to

$$t = \sup_{H\in\Sigma} J(\eta_{[0,-1]},F,H).$$

<u>Corollary 4.1:</u> Let F be the control law defined in Theorem 4.1, and let Σ be a monotone set excepting that condition (4.1.vi) is only fulfilled up to

$$t = \sup_{H\in\Sigma} J(\eta_{[0,-1]},F,H).$$

Then F is completely optimal.

□

The use of Corollary 4.1 requires an estimate of

$$\sup_{H \in \Xi} J(\eta_{[0,-1]}, F, H).$$

This can be obtained by simulations letting the output z be

$$z(t) = H(v_{[0,t-1]})(t) + \varepsilon.$$

In such a case it is extremely difficult to reject $H_\alpha \in \Xi$ for large α. This makes it necessary to use a low input v. The result is

$$j(z,F;H) = J(\eta_{[0,-1]}, F, H).$$

This is proved in Theorem 4.2. Thus, by making simulations for all $H \in \Xi$,

$$\sup_{H \in \Xi} J(\eta_{[0,-1]}, F, H)$$

can be determined.

<u>Theorem 4.2:</u> Consider the basic start-up problem. Let F be the control law defined in Theorem 4.1. Let Ξ be a monotone set except that condition (4.1.vi) is only fulfilled up to $t = j(z,F;H)$, where z is given by

$$z(t) = H(v_{[0,t-1]})(t) + \varepsilon$$

and v is such that $F \in \Phi_{t,v,z}$, $t \geq 0$. Then

$$J(\eta_{[0,-1]}, F, H) = j(z,F;H).$$

<u>Proof:</u> For some α_0, $H=H_{\alpha 0}\in\Sigma$. Let η and u be such that $H\in\Sigma_{t,u,\eta}$ and $F\in\Phi_{t,u,\eta}$.

It will be proved that for all $t < j(z,F;H)$,

$$\left.\begin{array}{ll} \text{(i)} & \displaystyle\sum_{s=0}^{t-1} v(s) \le \sum_{s=0}^{t-1} u(s), \\[2ex] \text{(ii)} & \left.\begin{array}{l} H_\alpha \in \Sigma_{t,u,\eta} \\[1ex] \alpha \ge \alpha_0 \end{array}\right\} \rightarrow H_\alpha \in \Sigma_{t,v,z}. \end{array}\right\} \quad (4.14)$$

For $t=0$ (4.14) is true. Assume (4.14) is true for $t\le p$. (4.14.i) is proved for $t=p+1$ similarly to how (4.7) was proved in the proof of Theorem 4.1.

Assume $H_\alpha\in\Sigma_{p,u,\eta}$ for $\alpha\ge\alpha_0$. It then follows from (4.1.vi) that

$$z(p+1) - H_\alpha(v_{[0,p]})(p+1) \le z(p+1) - H(v_{[0,p]})(p+1) = \varepsilon.$$

Further, Lemma 4.2 gives that

$$H_\alpha(v_{[0,p]})(p+1)-z(p+1) = H_\alpha(v_{[0,p]})(p+1)-H(v_{[0,p]})(p+1)-\varepsilon \le$$

$$\le H_\alpha(u_{[0,p]})(p+1)-\eta(p+1)+\eta(p+1)-H(u_{[0,p]})(p+1)-\varepsilon \le \varepsilon.$$

Thus (4.14.ii) is true for $t=p+1$.

Let s be the least number such that

$$H(v_{[0,s-1]})(s) \ge y_{ref}-\varepsilon_{ref}+\varepsilon.$$

Then

$$\eta(s) \geq -\varepsilon+H(u_{[0,s-1]})(s) \geq -\varepsilon+H(v_{[0,s-1]})(s) \geq y_{ref}-\varepsilon_{ref}$$

where the second inequality follows from Lemma 4.1 and (4.14.i). Thus s is the least upper limit of $J(\eta_{[0,-1]},F,H)$.

□

The condition (4.1.vi) in Definition 4.1 is essential. This will be shown in the following example where Σ is fulfilling all requirements in Definition 4.1 except (vi). It is shown that it is not optimal to take $u(t)$ as large as possible at each time t.

<u>Example: 4.2:</u> Set

$$h(t) = \begin{cases} 1 & t = 1 \\ 2 & t \geq 2 \end{cases}$$

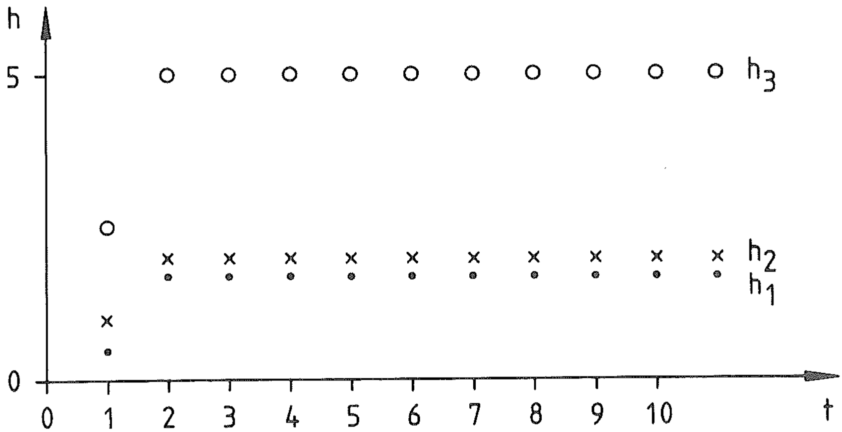

Fig. 4.1 The three impulse responses defining H_1, H_2 and H_3, when $\theta = 1/4$ and $q = 5/2$.

$$\delta(t) = \begin{cases} 2 & t = 1 \\ 1 & t \geq 2 \end{cases}$$

$$h_1(t) = h(t) - \theta \cdot \delta(t)$$
$$h_2(t) = h(t)$$
$$h_3(t) = q \cdot h(t)$$

where θ fulfils $0 < \theta \leq 1/4$. The basic start-up problem with

$$\Sigma = \{H_1, H_2, H_3\}$$

where $$H(u_{[0,t-1]}) = \sum_{i=0}^{t-1} h_j(t-i)u(i) \qquad j = 1, 2, 3$$

is studied for a given $\varepsilon > 0$. The parameters q, y_{ref}, and ε_{ref} are chosen such that (4.15) below is fulfilled.

Let $x = y_{ref} + \varepsilon_{ref} - \varepsilon$.

$$\left.\begin{array}{ll} \text{(i)} & q > 2 \\[2ex] \text{(ii)} & \theta x(1 - \frac{1}{2q}) < 2\varepsilon < \theta x(1 - \frac{1}{4q}) \\[2ex] \text{(iii)} & 0 < 4(\varepsilon_{ref} - \varepsilon) < \theta \cdot x. \end{array}\right\} \qquad (4.15)$$

From $\theta \leq 1/4$ and (4.15.i,ii) it follows that

$$2\varepsilon < \theta x(1 - \frac{1}{4q}) \leq \frac{x}{2}(1 - \frac{1}{q}). \qquad (4.16)$$

This inequality will be used later.

Observe that θ is such that Σ fulfils all requirements in

Definition 4.1 except for (vi). u_{max} is assumed to be so large that it does not limit u. The control law F in Theorem 4.1, i.e. taking u(t) as large as possible at each time t, will be examined. It is assumed that the process is described by H_1. Thus

$$u(0) = x/h_3(\infty) = x/(2q)$$

and

$$[h_3(1) - h_1(1)]\ u(0) \geq [h_3(1) - h_2(1)]\ u(0) =$$

$$= 2(q-1)x/2q > 2\varepsilon$$

where the last inequality follows from (4.16). H_3 can thus always be rejected at time t=1, but not always H_2, since from (4.15 i and ii)

$$[h_2(1) - h_1(1)]u(0) = \theta \cdot 2u(0) = \theta\ \frac{2x}{2q} < 2\varepsilon.$$

The control law may therefore give

$$u(1) = x/h(\infty) - u(0) = x/2 - x/(2q).$$

The difference between the outputs from H_2 and H_1 is

$$H_2(u_{[0,t-1]})(t) - H_1(u_{[0,t-1]})(t) \leq \theta[u(0) + 2u(1)] =$$

$$= \theta(\frac{x}{2q} + x - \frac{x}{q}) = \theta x(1 - \frac{1}{2q}) < 2\varepsilon.$$

As long as H_2 cannot be rejected, u(t)=0, t≥2, which proves the first inequality. The last inequality follows from

(4.15.ii). From

$$H_1(u_{[0,t-1]})(t) \le h_1(\infty)[u(0)+u(1)] =$$

$$= [2-\theta]\cdot x/2 = x - \theta\cdot x/2 < y_{ref} + \varepsilon_{ref} - \varepsilon - 2(\varepsilon_{ref} - \varepsilon) =$$

$$= y_{ref} - \varepsilon_{ref} + \varepsilon$$

it follows that

$$\sup_{H\in\mathbb{H}} J(\eta_{[0,0]}, F, H) = \infty.$$

Another control law G is defined by choosing u(0) as x/(4q) and at t≥1 the input u(t) is chosen as large as possible. Also with the control law G it will always be possible to reject H_3 at t=1 because

$$[h_3(1) - h_1(1)]\, u(0) \ge [h_3(1) - h_2(1)]\, u(0) =$$

$$= 2(q-1)x/4q > 2\varepsilon$$

according to (4.16). So u(1) can at least be chosen as

$$u(1) = x/h(\infty) - u(0) = x/2 - x/(4q)$$

which gives that

$$H_2(u_{[0,1]})(2) - H_1(u_{[0,1]})(2) = \theta[u(0) + 2u(1)] =$$

$$= \theta[x/(4q) + 2(x/2 - x/(4q)] = \theta\cdot x[1 - 1/(4q)] > 2\varepsilon.$$

Thus it is possible to reject H_2 at t=2. Therefore u(2) can be chosen so that the output y(t) is larger than $y_{ref} - \varepsilon_{ref}$ when t≥4. If H_2 generates the system, H_3 is still always rejected at $t=1$, and the risk of H_1 imposes no restriction on the choice of u. Neither H_1 or H_2 impose any restrictions when H_3 generates the system. It follows that $J(\eta_{[0,-1]},G,H)$ is finite for all $H \in \Xi$. The control law F in Theorem 4.1 is thus not optimal.

□

4.2 AN EXAMPLE OF ADAPTIVE START-UP CONTROL

The start-up problem mentioned in the introduction of Section 4.1 will be solved in a special case.

Example 4.3: The hot-plate process is studied once more where the unknown parameter is m_w instead of γ as in Example 4.1. To get simple notations the quantity of water will be measured by $c_w = m_w C_w$. Let Ξ be the set of models given by (B.1) with parameter values as

$$T_w(0) = T_p(0) = T_{room} = 0°C$$

$$c_w = [2000, 2500, 3000, \ldots\ldots\ldots\ldots, 11000] \; [Ws/°C]$$

$$c_p = 500 \; Ws/°C$$

$$\alpha = 0 \; W/°C$$

$$u \in [0, u_{max}] \qquad u_{max} = 1000 \; W$$

$$\Delta = 50 \; s \; .$$

Regard the basic start-up problem with

$$T_{wref} = 20°C \qquad \varepsilon_{ref} = 3°C \qquad \varepsilon = 1°C.$$

Let α in Definition 4.1 of monotone sets be $1/c_w$. It is seen straightforward that Ξ fulfils all conditions in Definition 4.1 except (4.1.vi).

Consider

$$h_{\alpha 2}(k) - h_{\alpha 1}(k) = \int_{\alpha 1}^{\alpha 2} \frac{\partial h(k)}{\partial \alpha} d\alpha = -\int_{c_{w1}}^{c_{w2}} \frac{\partial h(k)}{\partial c_w} dc_w$$

where the derivative

$$\frac{\partial h(k)}{\partial c_w} = -\int_0^{\Delta} \left\{ \frac{1}{(c_w+c_p)^2} (1-e^{-a(k\Delta-\tau)}) + \frac{\Lambda(k\Delta-\tau)}{(c_w+c_p)^2 c_w} e^{-a(k\Delta-\tau)} \right\} \gamma d\tau$$

is obtained from (4.2) in Example 4.1. According to Definition (4.1) of monotone sets, $\partial h(k)/\partial\alpha \geq 0$ should be increasing in k or $\partial h(k)/\partial c_w \leq 0$ and decreasing in k. Observe first that

$$\frac{\partial h(k+1)}{\partial c_w} - \frac{\partial h(k)}{\partial c_w} = \int_{k\Delta}^{(k+1)\Delta} \frac{d}{dt} \left\{ \frac{\partial h(t/\Delta)}{\partial c_w} \right\} dt.$$

From

$$\frac{d}{dt} \left\{ \frac{\partial h(t/\Delta)}{\partial c_w} \right\} = \int_0^{\Delta} -\left[\frac{a}{(c_p+c_w)^2} + \frac{(-\Lambda a(t-\tau) + \Lambda)}{(c_p+c_w)^2 c_w} \right] e^{-a(t-\tau)} d\tau$$

and from

$$\frac{a}{c_p+c_w} + \frac{\Lambda}{c_w^2} - \frac{\Lambda a t_{max}}{c_w^2} = 0$$

where $\quad t_{max} = \frac{c_w}{\Lambda} \geq 1000$ s,

it follows that (4.1.vi) is fulfilled for $t \leq t_{max}$.

The estimate from Theorem 4.2 is given in Figure 4.2. From Corollary 4.1 it follows that F, given in Theorem 4.1, is completely optimal.

Two different start-ups are shown in Figures 4.3 and 4.4 and the figures illustrate how Ξ_t evolves with time. The T_w-curves show the output of different members of Ξ. The curve of a model is discontinued when the model no longer belongs to the set Ξ_t. The measured output is assumed to be close to the model H_α with $1/\alpha = c_w = 8500$ Ws/°C in both start-ups. In the first start-up (Figure 4.3)

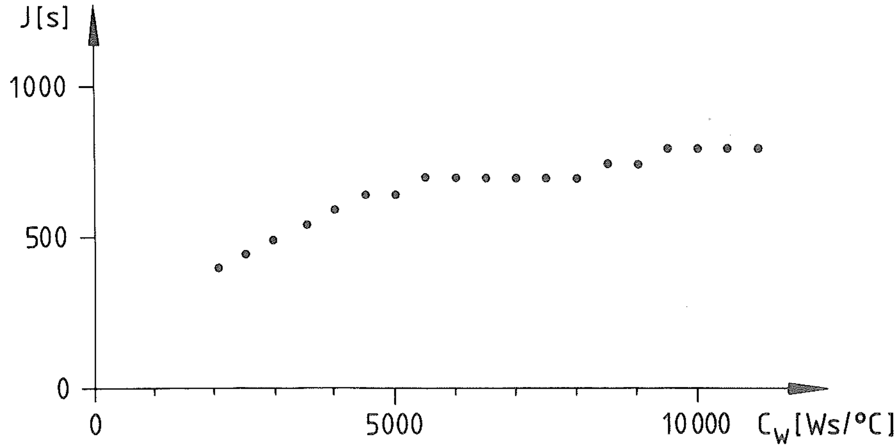

Fig. 4.2: Upper estimate of $J(\eta_{[0,-1]}, F, H)$ as a function of $H \in \Xi$ parametrized with c_w.

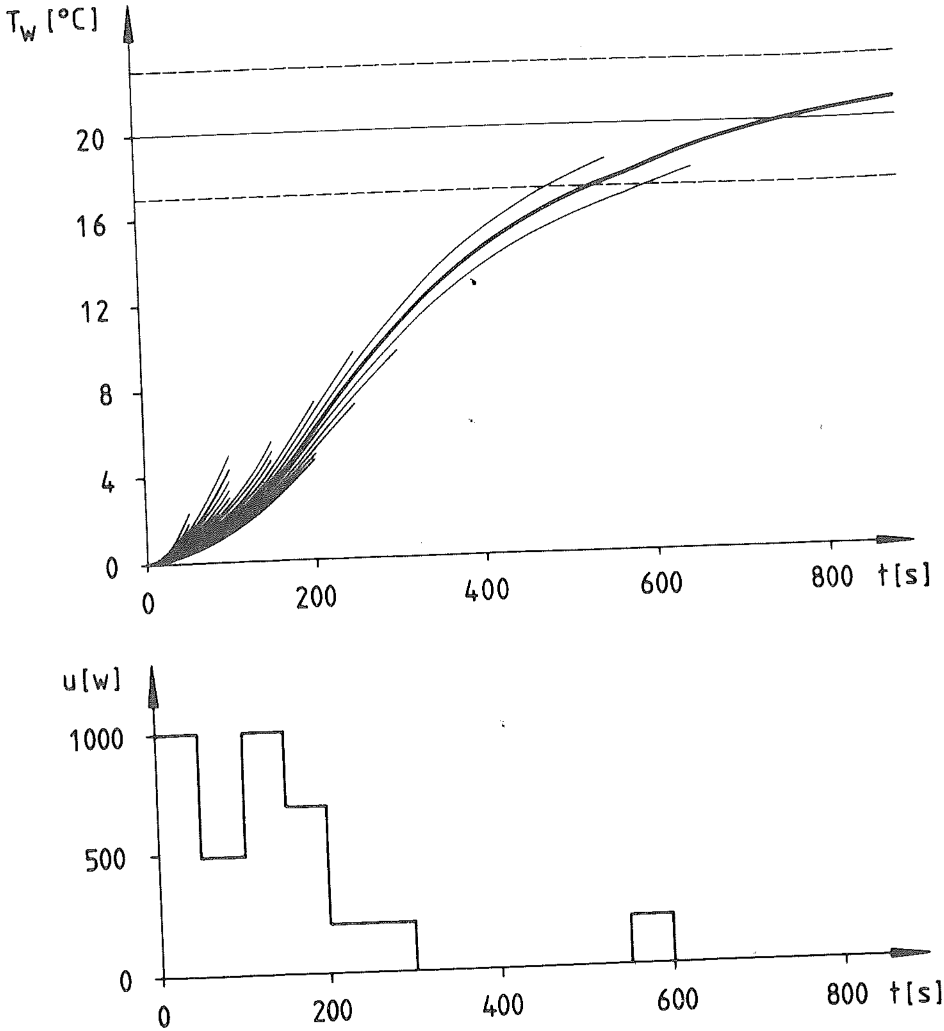

Fig. 4.3: The input and outputs from the models in Ξ when the optimal control law is used.

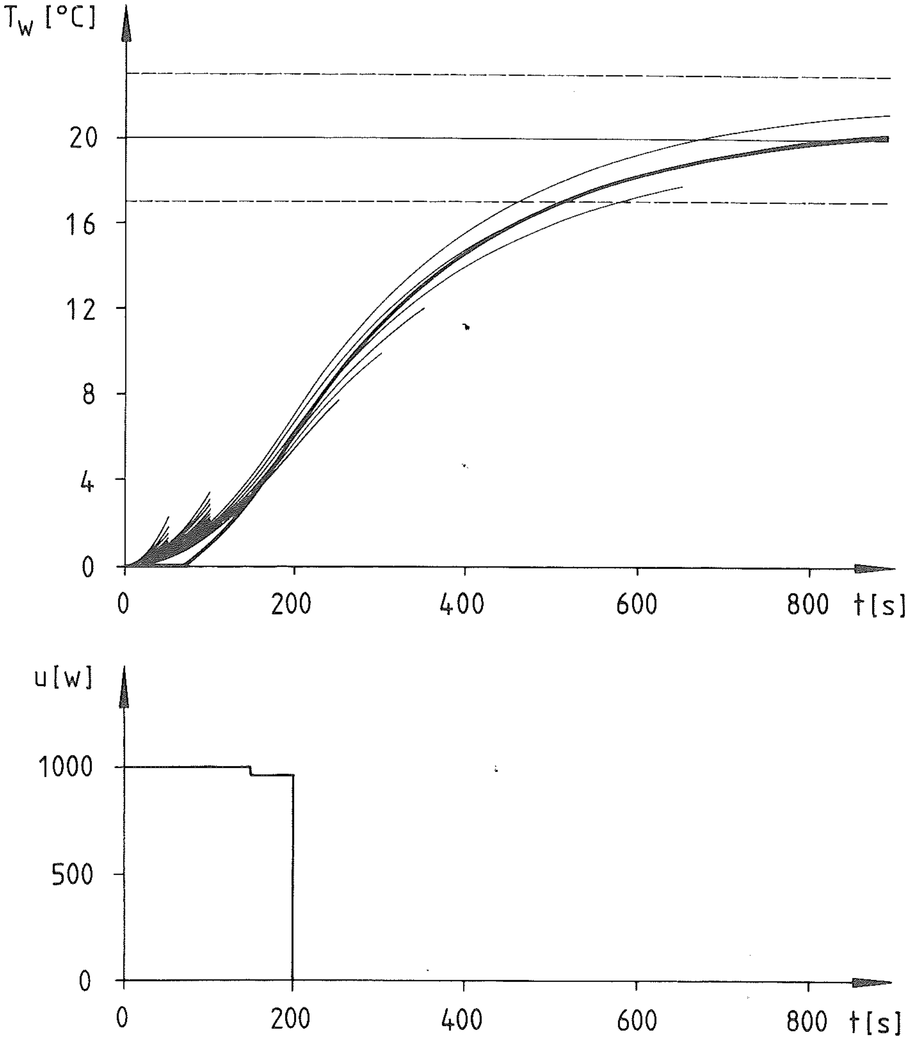

Fig. 4.4: As Figure 4.3 but with another output.

$$y(t) = H_{cw}(u_{[0,t-1]})(t)$$

and in the second one (Figure 4.3)

$$y(t) = \max(0, H_{cw}(u_{[0,t-1]})(t) - 1)$$

In the start-up shown in Figure 4.4, all models with $c_w < 8500$ Ws/kg are rejected almost at once, since $y(t)$ is on the ε-limit of H_α after approximately 70 s. The models with $c_w > 8500$ Ws/kg do not limit u any more than H_α does, so the start-up is made as if Ӿ had contained only H_α. The start-up shown in Figure 4.3 shows a completely different character. Here the input u is limited by models with $c_w < 8500$ Ws/kg which are not yet rejected.

□

4.3 ADAPTIVE START-UP CONTROL AND DYNAMIC PROGRAMMING

Examples 4.2 shows that for the basic start-up problem it is not always optimal to choose the input "as large as possible". It is of interest to find a systematic way to determine the optimal control law. This will be done by using dynamic programming (DP). The essential idea of DP will be presented here for discrete time. The loss J is assumed to be a sum of contributions in each time step:

$$J(x(t_0),u,t_0) = \sum_{t=t_0}^{\infty} J_t(x(t),u(t),t)$$

... the system and J is a positive

real-valued function. Let $V(x(t),t)$ be the optimal loss when starting in $x(t)$ at time t i.e.

$$V(x(t),t) = \min_{u_{[t,\infty)}} \sum_{s=t}^{\infty} J_s(x(s),u(s),s).$$

The idea of DP is to choose $u(t)$ such that

$$V(x(t),t) = \min_{u(t)} \{V(x(t+1),t+1) + J_t(x(t),u(t),t)\}. \qquad (4.17)$$

The precise meaning of the state $x(t)$ and the function J_t requires explaining. For known systems (i.e. in the non-adaptive case), the input sequence $u_{[0,t-1]}$ could be chosen as the state $x(t)$. Often it is, of course, possible to find a lower dimensional state representation, but $u_{[0,t-1]}$ is sufficient. In the adaptive case, information from the output segment has to be included to give a state. Thus a state could be given by

$$x(t) = (u_{[0,t-1]},\eta_{[0,t]}).$$

The next state $x(t+1)$ given $x(t)$ and $u(t)$ is not unique. It is only known that $\eta(t+1)$ is such that $\Xi_{t+1,u,\eta}$ is non-empty. One can denote the set of possible $x(t+1)$ by $\xi(t;x(t),u(t))$. The worst case should be considered, i.e. the $x(t+1)$ in $\xi(t;x(t),u(t))$ that results in the largest loss. (4.17) can thus be rewritten as

$$V(x(t),t) = \min_{u(t)} \max_{x(t+1)\in\xi(t;x(t),u(t))} \{V(x(t+1),t+1) + j_t(x(t),u(t),t)\}.$$

The criterion in the basic start-up problem is minimum time, and the loss function J is thus the time to reach the band $[y_{ref}-\varepsilon_{ref}, y_{ref}+\varepsilon_{ref}]$ and remain there. Introduce the set L_0 as the set of all those states that already have reached the band and can stay there. i.e.

$$L_0 = \{x(t)\colon V(x(t),t) = 0\}.$$

The incremental loss function j_t is thus

$$j_t(x(t),u(t),t) = \begin{cases} 0 & x(t) \in L_0 \\ 1 & \text{otherwise.} \end{cases}$$

It is natural to characterize the functions $V(x(t),t)$ by the sets L_i

$$L_i = \{x(t)\colon V(x(t),t) = i\}.$$

The ideas presented above will now be formalized by the use of the concepts in Chapter 3.

Definition 4.2: For the basic start up problem define L_i, $i\geq 0$, as the set of segment pairs

$$(u_{[0,t-1]}, \eta_{[0,t]})$$

such that

$$\left.\begin{array}{ll} \text{(i)} & \Xi_{t,u,\eta} \text{ is non-empty} \\ \text{(ii)} & \inf\limits_{F\in\Phi_{t,u,\eta}} \sup\limits_{H\in\Xi_t} J(\eta_{[0,t]},F,H) = t+i. \end{array}\right\} \qquad (4.18)$$

□

Remark 4.3: The range of J is the positive integers so that there is an $F'\in\Phi_{t,u,\eta}$ such that the infimum in (4.18.ii) is assumed if i is finite.

□

Theorem 4.3: If there is a $G'\in\Phi$ such that

$$\sup_{H\in\Xi} J(\eta_{[0,-1]},G',H) < \infty$$

for the basic start up problem, then there also exists an optimal control law F.

Proof: F is defined recursively. Assume (4.19) is true for a segment pair $(u_{[0,t-1]},\eta_{[0,t]})$:

$$\left.\begin{array}{lll} \text{(i)} & u(s) = f_s(u_{[0,s-1]},\eta_{[0,s]}) & 0 \le s \le t-1 \\ \text{(ii)} & (u_{[0,t-1]},\eta_{[0,t]}) \in L_i. & \end{array}\right\} \qquad (4.19)$$

Then (4.19.ii) and Remark 4.3 imply that there is a control law $F'\in\Phi_{t,u,\eta}$ satisfying

$$\sup_{H\in\Xi_{t,u,\eta}} J(\eta_{[0,t]},F',H) = t+i$$

Set $u(t)= f_t(u_{[0,t-1]},\eta_{[0,t]})=f'_t(u_{[0,t-1]},\eta_{[0,t]})$

Let $\eta(t+1)$ be such that $\Xi_{t+1,u,\eta}$ is non-empty.

It then follows that

$$\sup_{H\in\Xi_{t+1}} J(\eta_{[0,t+1]},F',H) \leq \sup_{H\in\Xi_{t}} J(\eta_{[0,t]},F'.H) = t+i.$$

F' is best for the "worst" $\eta(t+1)$ but there may be better laws if an "easier" $\eta(t+1)$, were obtained. This implies, according to Definition 4.2, that

$$(u_{[0,t]},\eta_{[0,t+1]}) \in L_{J}, \; J \leq \max(0,i-1). \tag{4.20}$$

From

$$\inf_{G\in\Phi_{0,u,\eta}} \sup_{H\in\Xi_{0}} J(\eta_{[0,0]},G,H) \leq \sup_{H\in\Xi_{0}} J(\eta_{[0,-1]},G',H) < \infty$$

it follows that

$$(u_{[0,-1]},\eta_{[0,0]}) \in L_{k}, \quad \text{some } k, \text{ finite.}$$

By induction in t it follows that F is well defined (see Remark 3.5).

It remains to be proved that F is optimal.

If for all $i\geq 0$, $(u_{[0,t-1]},\eta_{[0,t]}) \in L$ and $F\in\Phi_{t,u,\eta}$ imply that

$$\sup_{H\in\Xi_{t}} J(\eta_{[0,t]},F,H) \begin{cases} = t+i & i \geq 1 \\ \leq t & i = 0, \end{cases} \tag{4.21}$$

then it follows from Definition 4.2 that F is optimal. (4.19) implies (4.20) so (4.21) is true when i=0. Assume that (4.20) is true when $i\leq p$. Assume

$$(u_{[0,t-1]},\eta_{[0,t]}) \in L_{p+1},$$

$F\in\Phi_{t+1,u,\eta}$, and $\Xi_{t+1,u,\eta}$ are non-empty. (4.19) implies (4.20) so

$$(u_{[0,t]},\eta_{[0,t+1]}) \in L_J \qquad J \le p.$$

Hence (4.21) is true for $i\le p+1$.

□

The theorem does not say that it is possible to construct an optimal control law in finite time. However, by making restrictions on the problem the situation is altered.

Corollary: With assumptions as in Theorem 4.3, Ξ being a finite set, finite time horizon, and the input range of u containing only a finite number of values, it follows that an optimal control law can be constructed using a finite number of operations.

Proof: From Theorems 4.3 and 3.2 it follows that there is an optimal control law which is reduced to Ξ. The assumptions allow only a finite number of control laws which are reduced to Ξ.

□

To use dynamic programming to find an optimal control problem will, in most cases, lead to a vast amount of segment pairs $(u_{[0,t-1]},\eta_{[0,t]})$. Whether modern computers have ample capacity to handle this is a relevant problem. The difficulties will be demonstrated in the following example. The adaptive start-up problem in the example is quite trivial. The discretization of the problem is done as far as possible without reducing the problem to a matter of certainty.

<u>Example 4.4:</u> The basic start-up problem is studied in the case of a pure integration when $y_{ref} = 1.0$, $\varepsilon_{ref} = 0.09$ and Σ is given by

$$\dot{z} = u$$

$$z(0) = 0$$

$$y = a \cdot z$$

$$u \in [0.0, 0.2].$$

The set Σ consists of two members, H_1 (a=1.0) and H_2 (a=1.25). All systems $H^a \in \Sigma^a$ are such that the output $y(t) = H^a(u_{[0,t-1]})$ satisfies

$$|H(u_{[0,t-1]})(t) - y(t)| < \varepsilon = 0.06$$

for $H = H_1$ or H_2.

The sampling period is supposed to be one time unit and the input is discretized:

$$u = [0, 0.1, 0.2].$$

In Figure 4.5 the segment pairs are grouped together so that each ellipse represents a set of segment pairs for which

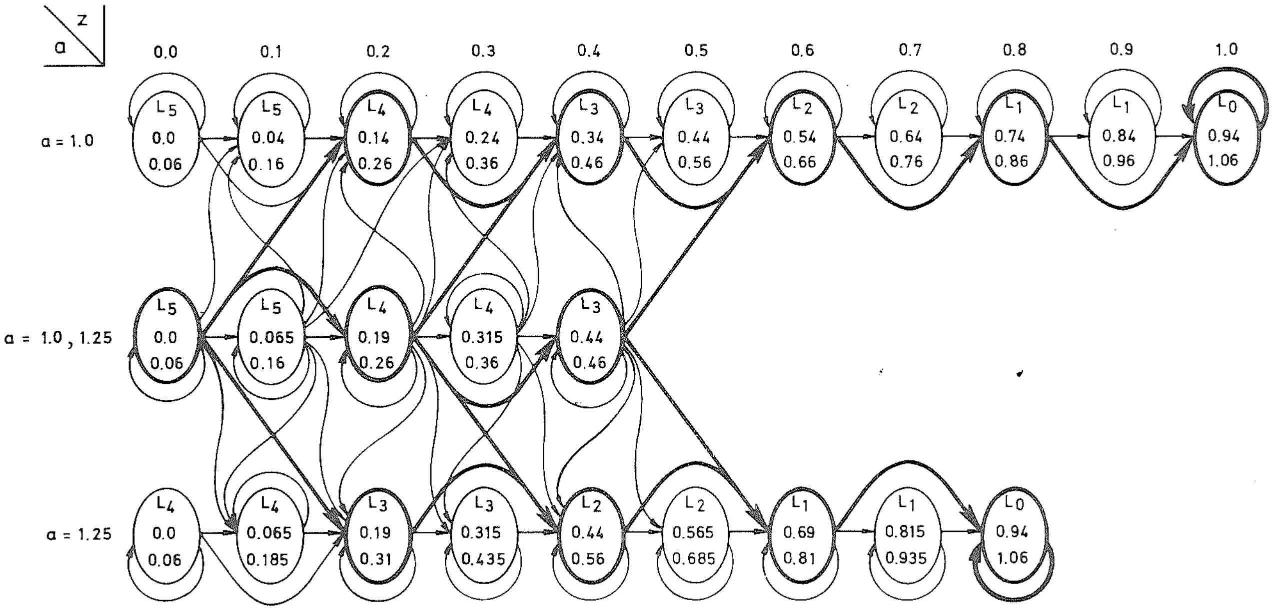

Fig. 4.5: The graph shows the start-up of the integrator in Example 4.4.

(i) $z = \sum_{i=0}^{t} u(i)$ are equal (given by the column in the graph)

(ii) $\Xi_{t,u,\eta}$ are equal (given by the row in the graph).

The minimum and maximum value of $y(t)$ is marked in each ellipse. Observe that t does not need to be the same for different segment pairs belonging to the same ellipse. The future behaviour only depends on the sum given in (i) and the set $\Xi_{t,u,\eta}$. This means that segment pairs belonging to the same ellipse belong to the same set L_i. Which set L_i a ellipse belongs to is marked. Possible transitions are marked with arrows. An optimal control law (in this case unique) is given by u=0.2 until $\Xi_5=\{a=1.0\}$ or $\Xi_4=\{a=1.25\}$. After that u=0.

The optimal control law is marked in Figure 4.5 by fat arrows. Dynamic programming applied to ordinary (not adaptive) start-up problems often produces so large a graph that it is impossible to handle it even in a large computer. Dynamic programming in the case of the adaptive start-up problem is, of course worse.

CHAPTER 5

MODEL BASED CONTROL

A method for start-up control has been given in Chapters 3 and 4 for special processes. Here it will be stated in general. Problems of choosing Ξ, ε, and ε_{ref} are discussed and illustrated by the start-up problem for the hot-plate process. It is argued in Section 5.2 that Ξ should consist of physical models. Different methods of generating ε are discussed in Section 5.3. In Section 5.4 the dependence between the discretization of the set Ξ and the needed ε is discussed.

Instead of discretizing Ξ, it can be regarded as a set of subsets. Such a method is discussed in Section 5.5. Finally a suitable control strategy based on the result of Theorem 4.1 is elaborated in Section 5.6.

In Chapter 7 references are frequently made to this chapter when presenting a controller for the hot-plate process.

5.1 A PROPOSAL FOR A METHOD OF START-UP CONTROL

According to Chapters 3 and 4 one method of start-up control is to choose a set of models Ξ and a set of model error functions $\varepsilon(t;u;H)$ fulfilling Assumption 3.1. Models are rejected as the outputs are received, which is formulated by the set of currently possible models, Ξ_t. From Theorem 3.2 it follows that it is sensible to let the control be based on old inputs and on the subsets Ξ_t. Choosing u according to the worst case is considered by requiring that the control is admissible. In Section 5.6 the implementation of the control law is further discussed.

The choices of Ξ, ε, and ε_{ref} are strongly dependent on each other. ε_{ref} will more or less be given by the problem.

Observe that near the goal $\varepsilon < \varepsilon_{ref}$, so a smaller than necessary ε_{ref} can result in an unnecessary work to find suitable models fulfilling Assumption 3.1. Further the examples of Chapter 4 show that the control depends critically on the relation between ε and ε_{ref}. In Example 4.2 it was even proved that if ε was close enough to ε_{ref} there were cases when the output never reached the goal for the reasonable control law of Theorem 4.1.

Even if the choice of ε_{ref} is important it is easy compared with the choice of ε and Ξ. These choices are completely dependent on each other as seen by Assumption 3.1.

5.2 Ξ GIVEN BY PHYSICAL MODELS

In principle, the functions H of Ξ can be regarded as a large table giving an output segment to every input sequence. However, the method of expressing H is of utmost significance. The model type determines the possibility of using the results of the science of material physics,. thermodynamics etc.. As an example, consider the use of Model (B.1) and the use of ARMA-models (see Ex. 5.1). Models of the type (B.1) will be called physical models. No strict definition will be given.

The main reason for using ARMA-models in adaptive control is the ease of computation and the advantage of generality. The ease of computation is due to the least square method of identification (see Goodwin and Payne, 1977). For the proposed method the idea is, however, to reject impossible models, not to choose what seems for the moment to be the best one. The generality of the ARMA-process is certainly of great advantage when it comes to selling canned controllers. The control problems addressed in the present work as exemplified in Section 2.1, are however, such that the use of much a priori information seems to be of great advantage. As an example consider the pulse-response shown in Figure 3.2. The use of a low order ARMA-model should require a very long sampling interval in order to avoid a completely wrong identification in the beginning. A high order ARMA-model would, on the other hand, be difficult to start-up. Add to this the difficulties in handling non-linearities. The difficulties of expressing physical quantities in ARMA-models will be illustrated by an example.

Example 5.1: Assume that the second order model (B.1) is transformed to a second order ARMA-model:

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_1 u(t-1) + b_2 u(t-2). \quad (5.1)$$

Not all sets of parameter values $\{a_1, a_2, b_1, b_2\}$ will describe a hot-plate process. Modelling errors can give parameter estimates that are inconsistent with the physical process. To avoid this, constraints on the parameters are introduced. Assume that the water quantity m_w is the only free parameter with $0.5 \text{ kg} \le m_w \le 2 \text{ kg}$. The a- and b-parameters will depend on the parameter m_w in the following way:

$$a_1 = -(1 - e^{-h/\tau})$$

$$a_2 = e^{-h/\tau}$$

$$b_1 = \frac{c_p}{c_w + c_p}\left[1 - \frac{c_w c_p}{\Lambda(c_w + c_p)}(1 - e^{-h/\tau})\right]$$

$$b_2 = \frac{c_p}{c_w + c_p}\left[1 + (h - \frac{c_w c_p}{\Lambda(c_w + c_p)})(1 - e^{-h/\tau})\right]$$

where $c_w = m_w C_w$; $c_p = m_p C_p$; $\tau = \dfrac{c_w c_p}{\Lambda(c_w + c_p)}$

h is the sampling period.

The constraints on a_1, a_2, b_1 and b_2 are complicated and the advantage of easy computation when using ARMA-models is thus lost.

□

Further, the use of physical models is of great importance when $\varepsilon(t;u;H)$ is to be given. This is discussed in the next two sections.

In the rest of this section some general problems of approximating the set Ξ^a with a set Ξ will be discussed.

(1) The structure of the members of Ξ^a is only partly known. The process is not necessarily a "black box", but rather "grey", in the sense that some structural parts are known while others may be considered completely unknown. In the hot-plate example the heat capacity of the plate and the water are known. The heat distribution in the kettle is

known in a qualitative way. The hydro-dynamics of the water is considered to be a black box.

(2) The structure is known but some parameters are unknown. The heat-transfer between the plate and the kettle is described by the heat-transfer ratio Λ. The structure is known or prescribed, but the value is not known.

(3) Unknown functions in members of Σ^a.
Physical models often contain algebraic functions which are unknown. Consider Model (B.1) of the hot-plate process when $\alpha \cdot T_w$ is replaced with $\varphi(T_w)$. Assume that the loss function φ is unknown. Even if it is known to be positive, increasing, and convex, it is not trivial to parameterize it. In principle, a correct parameterization would

5.3 MODELLING ERROR ε

Formally, the error $\varepsilon(t,u,H)$ is defined by Assumption 3.1, or in other words, the relation between the error $\varepsilon(t,u,H)$ and the model H must fulfil Assumption 3.1. In practical cases Σ^a is not known except through the real process, and consequently it is not possible to calculate a maximal modelling error $\varepsilon(t,u,H)$. It has to be an estimated upper bound of the real modelling error. For the case of Example 3.1 it follows that $\varepsilon(t,u,H)$ would be at least 3°C. If ε were constant the result would be a large ε_{ref}. However, when the process is in stationarity it seems possible to use an ε lower than 3°C. This implies that ε

should depend on u. Certainly the model error depends on the model H which is easily imagined in the hot-plate example.

A simple way to determine $\varepsilon(t,u,H)$ seems to be to run a lot of experiments for different cases. Consider the hot-plate process when Ξ is given by Model (B.1) with $\alpha \cdot T_w$ replaced with $\varphi(T_w)$. But $\varepsilon(t,u,H)$ should depend on u and H resulting in the need for many experiments. For each experiment the output is compared with the output from the model in Ξ with parameter values corresponding to those in the experiment. Whereas m_w is certainly known, φ and Λ are much the opposite.

When determining suitable $\varepsilon(t,u,H)$ it is of great advantage if the models in Ξ are physical models. This makes it easier to use the knowledge of science. Even then it can be very difficult to give a good estimate of $\varepsilon(t,u,H)$. As an example, the water temperature of the hot-plate process is studied. In Model (B.1) the temperature is assumed to be homogeneous in space. The error caused by such an assumption is very difficult to estimate if the circulation of the water should be taking into full consideration. On the other hand the simple approach of considering water not in circulation leads to results of no practical value. Due to the low value of heat conductivity of water, 0.56 W/mK, the difference between the top and bottom temperatures, would be more than 100°C during an ordinary start-up.

A systematic way of determining $\varepsilon(t,u,H)$ is hard to give. It has to be developed with respect to the actual process. In Chapter 7 an example is given of how to determine $\varepsilon(t,u,H)$.

Normally, $\varepsilon(t,u,H)$ has to be generated on-line due to the dependence of u as shown in Example 5.2.

Example 5.2: When modelling the hot-plate process the value of the heat transfer rate is critical. Assume that Λ is to be considered fixed and that Σ^a is given by Model (B.1) for some $\Lambda \in [1.5, 2]$ W/K. Assume further that $\alpha=0$ and $m_w \in [0.5, 2]$ kg. Let Σ be given by (B.1) when $\Lambda=2$ W/°C, $\alpha=0$ W/°C and $m_w \in [0.5, 2]$ kg. The modelling error $\varepsilon(t,u,H)$ is easy to generate. However, this has to be done for each model $H \in \Sigma$. ε_- will be generated in the case when $m_w=1$ kg. Let H_{inf} denote Model (B.1) with $m_w=1$ kg, $\Lambda=1.5$ W/°C and denote the states of that model by Y_w and Y_p. The states of model H ($m_w=1$ kg, $\Lambda=2$ W/°C) are denoted by T_w and T_p. It follows that with $\Lambda=2$ W/°C and $\Delta\Lambda=-0.5$ W/°C.

$$c_w \dot{T}_w = -\Lambda(T_w - T_p)$$

$$c_p \dot{T}_p = \Lambda(T_w - T_p) + u$$

$$c_w \dot{Y}_w = -(\Lambda+\Delta\Lambda)(Y_w - Y_p)$$

$$c_p \dot{Y}_p = (\Lambda+\Delta\Delta)(Y_w - Y_p) + u$$

$$T_w(0) = T_p(0) = Y_w(0) = Y_p(0) = 0°C.$$

With $Z_w = T_w - Y_w$ and $Z_p = T_p - Y_p$ the equations above give

$$c_w \dot{Z}_w = -\Lambda(Z_w - Z_p) + \Delta\Lambda(Y_w - Y_p)$$

$$c_p \dot{Z}_p = \Lambda(Z_w - Z_p) - \Delta\Lambda(Y_w - Y_p)$$

$$Z_w(0) = Z_p(0) = 0°C.$$

u is always positive, so $Y_w - Y_p$ is negative.

$$\left.\begin{aligned} Z_w &= -\frac{c_p}{c_w} Z_p \\ 2c_w \dot{Z}_w &= -2\Lambda\left(1+\frac{c_w}{c_p}\right) Z_w + 2\Delta\Lambda(Y_w - Y_p) \end{aligned}\right\} \Rightarrow Z_w \geq 0.$$

From the above it follows that the outputs from the members of Σ^a with m_w=1 kg will be lower or equal to T_w and greater or equal to Y_w. Thus $\varepsilon(t,u,H)$ is given by

$$\varepsilon_+ = 0°C \quad \text{and} \quad \varepsilon_-(t,u,H) = T_w(t) - Y_w(t)$$

where ε_+ and ε_- were defined in Remark 3.3. In Figure 5.1 a start-up is shown. Observe how ε_- decreases to zero for large t.

□

5.4 THE RESOLVING POWER OF Σ AND THE PARAMETER RESOLUTION

The modelling error ε depends both on errors in model structure and errors in parameter-values. The error caused by the structural error will give a natural limit of how fine the resolution in parameters has to be. For instance, in the hot-plate case it is of no use to specify the quantity of water m_w for each gram (0.001 kg) for a second order model. The change in the output caused by a change of m_w by one gram is negligible compared with the error caused by structural errors. It is natural to say that Σ has a resolving power which sets a limit on the needed resolution

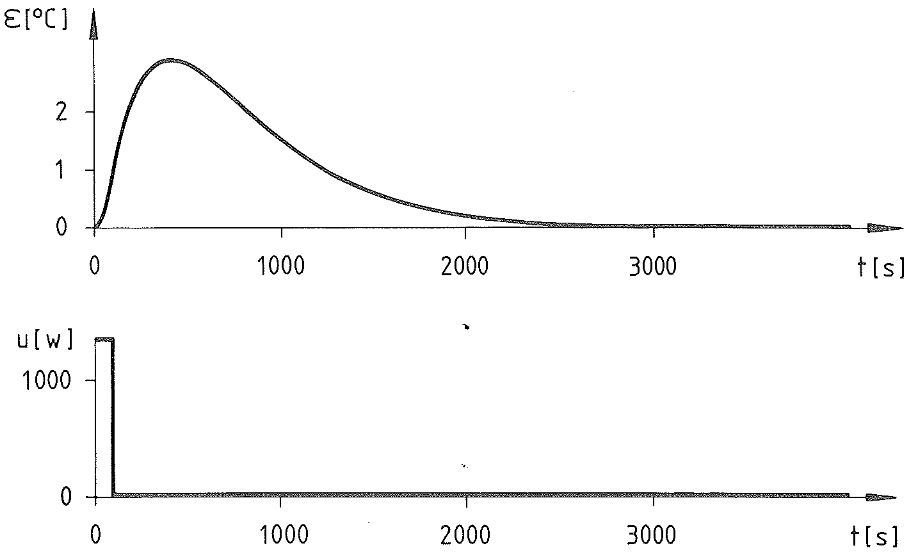

Fig. 5.1: Estimation of ε as the difference between the output from the models H and H_{inf}. The input u is also shown.

in the parameters. A proposal for a definition of the resolving power of Ξ is given in Appendix C.

Parameter resolution in relation to ε

Assume that a family of models is to be chosen in the hot-plate case. It is assumed that Model (B.1) is used to generate the members of Ξ. The only uncertainties are the quantity of water and the loss from the water. It is assumed that

$$m_w = [0.6, 2] \text{ kg} \qquad \alpha = [1, 10] \text{ W/°C}$$

for the real process.

In order to keep the administration and computing time down it is desirable to have few members in Σ. Few members mean that the resolution in m_w and in α is low. Good control requires, on the other hand, good resolution in m_w and in α. A compromise has to be made. This compromise and its relation to ε will be discussed here.

First the resolution in m_w is examined. Assume that Σ consists of the models obtained for

$$m_w = [0.6,\ 0.8,\ \ldots\ldots\ldots,\ 2.0]\ \text{kg} \qquad \alpha = 4\ \text{W/°C}.$$

For the purposes of this presentation, it is enough to study the two members which have $m_w = 1.0$ and 1.2 kg. By combining a pulse and a step a good start-up of the system is obtained (see Figure 5.2), where the outputs are given. It is observed that the outputs are converging on each other. This follows from the fact that the loss $\varphi(T_w)=\alpha(T_w-T_{room})$ is an increasing function of T_w. $T_{room}=0$°C in this Example.

Assumption 3.1 must also be true when $m_w = 1.1$ kg which gives a large error ε during the transient. In stationarity, however, the error caused by an error in the m_w-value is small. From Figure 5.2 it is also clear that the best chance to decide on the parameter m_w with good precision is during the transient.

Second the resolution in the loss $\alpha \cdot T_w$ will be examined. Assume that Σ consists of the models obtained for

$$\alpha = [1,\ 2,\ \ldots\ldots\ldots\ldots\ldots,\ 10]\ \text{W/°C} \qquad m_w = 1\ \text{kg}.$$

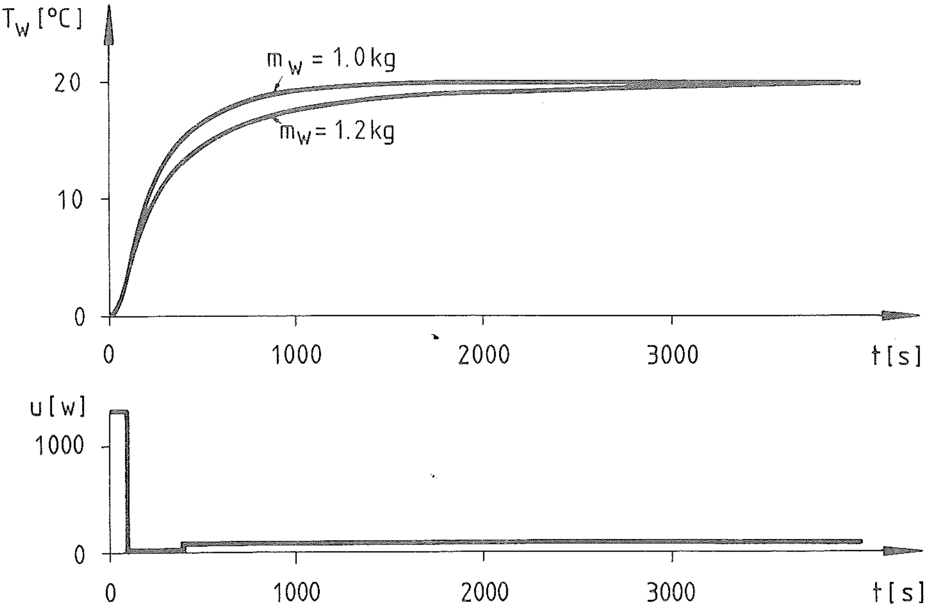

Fig. 5.2: The response from the system given by the second order model is shown for two cases of m_w: 1.0 and 1.2 kg.

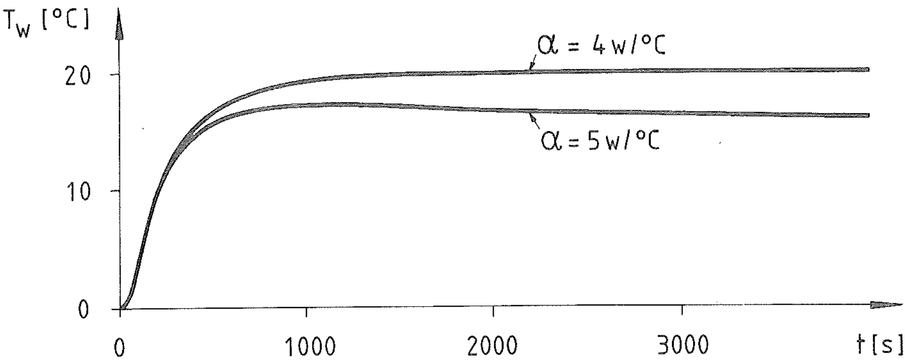

Fig. 5.3: For the input signal showed in Figure 5.1 the responses of systems are shown for the two cases when α = 4 and 5 W/°C.

By the same reason as for the case above, only the two members with $\alpha = 4$ and 5 W/°C will be studied. With the same input as above, good start-ups are obtained for both members. The outputs are shown in Figure 5.3. In this case the curves are diverging and the situation, with regard to ε, is reversed in time compared with the m_w-case.

Hence, if the temperature must be predicted with good precision during the transient, then the resolution in m_w has to be good and consequently ε must be small in order to estimate m_w. If the temperature has to be known with good precision in stationarity then it is sufficient with a good resolution in α, and ε has to be small in stationarity to estimate α.

The discussion will be concluded by assuming that Ξ consists of all members given by Model (B.1) and all combinations of

$$m_w \in [0.6, 0.8, \ldots, 2.0] \text{ kg} \quad \text{and} \quad \alpha \in [1, 2, \ldots, 10] \text{ W/°C}.$$

In the beginning of a start-up the temperature is, in a sense, uninteresting. Its value is of importance when the temperature is close to the reference temperature. In the beginning it is important to know m_w with good resolution in order to achieve a fast start-up. Thus it is desirable to have a good resolution in m_w but necessary to have it in α. Thus a Ξ with

$$m_w = 2.0 \text{ kg and } \alpha \in [1, 2, \ldots, 10] \text{ W/°C}$$

is possible to use but not a Σ with

$$m_w \in [0.6,\ 0.8,\ \ldots, 2.0]\ \text{kg} \quad \text{and} \quad \alpha \quad \text{fixed.}$$

To achieve a good estimate of m_w a small ε is required during the transient. Such a small ε requires a high order model. In the steady state ε must be small. Such an ε requires only a good estimate of the static gain; therefore, it could be sensible to have an ε which is small only in stationarity. This would allow a good estimate of α and good precision in the knowledge of the temperature when in stationarity. Of course, the estimate of m_w will be poor. This means that the control will be cautious in the beginning. So the use of a low order model is paid for in longer start-up times.

5.5 REGARDING Σ AS A SET OF SUBSETS

By regarding the set Σ as a set of subsets the means of estimating $\varepsilon(t,u,H)$ is simplified compared with the case in which Σ is discretized. Even the handling of Σ is simplified. The method will be discussed in this section but will not be used in the sequel.

Let Σ be generated by Model (B.1) where $m_w \in [0.5,\ 2]$ kg and $\alpha = 0$ W/°C. Divide Σ into disjoint subsets and let the subset M_i contain the models where

$$m_{i,1} \le m_w \le m_{i,s}.$$

Let $H_{i,l}$ and $H_{i,s} \in \Xi$ be the models with $m_w = m_{i,l}$ and $m_{i,s}$. It is easy to see that

$$H_{i,s}(u)(t) \leq H(u)(t) \leq H_{i,l}(u)(t).$$

Denote the estimate of the error caused by the structure error with $\varepsilon_{struc}(t,u)$ and let it be the same for $H_{i,l}$ and $H_{i,s}$. The whole subset M_i can now be rejected at time t if

$$y(t) < H_{i,l} - \varepsilon_{struc} \quad \text{or} \quad H_{i,s} + \varepsilon_{struc} < y(t).$$

The difficulties of using subsets instead of parametrizing Ξ may seem to be linguistic rather than practical. However, the idea applied to the third problem in Section 5.2, where $\alpha \cdot T_w$ in (B.1) is replaced by a loss function $\varphi(T_w)$ gives another result. How to discretize in such a case is not obvious. Certainly there are countably dense sets of functions among the loss functions φ. A suitable norm is given by the supremum norm. But which enumeration of the loss functions φ should be taken and how should it be administered? Here the idea of subsets can be of great value. Let, in this case, Ξ be generated by (B.1) where $m_w = 1$ kg and $\alpha \cdot T_w$ is replaced by $\varphi(T_w)$ and where φ is known to have the following properties:

$$\varphi(T_{room}) = 0 \text{ W}$$

$$\frac{d\varphi}{dT} > 0 \qquad \frac{d^2\varphi}{dT^2} > 0 \qquad \varphi(95°C) < 400 \text{ W}.$$

The loss curve must remain inside the shadowed region in Figure 5.4. The set of possible curves can be diveded into different groups. One group may be the group G_1 with $\varphi(50°C) > 20$ W. This group will have a smallest and a greatest member, which are depicted in Figure 5.5. The complement of G_1, here called G_2, has also a smallest and a greatest member. These have also been depicted in the figure by dashed lines. In principle, it is easy to make the classification with finer resolution but there is no

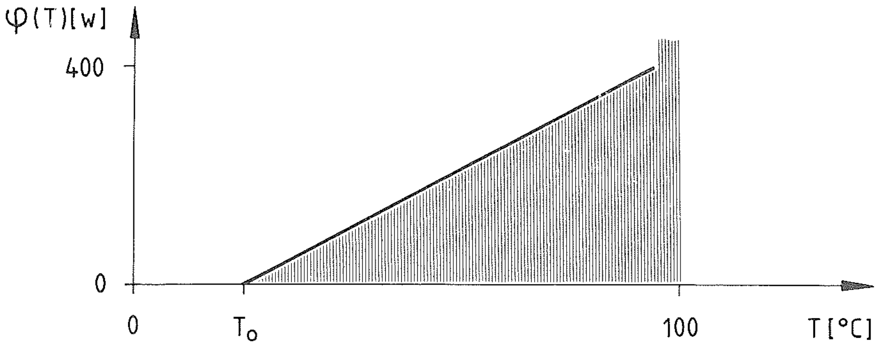

Fig. 5.4: The shadowed region shows where the loss curve can be.

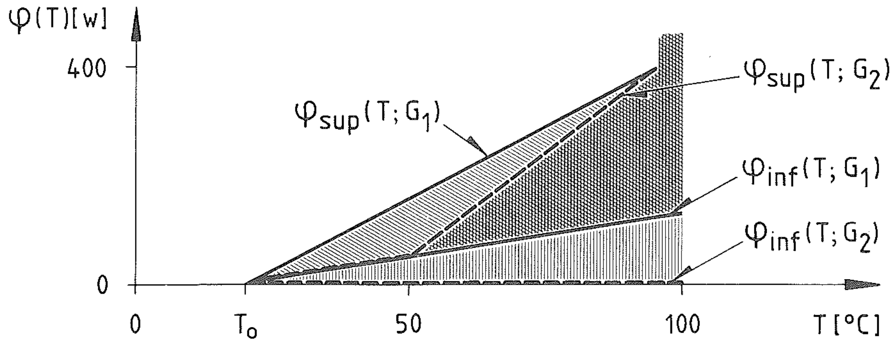

Fig. 5.5: The figure shows two classes G_1 and G_2 of loss function. The supremum and the infimum of the classes are marked.

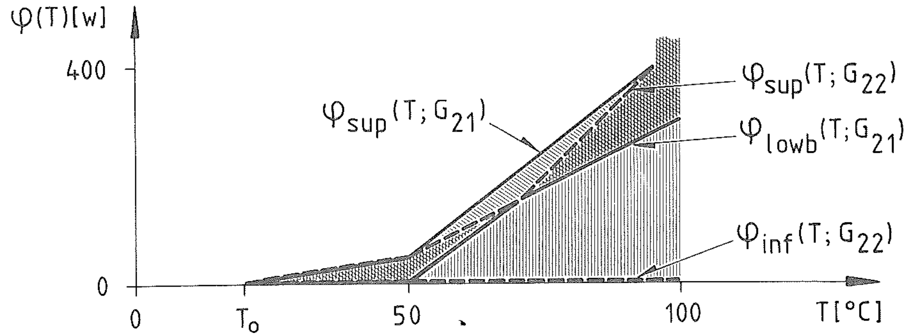

Fig. 5.6: The class G_2 which is shown in Figure 5.5 is divided in two groups G_{21} and G_{22}. The supremum and the infimum of the subclasses are marked.

assurance of a greatest or lowest member; yet upper and lower extremes can be given. With fine resolution, the upper and lower extremes of each class are close to each other. In Figure 5.6 the loss curves in G_2 have been divided in two groups.

The choice of the subsets should be made on-line in order to be efficient. For instance, if a start-up is made up to 70°C and it goes fast, then interest should be concentrated on temperatures around 70°C when dividing Σ into subsets. On the other hand, if the start-up is slow, it is natural to consider lower temperatures when dividing Σ into subsets.

5.6 CONTROL STRATEGY

The only presented method of finding an optimal control law in a general case is with the use of DP. However, DP will normally require so much calculations that it is impossible to find an optimal control law to a start-up problem of practical interest. In such a case the choice of the control law has to rely on reasoning. In order to illustrate the choice of a control law and how to implement it, consider the hot-plate process. Here ε is not constant, so the result of Theorem 4.1 is not applicable. On the other hand, the control law (4.3) of Theorem 4.1 seems to be reasonable even if ε is not constant. The implementation of it will here be discussed in the case when ε_+ is constant but $\varepsilon_-=\varepsilon_-(t;u;H)$.

<u>Calculation of maximal u(t)</u>

If the model set Ξ were generated by Model (B.1) when $\alpha=0$, it would be easy to find the largest possible $u(t)$ for a certain model. The water temperature is increasing and the limit is given by

$$\lim_{t\to\infty} T_w(t) = \frac{h}{c_w + c_p} \sum_{r=0}^{\infty} u(r).$$

If $\alpha \neq 0$ in Model (B.1) there is no such easy formula. One solution is to use an iterative method. For each t and in principle each $H \in \Xi_{t,u,y}$ the following iterative procedure is made: The control signal is chosen as

$$u(s) = \begin{cases} u_{test} & s = t \\ 0 & s > t. \end{cases}$$

The model output $T_w(t)=H(u)(t)$ is calculated until a maximal $T_w(t)$ can be distinguished. If it is not close enough to $T_{wref}+\varepsilon_{ref}-\varepsilon$ another attempt is made with a new u_{test}, if the input constraints allow. Using such an iterative method it is rather easy to find the highest possible u(t). For the implementation described in Chapter 7, the number of iterations never exceeds eight. The model output T_w will of course reach T_{wref}, and if the model describes the process well, so will the process output.

Local Stability of the proposed control law

The losses will give a stationary point $T_{wref} > T_{room}$ for $u=u_0 > 0$. In general, high order models of the hot-plate will give sampled versions which, linearized around the stationary point, has a zero outside the unit circle (see Åström, Hagander, Sternby, 1980). For instance, for a fourth order model (the model in Example 6.4 with a loss that is linear in T_w) and the sampling period h=10 s, one zero was found near z=3.4. Above it was said that the maximum of T_w should if possible be "close enough" to $T_{wref}+\varepsilon_{ref}-\varepsilon$. If "close enough" is altered to "exactly equal to" the proposed control law becomes an output dead beat controller, which means that the closed system will be locally unstable when the model has a zero outside the unit circle. In the hot-plate case such instability is quite harmless. The input signal will oscillate, but the limits on u will prevent high amplitudes. In the worst case an energy pulse of about 1350 Ws is inserted to the hot-plate. The process is so slow that it will take several hundred seconds before such a pulse of energy has reached the water. At that time the losses have removed most of the energy. If the losses are small, u_0 is small, which implies that the amplitude will be limited by u_0 and not by $u_{max}-u_0$.

If the control is not acceptable there are many possible things to do. For instance when T_w is near T_{wref}, the control law can be changed to a state feedback controller in order to get a stable system.

CHAPTER 6

SOLVING START-UP PROBLEMS WITH PD-RELATED CONTROLLERS

In Chapter 4 it was argued that in realistic cases it is usually not possible to find the optimal control to an adaptive start-up problem. In this chapter a discussion will be undertaken on how well a PD-controller solves certain start-up problems.

In industry, start-up control is usually associated with starting processes in paper mills, oil refinery and power plants etc.. In such cases the start is usually made in small steps or the reference value is ramped slowly. In the case of small steps a new one is not taken before the last step has reached stationarity. During a step the process is usually controlled by a PID-controller or by a simple on-off controller.

In this chapter the change in reference value is supposed to be done in one step. It is shown that a high gain PD-controller is optimal in some cases and gives good control for some other adaptive start-up problems. The I-part does not play the same fundamental roll as the P- and D-parts.

In Section 6.2 it is shown that a PD-controller solves the start-up problem for the hot-plate process in a certain case. In fact the controller is nearly optimal. However, in order to succeed, some ad hoc features have to be added. In Section 6.3 it is argued that a PD-controller for the hot-plate process would, in a practical case, require further ad hoc features. The general conclusion is that it may be possible to solve start-up problems based on such a simple controller as a PD-controller. However, in order to solve practical start-up problems ad hoc features have to be used. They rely on specific knowledge about the actual process but the knowledge is not used in an efficient or systematic way.

6.1 GENERAL ASPECTS OF PD-CONTROLLERS

In automatic control PID-controllers dominate the scene nearly totally. The advent of the computer has not changed this picture for several reasons. First, it is easy and cheap to make PID-controllers in analog technique. Not only electrical implementations but also mechanical, pneumatic, and hydraulic implementations are possible. Second, it is easy to understand and use them. An ordinary PID-controller has four knobs to adjust (the fourth knob gives the filter constant), and it is easy to learn how to use them. Third, PID-controllers usually work very well. There are drawbacks too, like integral wind-up. The problem is, however, sometimes easy to circumvent, when the regulator is implemented digitally.

The advent of the digital computer on a chip has accelerated the trend of adding ad hoc features to the PID-controller. It is worth noting that computers are mainly replacing analog PID-controllers with digital ones. There are few installations of digital controllers that rely on control principles which do not go back to the principles of the PID-controller. Considering the level of modern control theory and modern computer technology, it can be said that the PID-controller is doing well.

One reason for the continued use of the PID-controller is that other controllers are difficult to design. Even for what seems to be a simple adaptive start-up problem, it can be very difficult to produce an optimal controller as has been shown in Chapter 4. Another reason is that the

PD-controller comes close to the optimal solution of some common basic start-up problems. The optimality in the cases discussed here critically depends on the saturation of the input $u(t)$ and on the form of the terminal condition on the output $y(t)$. Four examples will be given.

<u>Example 6.1:</u> Consider the basic start-up problem with the models of Σ given by pure integrators:

where $0 \leq u \leq 1$ and $a \in [0, a_{max}]$. A saturated P-controller would give

$$u = \text{sat}[K(1-y)] \qquad (6.1)$$

where $\text{sat}(x) = \min[0, \max(x,1)]$. This control is actually optimal for high gain K and short sampling interval h, if ε_{ref} is large enough. A formulation along the lines of Chapter 3 will be given. All outputs $y(t)$ are assumed to satisfy Assumption 3.1:

$$|y(t) - H(u_{[0,t-h]})(t)| \leq \varepsilon \qquad (6.2)$$

where ε is constant. Assume moderate ε_{ref}-requirements:

$$\varepsilon_{ref} > 4\varepsilon.$$

That the control (6.1) is admissible follows from the special structure of the system. Assume

$$y_{ref}+2\varepsilon \leq y(t) < y_{ref} + \varepsilon_{ref} - 2\varepsilon.$$

Then, from (6.2) and $H(u)(s) \geq H(u)(t)$ when $s>t$, it follows that $y(s) \geq y_{ref}$ for $s \geq t$, so that $u(s)=0$, $s \geq t$, and

$$y(s) < H(u)(s)+\varepsilon = H(u)(t)+\varepsilon < y(t)+2\varepsilon < y_{ref}+\varepsilon_{ref}.$$

If on the other hand $y(t) \leq y_{ref}+2\varepsilon$, then

$$y(t+h) \leq H(u)(t+h)+\varepsilon \leq H(u)(t)+\varepsilon+a_{max}\cdot h <$$

$$< y(t)+2\varepsilon+a_{max}\cdot h < y_{ref}+\varepsilon_{ref}-2\varepsilon,$$

for a short enough h.

Σ is a monotone set, so a completely optimal control law is given in Theorem 4.1. In fact, for large K, the control law (6.1) is also completely optimal. Assume that for some t, $u(t)<1$. Then $y(t) > y_{ref}-1/K$ and for $s \geq t$

$$y(s) \geq H(u)(s)-\varepsilon \geq H(u)(t)-\varepsilon \geq y(t)-2\varepsilon >$$

$$> y_{ref}-1/K-2\varepsilon > y_{ref}-\varepsilon_{ref} \quad \text{for } K>1/2\varepsilon.$$

□

The P-controller in Example 6.1 could be considered as an on-off controller. By introducing a time delay in the process it can easily be seen that the P-controller does not solve the problem. However, a derivative control will help.

<u>Example 6.2:</u> Consider the basic start-up problem as in Example 6.1, but let the models of the process be given by

Assume that $u \in [0, 1]$, $a \in [0, a_{max}]$, and $a_{max} T < y_{ref}$. The last condition is essential because also the PD-controller will act as an on-off controller. Let the output from the PD-controller be

$$u(t) = \begin{cases} \mathrm{sat}[K(y_{ref} - y - T_D y')] & y < y_{ref} - d \\ 0 & \text{otherwise.} \end{cases}$$

Assuming an exact model and d=0 the controller is turned off approximately when

$$K(y_{ref} - y - T_D y') = K(y_{ref} - a(t-T) - T_D a) = 0$$

giving $t = y_{ref}/a$ if $T_D = T$ and K large. This is what is desired. A closer look shows, however, that y will have a plateau $y = y_{ref} - 1/(2K)$ during the time $t = y_{ref}/a$ until $t = y_{ref}/a - 1/aK + T$. During the same time u is equal to 0.5 giving that y in the end becomes about $y_{ref} + aT/2$. K is assumed to be large. Adding the dead zone to the controller, the problem of the overshoot is solved if $d > 1/(2K)$. The controller is, however, not admissible in the meaning of Chapter 3 unless Assumption 3.1 is augmented. The problem is due to the use of the derivative y'. If Assumption 3.1 is augmented to exclude high frequency noise and ε is constant, it is possible to prove that the PD-controller is completely optimal. The proof is technical and is given in Appendix D. The control law (4.3) in Theorem 4.1 is completely optimal when Assumption 3.1 is not augmented, and it works even if $a_{max}T > y_{ref}$.

□

These two examples show that the PD-controller can be quite successful in two interesting cases. Two more examples will be given here. In spite of not being optimal the PD-controller makes good control.

<u>Example 6.3:</u> The hot plate process will be studied once more when the water quantity m_w is unknown. Use of the Laplace transform on Model (B.1) with α=0 W/°C and $T_w(0)=T_p(0)=0$°C gives

$$T_w(s) = \frac{A}{s(1+Ts)} U(s)$$

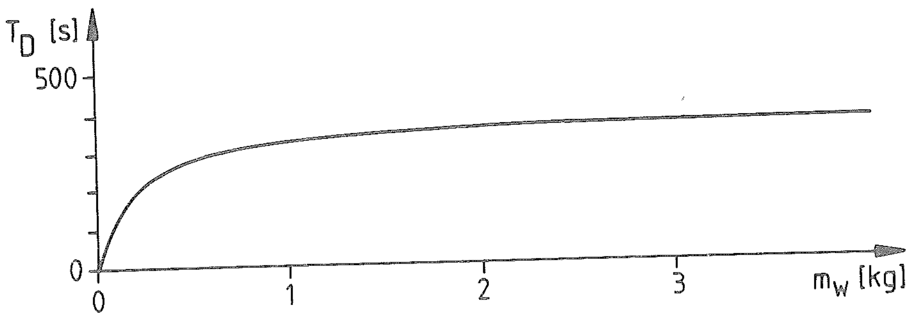

Fig. 6.1 (Ex. 6.3) $T_D = T = \dfrac{m_w C_w m_p C_p}{\Lambda(m_p C_p + m_w C_w)}$ as function of m_w.

where $A = \frac{1}{(m_p C_p + m_w C_w)}$, $T = \frac{m_w C_w m_p C_p}{\Lambda(m_p C_p + m_w C_w)}$.

With $u = u_{max}$,

$$T_w(t) = A\,[t - T(1-e^{-t/T})]\,u_{max}. \tag{6.3}$$

Assume that a PD-controller is used. The power is turned off when

$$u = K(T_{wref} - T_w - T_D \dot{T}_w) = 0 \rightarrow T_D \dot{T}_w(t) = T_{wref} - T_w(t). \tag{6.4}$$

Assuming full power until turn off (i.e. K is large) implies that the turn off time τ has to satisfy

$$T_{wref} = A \cdot \tau \cdot u_{max}. \tag{6.5}$$

Equations (6.3), (6.4), and (6.5) give

$$T_D A(1- e^{-\tau/T})\,u_{max} = T \cdot A(1-e^{-\tau/T})\,u_{max}.$$

Elimination gives $T_D = T$. Observe that

$$T \cdot \dot{T}_w(t) < T_{wref} - T_w(t) \qquad t < \tau$$

so if K is large the PD-controller will give $u=u_{max}$ until $t=\tau$ and it can be shown that $T_{wref} - T_w - T \cdot \dot{T}_w$ remains equal to zero for $t > \tau$.

If $T_D < T$ the turn off will be made too late giving an overshoot. If $T_D > T$ the turn off is made to early. The result is a slower start-up than when $T_D = T$ but no overshoot is obtained.

The above analysis have been simplified by assuming K to be infinite at suitable instances. Assuming K to be finite but large will, however, give the same result but certainly a more complicated analysis. In this example T_D depends on how much water there is in the kettle. The dependence is shown in Figure 6.1. If half a litre of water is minimum, T_D belongs to a rather narrow interval [270 s, 375 s], and in order to completely avoid a overshoot T_D should be chosen to 375 s.

□

<u>Example 6.4:</u> A better model of the hot-plate (see Section 7.1) is

$$m_w C_w \dot{T}_w = -\Lambda_{wk}(T_w - T_k)$$

$$m_k C_k \dot{T}_k = \Lambda_{wk}(T_w - T_k) - \Lambda_{kp}(T_k - T_{p1})$$

$$m_{p1} C_p \dot{T}_{p1} = \Lambda_{kp}(T_k - T_{p1}) - \Lambda_{pp}(T_{p1} - T_{p2}) + u$$

$$m_{p2} C_p \dot{T}_{p2} = \Lambda_{pp}(T_{p1} - T_{p2})$$

$$T_w(0) = T_k(0) = T_{p1}(0) = T_{p2}(0) = 0°C$$

where $m_k = 0.4$ kg $m_{p1} = 0.9$ kg $m_{p2} = 0.6$ kg $u_{max} = 1350$ W
$C_w = 4170$ Ws/Kg $C_k = 900$ Ws/Kg $C_p = 500$ Ws/kg
$\Lambda_{wk} = 10$ W/°C $\Lambda_{kp} = 3.5$ W/°C $\Lambda_{pp} = 5$ W/°C.

Assume that K is large and that when the power is turned off, it will remain off. Then T_D can be determined by simulations of the fourth order model with $u=u_{max}$. T_D is given as function of $T_{wref}=T_{inf}$ by

$$T_{inf} = \frac{t \cdot u_{max}}{m_w C_w + m_k C_k + m_{p1} C_{p1} + m_{p2} C_{p2}}$$

$$T_D = \frac{T_{wref} - T_w}{\dot{T}_w}$$

In Figure 6.2 T_D as function of $T_{inf} = T_{wref}$ is given for two cases of m_w. T_D is, in this case, not only dependent on m_w but also on the final temperature T_{inf}. Above, $T_{inf} = 10°C$. T_D is, however, nearly constant. This means that if one wants to raise the temperature with more than 10°C, T_D can be chosen independently of reference value and water quantity m_w.

It is now easy to choose the parameters for the PD-controller. K is chosen so that the controller is proportional well beyond the knees in Figure 6.2. $T_{inf} = 10°C$ gives $K \cdot T_{inf} = u_{max}$, which means that K = 135 W/°C. Figure 6.2 gives T_D = 300 s (or 375 s to be

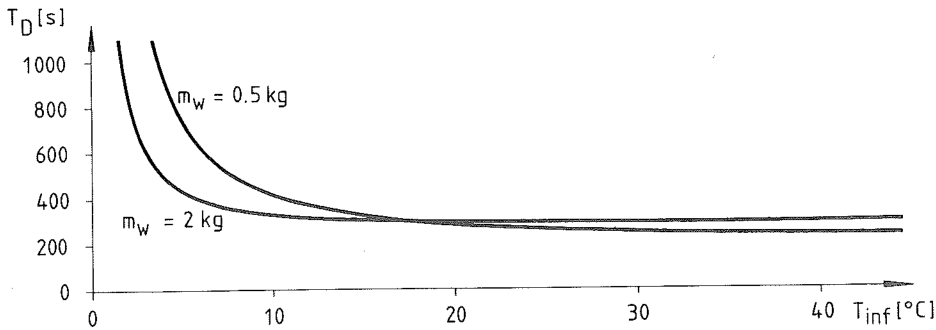

Fig. 6.2: (Example 6.4). T_D as a function of the final temperature T_{inf} for two different cases of quantity of water, m_w=0.5 and 2 kg.

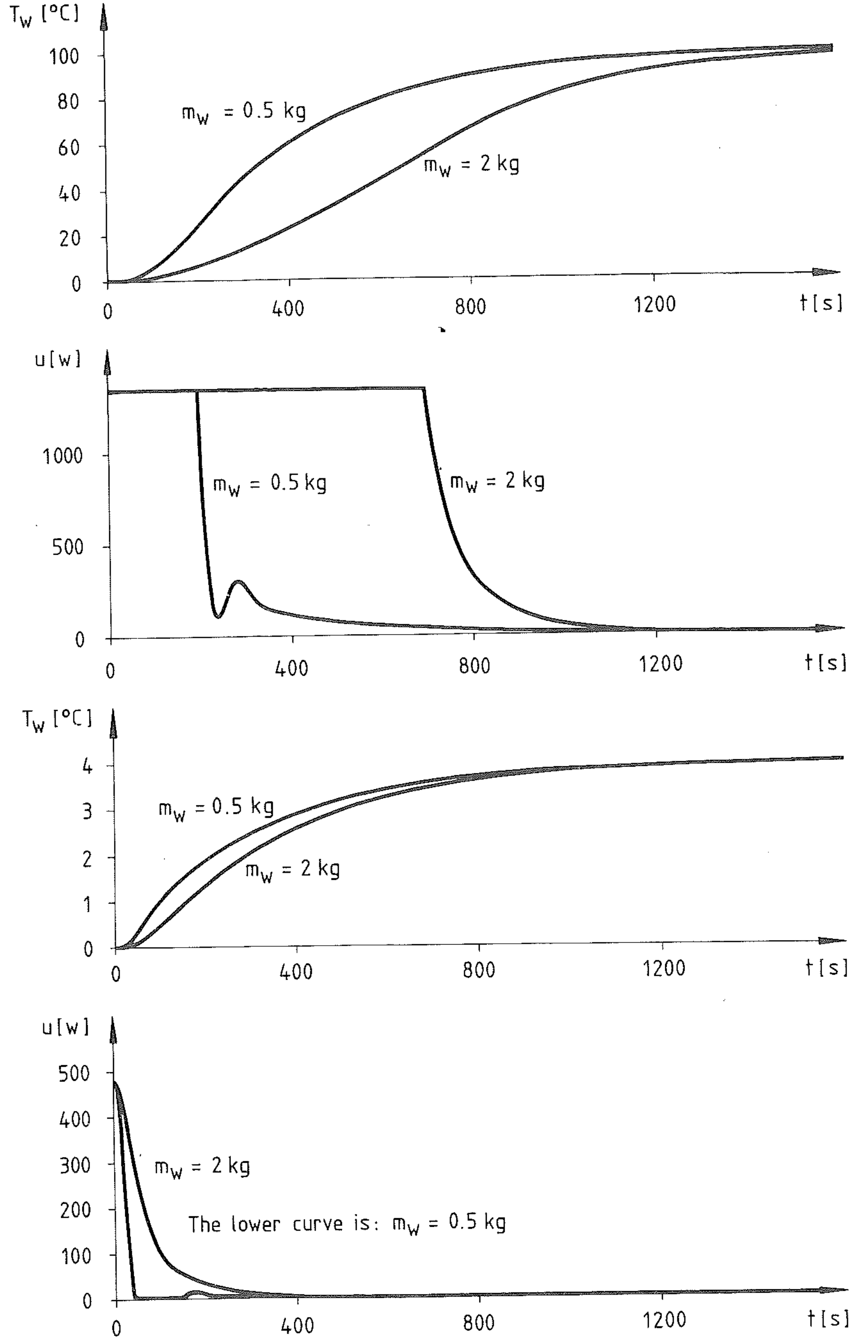

Fig. 6.3: (Example 6.4). Simulations of the fourth order system when controlled by a modified PD-controller. $K = 135$ W/°C and $T_D = 300$ s.

exact). Simulations of the closed system show good results (see Figure 6.3). The simulations show that the turn off is almost definitive. Thus the assumption made above is relevant. When $T_{wref} = 4°C$ the low gain K is responsible for the success. A high gain would have demanded a much higher T_D according to Figure 6.2. However, the control is not at all optimal. u could be held at u_{max} at least for the time

$$\tau = T_{wref} \cdot [m_w \cdot C_w + m_k C_k + (m_{p1} + m_{p2}) C_p] / u_{max} .$$

With $T_{wref} = 4°C$ and $m_w = 0.5$ kg, $\tau = 10$ s.

□

These examples show that fixed PD-controllers are optimal in some important cases and show good control in other cases. Observe, however, that the use of the PD-controller requires good feeling for how to choose the values of the parameters <u>based on good knowledge of the process</u>.

6.2 IMPLEMENTATION OF A DIGITAL PD-CONTROLLER

The basic start-up problem will be considered for the hot-plate process when $T_{wref} = 40°C$ and $\varepsilon_{ref} = 2°C$. An implementation of a PD-controller will be made and tested. When the hot-plate process is described by simple models, it is easy to find good PD-controllers. In practice there are, however, problems. First of all, features like gain-scheduling are necessary (see next section). Therefore the controller must be implemented digitally. An analog PD-controller with filter is given by the Laplace transform

$$K\left(1 + \frac{T_D s}{1 + T_D/N\, s}\right). \tag{6.6}$$

With h denoting the sample period, a discrete version of (6.6) is

$$u(n) = \alpha(y_{ref} - y(n)) + \beta u(n-1) + \gamma(y(n-1) - y(n))$$

$$\text{where} \quad \alpha = \frac{K \cdot N \cdot h}{T_D + h \cdot N}, \quad \beta = \frac{T_D}{T_D + h \cdot N}, \quad \gamma = K\,\frac{T_D + T_D N}{T_D + h \cdot N}.$$

No derivation is done on the reference value.
From the previous section it follows that reasonable values of the parameters are

$$K = 135 \text{ W/°C} \quad \text{and} \quad T_D = 250 \text{ s}.$$

Observe that the losses make it possible to use a lower T_D than could be deduced from Examples 6.3 and 6.4. What N should be is more difficult to know. The result of using no filter (i.e. $N=\infty$ in the formulas above) was a negative steady-state error of about 5°C, i.e. $T_w = y = 45$°C. The negative steady-state error depends on the discretization of the output signal. Each time the output passes down a discretization level a large positive input signal is obtained. Certainly a large negative input signal is obtained when the output passes beyond a discretization level, but negative signals have no effect on the hot-plate. With a filter, the controller will not give such large signals, but even such, there are problems with negative stationary error. The problem is solved by introducing a non-linearity in the PD-controller, such that the input u is

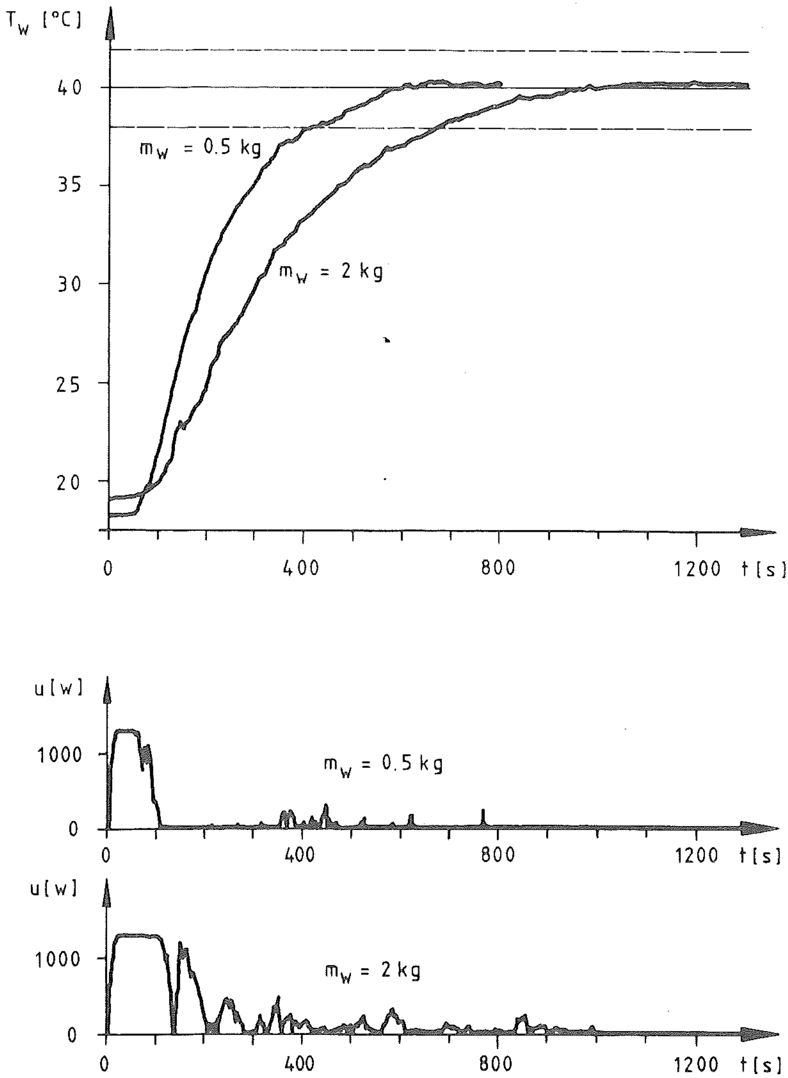

Fig. 6.4: The curves show water temperature and the control signal u, when the hot-plate is controlled by a PD-controller with K = 135 W/°C, T_D = 250 s, and N = 10.

zero when the output T_w is larger than the reference value. Another possibility had been to use integration in the controller. It is, however, difficult to avoid integral wind-up and, at the same time, assuring that the first part of the start-up is not influenced. Further, the losses are small (about 50 W) when $T_w = 40°C$, so a possibly positive stationary error is small due to the large gain. The method used is much simpler. For N=10 the hot-plate process was controlled in two cases:

Case 1: $m_w = 0.5$ kg $\quad T_{wref} = 40°C$

Case 2: $m_w = 2.0$ kg $\quad T_{wref} = 40°C$

In both cases the result was good as seen in Figure 6.4. Maybe T_D could be a little less and K a little higher, but it is difficult to make it significantly better. This can be seen in the case where m_w was 0.5 kg. The overshoot is about 0.5°C in both cases. It can thus, on good grounds, be judged that the PD-controller is admissible if $\varepsilon_{ref} = 2°C$ and $T_{wref} = 40°C$. It would maybe even be possible to have it lower.

6.3 PROBLEMS WITH A PD-CONTROLLER

The PD-controller seems to work rather well both in theory and in practice. Problems have, however, been demonstrated. One problem was the time delay of the process (see Figure 6.5). If the change $T_{wref}-T_w(0)$ is not limited from below, the gain must probably be very low. An unacceptably slow response would result in cases where $T_{wref}-T_w(0)$ is

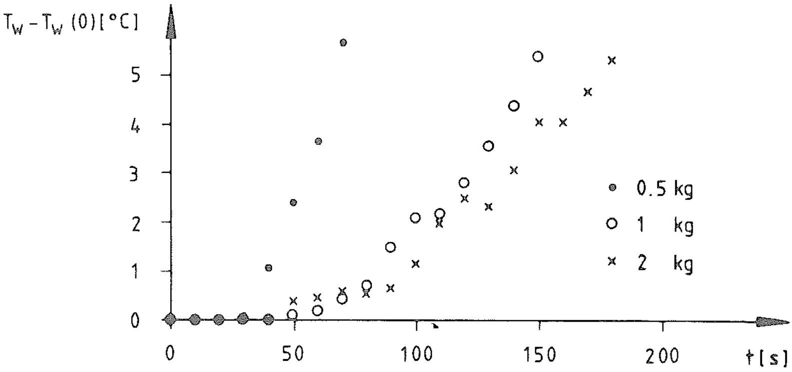

Fig. 6.5: Step responses in the cases of 0.5, 1 and 2 kg of water.

large, and for high temperatures the high losses would give a steady-state error which is too large already with K = 135 W/°C.

The PD-controller cannot handle such situations but gain-scheduling could be used. The gain would be a function of the difference $T_{wref}-T_w(0)$. If the PD-controller is implemented by a computer, this is surely possible. The real problem is to decide upon the form of the gain-scheduling. Even if gain-scheduling is used there are problems. For an iron kettle the time delay is so long that when $T_{wref}-T_w(0)$ = 10°C the gain must be lower than 35 W/°C. If a big kettle with 40 litres of water (there are such kettles!) were used, the result would be such that it takes about three hours to reach [T_{wref} −1]°C when $T_{wref}-T_w(0)$=10°C. This result is far from optimal.

Another problem is when the reference value is high and it is not known if a lid is used. Without the lid the losses are so large that the analysis made in Section 6.1 is not applicable. The high losses require a large gain or some form of reset, for instance an integrator. Simulations were made on a fourth order non-linear model (7.5) and an integrator was added to the controller. The input was, however, limited in order to prevent reset wind-up (see Åström and Wittenmark, 1982). To get an over all good performance K and T_I were chosen to K=270 W/°C and T_I=325 s.

With little water, m_w = 0.2 kg, the best value of T_D is about 32 s. If m_w = 2 kg and a lid is used (simulated by multiplying $\varphi_w(T_w)$ in (7.5) with 0.2) T_D has to be about 100 s if large overshoots are to be avoided. With m_w = 0.2 kg and no lid is used the start-up takes nearly three times longer time with T_D=100 s than with T_D=32 s.

Until now only the hot-plate has been considered. If other problems, like the one of the start-up of a batch furnace are considered, new problems turn up. In such cases there are often requirements on temperature gradients in both space and time. The gradients, in themselves, are not measurable. The use of PD-controllers in such cases, if it is possible, requires even more tailoring.

6.4 CONCLUSIONS

For "straightforward" start-up problems the PD-controller can be easy and efficient to use. However, in many practical cases much tailor made work has to be done to enable the the use of the PD-controller.

The main problem is not that PD-controllers are impossible to use but that the use of them seems to emphasize searching for tricks, which are often specific for the actual process. The main advantage with PD-controllers - the limited requirement of process knowledge - is no longer true.

CHAPTER 7

A CASE STUDY: THE HOT-PLATE PROCESS

In Chapter 5 was undertaken a general discussion about control principles for the hot-plate process using the concepts of Chapters 3 and 4. Now the results of Chapter 5 are used in making a controller for the hot-plate process.

Second and fourth order models for the hot-plate process are described in Section 7.1. The set Σ, the generation of ε, and the control law are discussed in Sections 7.2 - 7.4. In Section 7.5 an implementation of the proposed controller is described. Results are given in Section 7.6, where the controller is compared with the PD-controller used in Section 6.2.

7.1 MODELLING THE HOT-PLATE

One of the basic ideas behind the control strategies proposed in Chapter 5 is the definition of a set Σ of physical models. Such models for the hot-plate process will be presented and discussed.

To have only one state, i.e. one temperature when describing the system is inadequate. The natural subsystems are the plate, the kettle, and the water. A crucial question is if the temperatures can be considered constant in each subsystem. In such a case three states can describe the system. The different subsystems and their contact surfaces will now be discussed from a thermodynamic viewpoint. Measurement data from the hot-plate process are shown in Figures 7.1, 7.2 and 7.4 - 7.9.

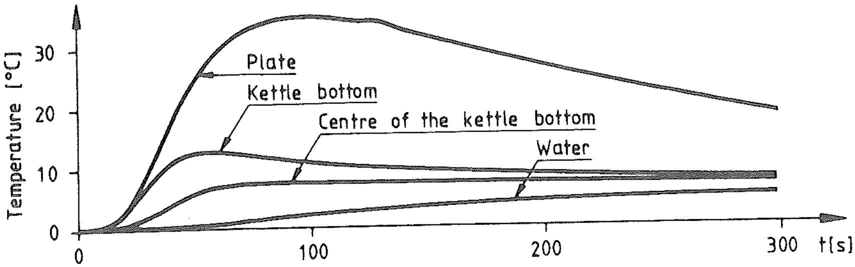

Fig. 7.1: Pulse response for the system. The pulse is 30 seconds long and the amplitude is 1350 W.

Plate:

First the temperature difference ΔT between the heat elements and the plate surface will be estimated. Assume that the maximum power, 1350 W, is transferred from the heat elements up to the plate surface. In stationarity the following equation gives ΔT.

$$\frac{\lambda_{Fe} A_{pr} \Delta T}{d_{pr}} \approx \frac{60 \cdot 0.02}{0.004} \Delta T \text{ W/°C} \approx 1350 \text{ W} \quad ==> \quad \Delta T \approx 4.5\text{°C}$$

where λ_{Fe} is heat conductivity of iron. Used value: 60 W/(m·K).

A_{pr} is the area of the heat elements. These are considered as a homogeneous thin plate. Used value: 0.02 m^2.

d_{pr} is the distance from the heat elements to the upper surface of the plate. Used value: 0.004 m.

The small ΔT and Figure 7.1 imply that only one state can represent the temperature in the plate. However, the case is made more complex by the fact that the plate is not homogeneous. The centre of the plate contains no heat elements. There is also a pit having a diameter of 55 mm. Moreover, the iron is thicker in the centre. In Figure 7.2 a pulse response for just the plate is shown. One temperature sensor was placed directly on the upper surface of the plate and another was mounted from below in the centre of the plate. This indicates that a good model requires at least two states.

The heat capacity of the plate is $m_p C_p$ = 1.5 kg ·500 Ws/kg = 750 Ws. This is small compared to the heat capacity of a normal quantity of water. The plate temperature during the critical stage of a start-up is, however, much higher than that of the water. During the critical stage of a start-up the plate will therefore contain more heat than the water. A good model of the plate is therefore important.

Kettle:

Assume that maximum power, 1350 W, is transferred through the bottom of the kettle. In stationarity the temperature fall ΔT across the kettle bottom is given by the equation

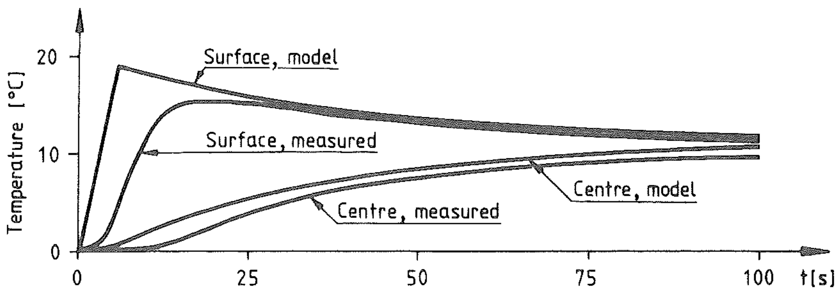

Fig. 7.2: Pulse response for the plate only. The pulse is 6 seconds long and the amplitude is 1350 W.

$$\frac{\lambda_{Al} A_{bottom} \Delta T}{d_{bottom}} \approx \frac{210 \cdot 0.025}{0.005} \Delta T \text{ W/°C} \approx 1350 \text{ W} \Rightarrow \Delta T \approx 1.2°\text{C}$$

where λ_{Al} is heat conductivity of aluminum. Used value: 210 W/(m·K)

A_{bottom} is the bottom area of the kettle. Used value: 0.025 m^2.

d_{bottom} is the bottom thickness of the kettle. Used value 0.005 m.

Figures 7.1 and 7.2 show that the temperature difference between the plate and the water must be hundreds of degrees centigrade in order to transfer the max. power through the kettle bottom. The temperature fall ΔT across the kettle bottom should thus be negligible compared with other temperature differences. Thus it seems natural with only one state representing the temperature of the kettle. On the other hand Figure 7.1 shows different temperatures at different locations on the bottom. The sensor was just placed in the water on the bottom. The electrical conductivity in water is so low compared to the conductivity in the sensor that it does not disturb the measurement. There are two reasons for different temperatures. First, the bottom of the kettle is not plane and second there is the small pit in the centre of the plate mentioned above. Consequently, the bottom ought to be described by different regions. One suitable region might be the part of the bottom in direct contact with the plate. The rest of the bottom can be the other regions. The heat transfer to the regions without direct contact with the bottom is going in a horizontal direction along the bottom. To see what this means for heat transfer, consider a square (5x5 cm) of the bottom. If the temperature difference between two sides of the square lying opposite each other is 1°C, the amount of heat passing in the square is about 1 W. This shows that a large temperature difference is required to transfer a significant amount of power. A good model of the kettle would therefore require many states. However, compared with the plate, the situation is different because the energy located in the kettle is small compared with what is located in the plate and in the water. This is seen by studying the temperature of the kettle and considering the low heat capacity of the kettle (540 Ws).

The temperature difference between the kettle and the water is significantly smaller than the difference between the kettle and the plate. This is observed in Figure 7.1. Note that the plate centre temperature shown in Figure 7.1 is lower than the surface temperature. Naturally the latter is higher in the beginning since the heat elements are located just under the surface. The heat is then propagating towards the centre. With the kettle on the plate the situation is other than it was when the temperature curves in Figure 7.2 were measured. Energy considerations show, however, that still most of the available energy content is located in the plate surface during the first part of the start-up. For the overall performance it is thus concluded that it is more important to have a good model for the plate than for the kettle.

<u>Water:</u>

During a step response, one sensor was placed in the water close to the bottom at the side and another was placed in the centre just below the water surface. After 60 seconds the first sensor showed a temperature increase of 3.7°C while the other showed only a 0.5°C increase. After some time the differences were typically less than 0.5°C. Obviously, one state does not describe the water content dynamics during the first stage of a start-up. However, the poor accuracy due to the low order model is accepted in a short time-range, because in a longer time-range the control philosophy can compensate for the inadequacies of the model. Compare with the end of Section 5.4. This means that the model error ε has to be larger than for a more elaborate model. However, by making ε depend on the input u, it is possible to use ε which will decrease after some time. Most of the start-ups will take longer time than for a more elaborate model. This is the price paid for a simpler model.

The heat transfer rates between contact surfaces

The discussion above, together with Figure 7.1, shows that it is clear that the large temperature gradients must occur in the contact surfaces plate-kettle and kettle-water. This is in agreement with Fenech et al (1964). Typically, the

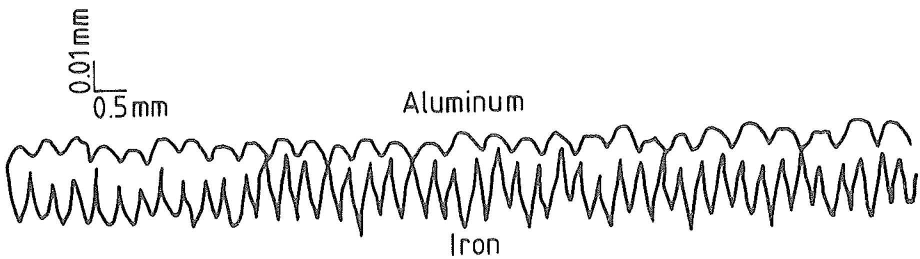

Fig. 7.3: Contact surfaces between two metals in great magnification. Modification of a figure in Fenech

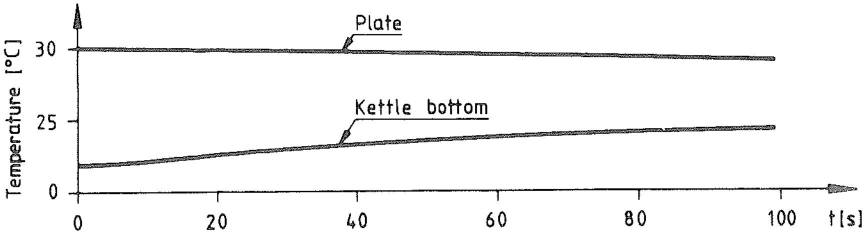

Fig. 7.4: Temperature curves when the empty kettle was placed on a warm plate. From these curves the heat transfer coefficient Λ_{kp} was estimated at 4.5 W/°C.

contact surfaces between metals can be illustrated by Figure 7.3. The contact surface plate-kettle is probably even rougher. As can be seen by the naked eye, both surfaces are turned on a lathe.

To get an estimate of the heat-transfer ratio Λ_{kp} between the kettle and the plate, the kettle was put on the plate. The temperature of the plate was about 30°C and that of the kettle about 20°C. The temperatures of the kettle bottom and the plate were measured. The result is shown in Figure 7.4. The definition of Λ_{kp} and Figure 7.4 give:

$$\Lambda_{kp}(T_p - T_{bottom}) = m_k C_k \frac{dT_{bottom}}{dt} \quad ==> \quad \Lambda_{kp} \approx 4.5 \text{ W/°C.}$$

The plate temperature was measured by the sensor mounted in the centre of the plate. The surface will have a lower temperature giving a higher Λ_{kp}. On the other hand, it is doubtful to use the whole mass of the kettle in the equation. It would probably be more correct to include only the bottom mass which gives a lower Λ_{kp}. If the max. power

(1350 W) is assumed to be transferred into the water, it would require a temperature difference of 270°C between the plate and the kettle.

In Bird et al (1960) the heat transfer coefficient between a liquid over a relatively warmer solid surface is given as.

$$100 - 600 \text{ kcal/m}^2\text{hr K} \quad \text{or} \quad 115 - 695 \text{ W/(m}^2\text{K)}.$$

This gives $0.025 \cdot 115 = 2.9 < \Lambda_{kp} < 0.025 \cdot 695 = 17.4$ [W/°C].

By following the definition of Λ_{wk} the energy inserted to the water during the time t_1 to t_2 can be expressed by the left part of (7.1). This amount of energy is absorbed by the water and can be expressed by the right hand side of (7.1),

$$\Lambda_{wk} \int_{t_1}^{t_2} [T_k(t) - T_w(t)]\, dt = [T_w(t_2) - T_w(t_1)]\, m_w C_w . \qquad (7.1)$$

The following table can be derived from Figure 7.1 and (7.1). $T_k(t)-T_w(t)$ is supposed to be constant in the interval $[t_1, t_2]$.

t_1 [s]	t_2 [s]	T_k-T_w [°C]	$T_w(t_2)-T_w(t_1)$ [°C]	Λ_{wk} [W/°C]
60	80	11	0.8	15
140	180	6	0.6	10.5
250	350	2.4	0.6	10

The achieved values of Λ_{wk} are in good agreement with the values given by Bird et al (1960). The value of Λ_{wk} is more than three times greater than that of Λ_{kp}. The kettle temperature will therefore be closer to the water temperature than to the plate temperature. Moreover, in Figure 7.1 the pulse length was shorter than usual. With a longer power pulse the temperature differences will be larger which will increase the circulation of the water. The result is a higher Λ_{wk} *. Λ_{kp} is also increasing with temperature differences but not to the same extent as Λ_{wk}, which can be seen in Figure 7.7 below.

Second order model:

To this point a second order model of the hot-plate has been used extensively. There are good reasons for not increasing the model complexity in a practical case, the main ones being:

(1) The temperature difference between the plate and the kettle dominates if short transients are disregarded. The heat capacity of the kettle is small compared to that of the water. A second order system with one state for the plate and one for the combined kettle and water is therefore natural. Compare with the situation in which the kettle's heat capacity is in the range of that of the water and the heat transfer ratio Λ_{kp} is equal to the heat transfer ratio Λ_{wk}. In such a case, a special state is required for the kettle, giving a third order system.

* Boiling water has a heat transfer coefficient of about 1000 - 20000 kcal/m^2hr K. See Bird et al (1960). Compare

(2) The parameters of a second order model are easy to identify compared to how it should be by using a higher order model. The intended identification is done off-line.

(3) It is easier to implement a control structure based on a low order model. Moreover, the actual computer has neither memory- or CPU-capacity to handle more than second order models.

The resulting second order model is:

$$\left.\begin{aligned} m_w C_w \dot{T}_w &= -\Lambda\cdot(T_w - T_p) - \varphi_w \\ m_p C_p \dot{T}_p &= \Lambda\cdot(T_w - T_p) - \varphi_p + \frac{V^2}{R}u \\ T_w(0) &= T_p(0) = T_{room} \end{aligned}\right\} \qquad (7.2)$$

where $\Lambda = \Lambda(T_p - T_w)$

$R = R(T_p)$ is the resistance of the plate

V is the voltage of the AC-net

$\varphi_w = \varphi_w(T_w - T_{room})$ is the loss from the water

$\varphi_p = \varphi_p(T_p - T_{room})$ is the loss from the plate.

The input has been expressed in a form suitable for the experimental set up. Earlier it was assumed that the power was controlled directly. On the laboratory set-up, the power is, however, controlled by a triac. The triac is an on/off device for the control of the half periods of the AC-voltage sine curves. By turning a half period on or off, the overall power may be controlled quasi-continuously. The plate is assumed to be made up of iron with C_p=500 Ws/kg. At 20°C a proper value is about 450 Ws/kg and at 400°C about 600 Ws/kg. The choice of C_p is commented upon in Section 8.2.

The second order model was identified from measurements of the loss, both for the whole system and for the separate plate. For every sampling interval of 20 seconds, the input power is defined by the computer as the number of AC voltage sine half periods turned on. When stationarity had been reached, the room, the water (in applicable cases), and the plate temperatures, as well as the heat element resistance R and the AC-net voltage V were measured.

Figure 7.5 shows measured losses for the total system φ_{tot} and Figure 7.6 for the plate φ_p. In the latter case, the plate was insulated from above. Due to effects like the heat exchange between the kettle and the iron sheet around the iron plate, the measured losses for the plate are only an

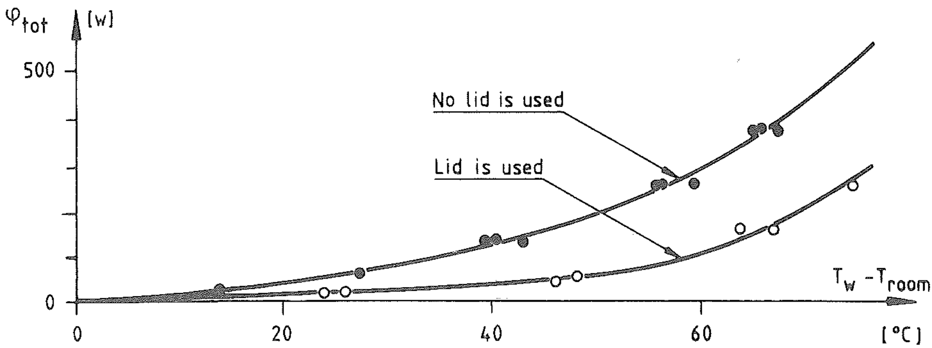

Fig. 7.5: The upper curve is the total loss when no lid is used. The lower one is the total loss when lid is used. The data points producing the curves are also given.

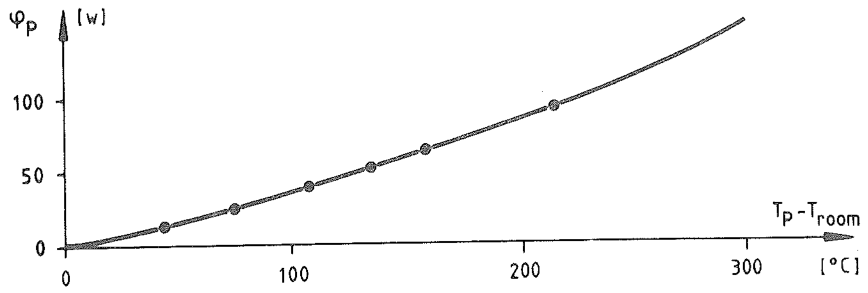

Fig. 7.6: Losses for the plate only. During the measurements an insulator was placed on the plate. The data points producing the curve are also given.

approximation of the real loss. In practice this does not matter. When total loss was measured T_p was also recorded. The room temperature was not the same in all experiments. It affects the loss which means that the loss depends on two temperatures. For low temperatures the proposed way of regarding the loss as a function of the difference is correct, but when the water is near boiling, it is incorrect. This is easily seen by assuming T_{room} to be 50°C. The water will not be warmer than 100°C even if full power is used.

$\varphi_w(T_w-T_{room})$ is defined in discrete points by

$$\varphi_w(T_w-T_{room}) = \varphi_{tot}(T_w-T_{room}) - \varphi_p(T_p-T_{room}) \qquad (7.3)$$

where $\varphi_p(T_p-T_{room})$ is given by Figure 7.6.

The heat transfer ratio can now be calculated as

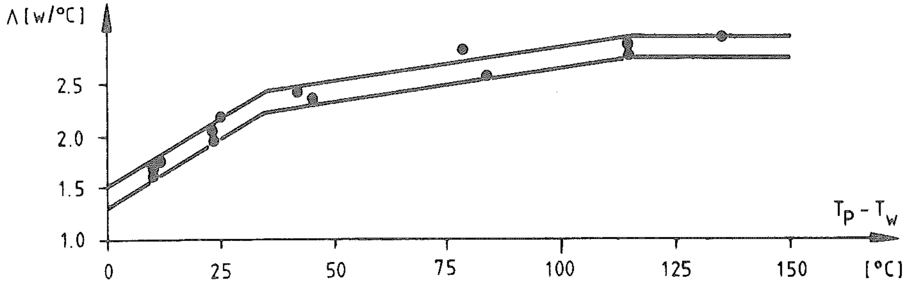

Fig. 7.7: Λ calculated from measurements. Λ is plotted versus the differences $T_p - T_w$. The full lines are used as estimates of Λ. The upper line is used in the second order and the lower in the fourth order model.

$$\Lambda = \frac{\varphi_w (T_w - T_{room})}{T_p - T_w}.$$

In Figure 7.7 Λ is plotted versus the difference $T_p - T_w$.

The heat transfer between the plate and the kettle can be separated into three parts: conductivity, convection and radiation. Heat transfer through conductivity is essentially linear as a function of the temperature difference. Convection is in classical thermodynamic theory shown (see Jacob, 1949) to be proportional to

$$(T_p - T_w)^{5/4}.$$

Heat transfer due to radiation is proportional to

$$T_p^4 - T_w^4.$$

Figure 7.7 indicates that a considerable part of the heat transfer takes place through convection. This is in agreement with the previous discussion concerning the form of the contact surface between the plate and the kettle.

A conservative estimate of Λ is given by the upper curve in Figure 7.7. In Section 7.3 it will be argued that the upper and the lower estimates shown in Figure 7.7 are satisfactory.

With the heat transfer ratio Λ given, $\varphi_w(T_w - T_{room})$ was finally calculated from the equations

$$\left.\begin{aligned} \varphi_{tot}(T_w - T_{room}) &= \varphi_w(T_w - T_{room}) + \varphi_p(T_p - T_{room}) \\ \Lambda(T_p - T_w) &= \varphi_w(T_w - T_{room}). \end{aligned}\right\} \qquad (7.4)$$

In stationarity a second order model will give the same water temperature as the true system if the total loss of the system is correctly given by $\varphi_{tot}(T_w - T_{room})$.

In Figure 7.8 the plate resistance is plotted versus T_p.

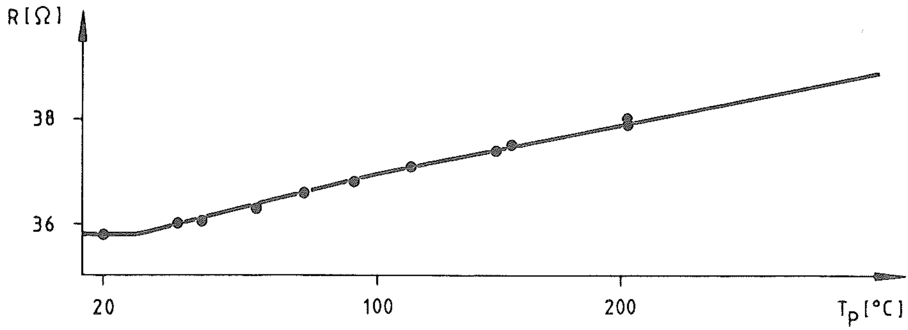

Fig. 7.8: The resistance R of the heat elements in the plate plotted versus T_p. Data points producing the curve are given.

An experiment was done on the hot-plate with $m_w = 1$ kg. The result is shown in Figure 7.9 together with a simulation of the described second order model. It is of great interest to compare the accuracy of a complex (fourth order) model with that of the simple second order model. A simulation of the fourth order model, described below, is shown. The fourth order model will be used later to estimate the modelling error ε.

Fourth order model:

The heat capacity of 0.1 kg of water is the same as that of the kettle. The kettle can therefore be represented satisfactorily by just one state. The mass of the side of the kettle is included in the water mass which reduces the model error. From now on m_k is supposed to be 0.4 kg. The remaining 0.2 kg of the kettle has a heat capacity of only 180 Ws, which is less than the heat capacity of 0.05 kg water. With two states for the plate the resulting order is four. The extension from second to fourth order is made in a straightforward way. No loss term is used for the state T_k describing the kettle temperature since T_k in fact represents only the bottom of the kettle. The main contribution to the losses from the kettle comes from the sides which are included in the water content. The second state of the plate T_{p2} also lacks loss term. This state describes the temperature of the centre of the plate. The centre part is much more compact than the rest of the plate. Consequently, it is believed that the main part of the losses for the plate is associated with the first state T_{p1}.

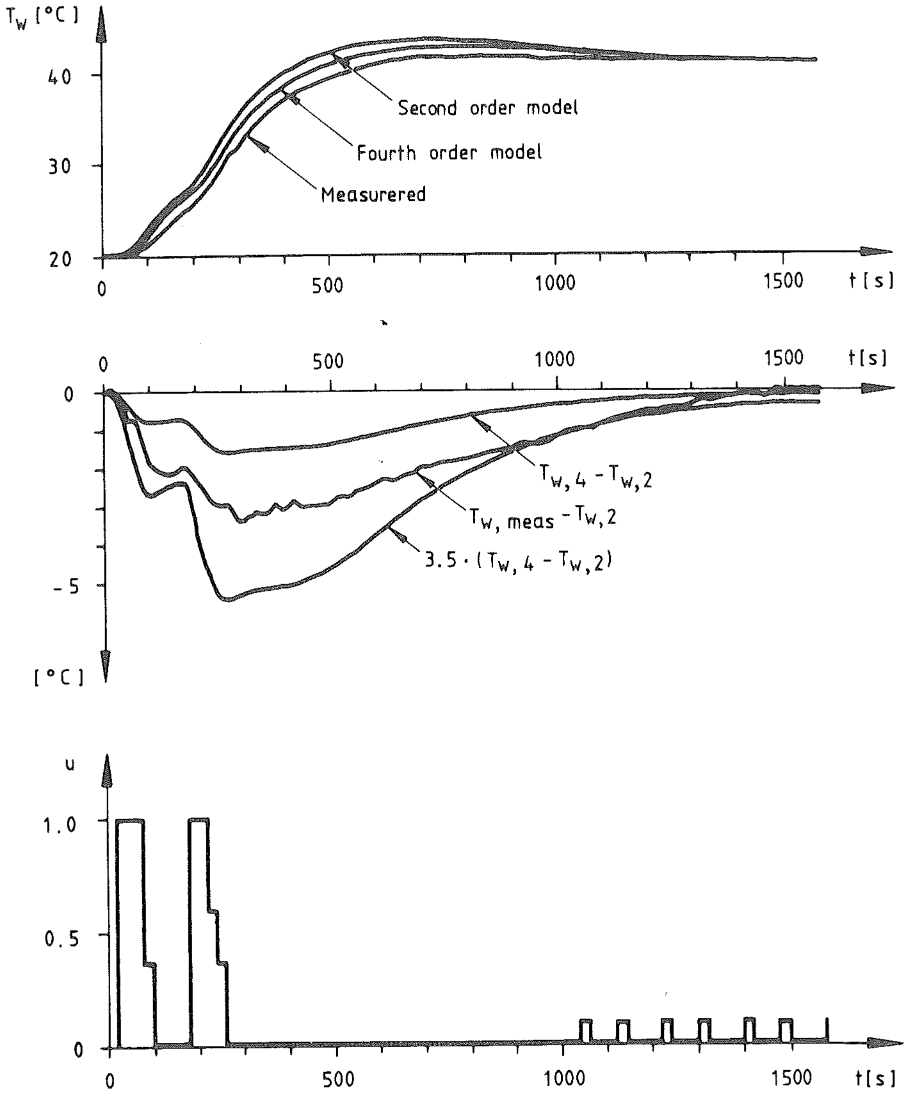

Fig. 7.9: The output from the real system ($T_{w,meas}$), the second order model ($T_{w,2}$), and the fourth order model ($T_{w,4}$). In the middle figure some differences between outputs are shown. The amplified difference $T_{w,4}-T_{w,2}$ is referred to in Section 7.3.

$$\left.\begin{aligned}
m_w C_w \dot{T}_w &= -\Lambda_{wk}(T_w - T_k) - \varphi_w \\
m_k C_k \dot{T}_k &= \Lambda_{wk}(T_w - T_k) - \Lambda_{kp}(T_k - T_{p1}) \\
m_{p1} C_p \dot{T}_{p1} &= \Lambda_{kp}(T_k - T_p) - \Lambda_{pp}(T_{p1} - T_{p2}) - \varphi_p + \frac{V^2}{R} u \\
m_{p2} C_p \dot{T}_{p2} &= \Lambda_{pp}(T_{p1} - T_{p2})
\end{aligned}\right\} \quad (7.5)$$

$$T_w(0) = T_k(0) = T_p(0) = T_{p1}(0) = T_{room}$$

where $\Lambda_{kp} = \Lambda_{kp}(T_p - T_k)$

$\varphi_w = \varphi_w(T_p - T_w)$

$\varphi_p = \varphi_p(T_p - T_w)$

$R = R(T_p)$.

By setting $\Lambda_{kp} = 0$ only the plate was simulated (see Figure 7.2). The values

$$\Lambda_{pp} = 5 \text{ W/°C}; \quad m_{p1} = 0.9 \text{ kg}; \quad m_{p2} = 0.6 \text{ kg};$$

were found to be satisfactory.

In stationarity, the total heat transfer ratio Λ from plate to water is given by

$$\left.\begin{aligned}
\Lambda(T_p - T_w) &= \Lambda_{wk}(T_k - T_w) \\
\Lambda(T_p - T_w) &= \Lambda_{kp}(T_p - T_k)
\end{aligned}\right\} \Longrightarrow \quad \frac{1}{\Lambda} = \frac{1}{\Lambda_{kp}} + \frac{1}{\Lambda_{wk}} . \quad (7.6)$$

According to a previous discussion good estimates of Λ_{wk} and Λ_{kp} are 10 W/°C and 4.5 W/°C. These values and (7.3) give $\Lambda = 3.1$ W/°C. In comparison with Figure 7.7, this seems to be somewhat too high but still reasonable.

It is crucial to have a good model accuracy in stationarity. Thus it is better to define Λ_{kp} as the solution of (7.6) for a given Λ, since Λ is known with better precision than Λ_{kp}. Λ is chosen according to the lower curve in Figure 7.7. The reason is explained in Section 7.3. The loss from the water was previously defined as depending on the heat transfer ratio Λ. Since it is not the same for the two models, φ_w has to be calculated specially for the fourth order model.

As already mentioned, a simulation of the fourth order model is shown in Figure 7.9.

7.2 CHOICE OF Ξ

In the previous section it was argued that the set Ξ should contain second order models of type (7.2) of the hot-plate process. A discretization of Ξ will be made, so the six different quantities, m_w, φ_w, φ_p, Λ, R, and V, will have to be determined. From Figure 7.9 it follows that $\varepsilon\_$ should be large during transients, given that the resolving power of Ξ is poor for the same time. From the discussion in Section 5.4 it follows that finer discretization in the value of the quantity of water m_w, which is supposed to be in the interval [0.5 kg, 2 kg], are of no use. Therefore, m_w is given for each 0.1 kg.

In Section 5.5 was discussed how a loss function could be represented. The knowledge in this case is certainly better than it was assumed in 5.5, but the same idea can be used. The resulting Ξ would, however, be too large to be used in the actual computer. It is therefore assumed that

$$\varphi_w(T_w) = \alpha\,\varphi_{nw}(T_w) \qquad 0.9 \le \alpha \le 1.1$$
$$\varphi_p(T_p) = \alpha\,\varphi_{np}(T_p)$$

where φ_{nw} = nominal function φ_w given by (7.4)

φ_{np} = nominal function φ_p given by Figure 7.6.

This implies that

$$0.9 \cdot \varphi_{ntot}(T_w) \le \varphi_{tot}(T_w) \le 1.1 \cdot \varphi_{ntot}(T_w)$$

where φ_{ntot} = nominal function given by Figure 7.5.

The inequality is not exact depending on the non-linear Λ, but in practice this fact can be neglected. From the loss data it seems enough with the proposed interval for α.

The poor resolution of Ξ during transients also means that the parameters of models in Ξ are difficult to determine except for the static gain, which is a function of α, Λ, R, and V.

Λ is difficult to determine. In principle, the ideas in Section 5.5 about loss function could be applied. This would, however, increase the amount of models in Ξ too greatly. Further, it is very difficult to get any resolution in Λ of practical value. Therefore Λ is supposed to be the upper line in Figure 7.7.

R seems to be known with good precision as a function of T_p. The function is represented by the line in Figure 7.8.

It was found that V was nearly always varied in the interval [223 V, 227 V]. Also here, the resolving power in Ξ makes it nearly impossible to get better precision in V. V is supposed to be 225 V.

<u>In summary:</u>

Ξ is chosen as the set of models defined by model (7.2), where

R is given by the curve in Figure 7.8

Λ is given by the upper curve in Figure 7.7

$V = 225$ V

φ_{np} is given by the curve in Figure 7.6

φ_{nw} is given by (7.4) where φ_{ntot} is from Figure 7.5 and φ_{np} from Figure 7.6

$\alpha \in [0.9, 0.95, 1.0, 1.05, 1.10]$

$m_w \in [0.5, 0.6, \ldots\ldots\ldots 1.9, 2.0]$ kg.

In total, 80 different second order models are used in Ξ.

7.3 GENERATING ε

The error quantity ε is used when deciding if a model is possible or not as described in Chapter 3. In Chapter 5 the main problems in determining ε were discussed. In Example 5.2 one way to determine an ε on-line was proposed. A similar means will be used here. The main part of the error is due to the low order of the model as seen in Figure 7.9. The way used to choose ε is to multiply the difference between the fourth and second order models with a constant as suggested in Figure 7.9.

The reason for the choice of Λ in the second and fourth order models is now discussed. It seems natural that a high Λ gives a high T_w. In the case of no losses, constant Λ, and a second order model, this was demonstrated in Example 5.2. Let all parameters be fixed except Λ, which is limited by the two curves in Figure 7.7. Then it is plausible that the chosen second order model gives maximal T_w among the possible second order models. In the same way it is plausible that the fourth order model gives the lowest T_w among all possible fourth order models. Furthermore, all simulations show that for equal Λ, the second order model gives a higher value of T_w than the fourth order model. Hence it can be assumed that this is true in general.

The second order model gives an upper estimate of the temperature T_w. Of course, measurement errors make it necessary to assume that $\varepsilon_+ \neq 0$. For the same reason ε_- is not allowed to decrease to zero.

$$\varepsilon_+ = 1°C$$

$$\varepsilon_- = \max[1.0, 3.5 \cdot (T_{w,2} - T_{w,4})]$$

where the subscripts 2 and 4 stand for second order model and fourth order model.

The chosen value of ε_- is optimistic as can be seen from Figure 7.9. On the other hand ε_+ is pessimistically chosen.

7.4 THE CONTROL LAW

The control law proposed in Section 5.6 will be used. This means that u(t) is chosen so that the second order model, with $u(t+j)=0$ $j\geq 1$, gives a predicted output y satisfying

$$y(s) + \varepsilon_+ \leq y_{ref} + \varepsilon_{ref} \qquad s\geq t.$$

ε_+ was chosen at 1.0°C in Section 7.3. If ε_+ were equal to ε_-, it would also have been necessary to calculate the fourth order model when calculating the predicted output with the result that an already overloaded CPU might have had a 50% load increase.

7.5 IMPLEMENTATION

The computer program implementing the control strategy is called HOTCON (HOT-plate CONtroller). In Section 7.2, the set Σ was defined. To every member there is associated a data-structure:

m_w, α	; These are the parameters which are specific for each member.
T_w, T_p	; states in the second order model
T_w, T_k, T_{p1}, T_{p2}	; states in the fourth order model.

The sampling interval is 20 seconds. The control u(t) is decided as the largest u for which no member in Ξ_{t-1} gives an overshoot. Ξ_t is not possible to use due to calculation times. It is not generally necessary to examine all the members in Ξ_{t-1}. For instance, if it is known that the model with m_w=1.0 kg and α=0.95 does not give an overshoot for u(t) then it is known that this applies to the model with m_w=1.1 kg and α=0.95. By generalizing this idea, a subset of Ξ_t called WorstCases can be defined. This set includes the only members of Ξ which need to be examined when the control signal u(t) is calculated.

A correct implementation of the control strategy requires that:

(1) time keeping is done perfectly,

(2) right amounts of power are switched on to the plate at the right time,

(3) T_w is measured at the sampling instance,

(4) the control signal u(t+1) is calculated,

(5) models are rejected in order to know the sets Ξ_t.

The timing is very important. Missing 0.1 seconds of full power each sampling period (20 seconds long) would mean that an average of about 7 W was missed. The total loss from the process when T_w = 40°C is about 40 W.

Rejection of models requires simulation of all models in Ξ_t. In the beginning of a start-up when no models are rejected, the simulation time is twice the simulated time (Euler approximations are used with Δt=2 s). To get good control under such a condition seems to be impossible. The tasks are, however, possible to perform with only one CPU because they do not require to be executed with equal precision in time. The philosophy is that the model rejection is not necessary. It only speeds up the start-up. Therefore it can be done with low priority.

By the use of concurrent programming (see Brinch-Hansen 1973) the problem is, however, solved by letting each of the 5 tasks described above be executed by its own program, subsequently called process. A process is executed by the CPU in the order of its priority. The list above is in fact ordered in priority. The processes interact with each other via monitors*. The access graph of HOTCON is shown in Figure 7.10. The program is written i Pascal using certain external procedures with which creation and synchronization of processes are done (see Elmqvist and Mattsson, 1981). The processes will be described in their order of priority. A sixth process is added that performs logging of the experiments onto disk storage.

<u>Main:</u>

In Main the other processes are created. This means that each process gets its own stack, heap, and program counter. The monitors are initialized. Main is then working as a software clock. Every 20 seconds Main tells the monitor Synch to release all processes which are awaiting in Synch.

* A monitor is a set of data and a set of procedures which are working on the set of data. A process in real time programming can be viewed as a complete program which interacts with other processes via the monitors (see Brinch Hansen, 1973).

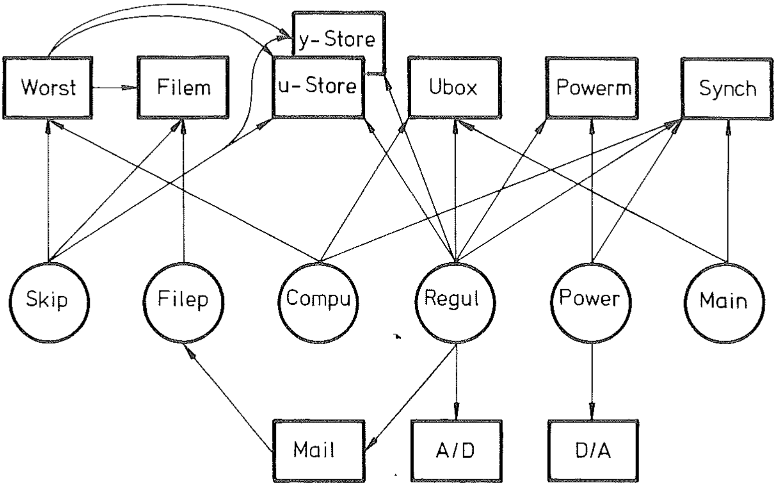

Fig. 7.10: Access graph for HOTCON. The squares are monitors and the circles are processes.

Power:

Power is awaiting in Synch for a new sampling interval. It starts with reading in PowerMon how many ticks (tick= 20 ms is the time-unit of the computer) the power is to be switched onto during the present sampling interval. The task is executed exactly. Then Power awaits in Synch for the next sampling instant.

Regul:

Regul is awaiting in Synch for a new sampling instant. It starts by measuring T_w and by writing T_w in y-store. Then Regul asks Ubox for a new input u. Regul will be awaiting in Ubox until a new u is received or one second is left of the sampling interval. If Regul is released because of the last reason, u is set to zero which for the hot-plate process is a safe value. The u-value is converted to a number of ticks (pticks=trunc(u*nticks), nticks is the number of ticks during a sampling period) which is written in PowerMon. It is also written in u-store. (In fact u=pticks/nticks is stored in order not to make a discretization error which would be 0.7 W in mean). Observe that Regul will always write the converted u-value in PowerMon in good time before the sampling instant. Consequently, it is sure that Power will switch on the power to the plate as asked by Regul. Regul also sends the u- and y-values to Filep.

<u>Compu:</u>

Compu is awaiting in Synch for the next sampling instant. Then it waits for 1 second more in order to give Skip the possibility to use the latest data when updating the set WorstCases. Compu then starts by updating the members in WorstCases and then computes the new u. u is placed in Ubox where Regul gets the value.

<u>Filep:</u>

Filep takes care of logging, both that made on screen and on disk storage.

<u>Skip:</u>

Skip calculates Σ_t and changes the set WorstCases if necessary.

7.6 RESULTS

HOTCON was tested in three cases m_w = 0.5 kg, 1.0 kg and 2.0 kg. T_{wref} was 39°C and ε_{ref} = 2°C. Observe that the control law has $T_{wref}+\varepsilon_{ref}-\varepsilon_{-}$=40°C as the target. In this way both HOTCON and the PD-controller of Chapter 6 are aiming towards the same temperature.

HOTCON gave good control in all three cases. Maximal values in the three cases were 39.8, 39.6 and 39.7°C. For the first two cases HOTCON identified, however, that m_w was 0.6 kg and 1.1-1.2 kg. This shows that ε_{-} was too small. Already in Figure 7.9 it can be seen that the choice is optimistic. The start up times and Ξ_{∞} were in the three cases:

m_w [kg]	$\Xi_{\infty}(m_w,\alpha)$	t_m [s] (37°C)	PD-controller
0.5	(0.6,1.0 -1.1)	400	348
1.0	(1.1,1.05-1.1),(1.2,0.95)	540	
2.0	(2.0,1.0 -1.1)	740	596

For comparison the corresponding times are shown when the PD-controller was used.

In Figure 7.11 the start up is showed when m_w = 2 kg. Measured data are shown. The outputs from members in Ξ_t which limit u are also shown. The index t in Ξ_t is not the time shown by the time axis. The reason is that the process Skip does not keep up with real time. The time lag is shown in Figure 7.11. In Figure 7.12 the result, when using HOTCON, is compared with the result from Section 6.2 when a PD-controller was used. There are several reasons why the

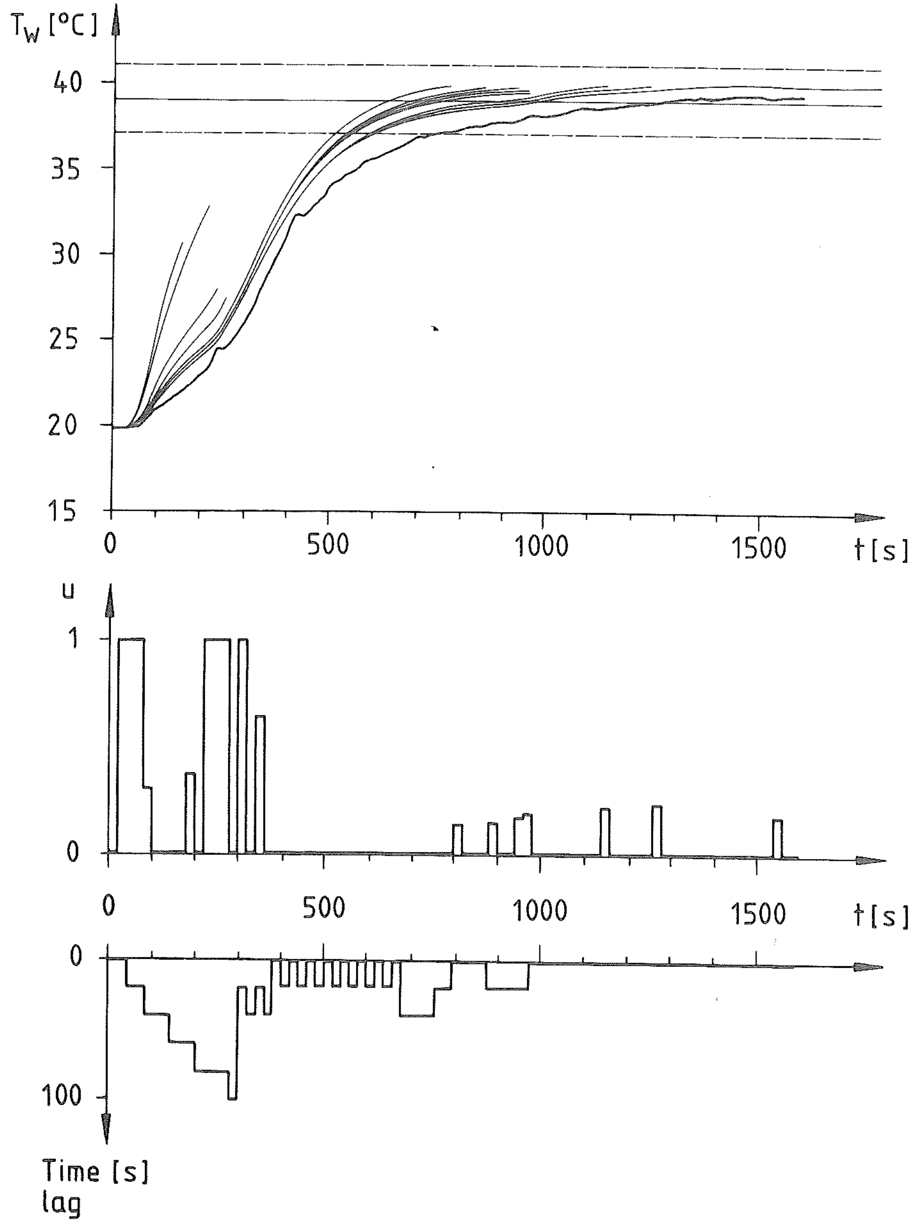

Fig. 7.11: The upper graph shows the measured output. The outputs from the models which at any time have limited u are also shown. A model output curve is discontinued when the model no longer belongs to the set Ξ_t. Ξ_t contains also models whose output curves are not shown. Their outputs are below the outputs shown in the figure. The middle part shows the input. The lower part shows time lag of the process SKIP as a function of time.

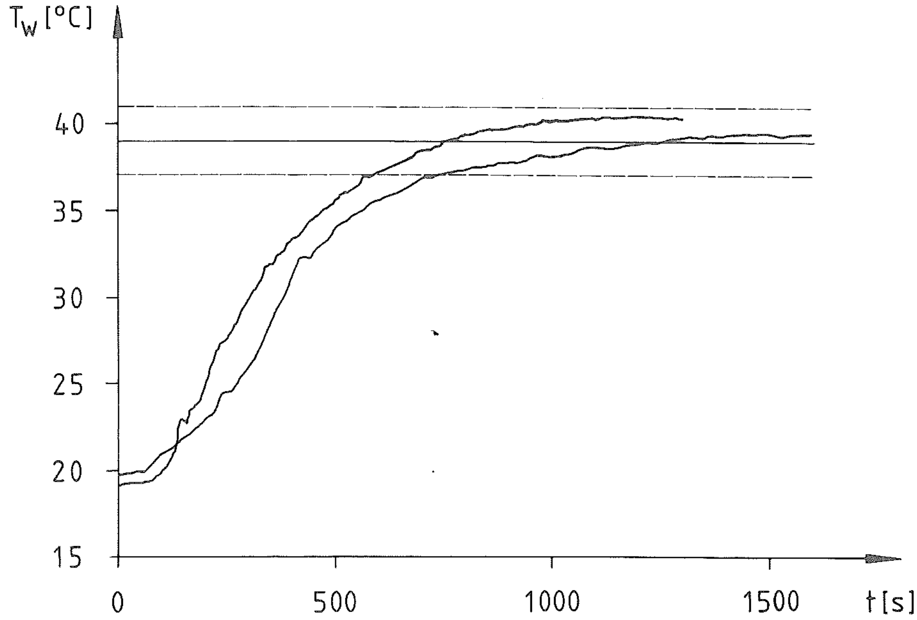

Fig. 7.12: The output for two cases of control. The upper most curve is when the PD-controller in Section 6.2 was used. The lower curve is when HOTCON was used.

PD-controller gives a better start-up.

First, the PD-controller works very well in cases when $m_w \in$ [0.5 kg, 2.0 kg] and T_{wref} is about 40°C. A perfect start-up would have been a pulse and then u=0 until the value of u becomes small in order to compensate for the loss in stationarity. In the two cases shown in Figure 6.4, the case m_w = 0.5 kg is nearly perfect. 90% of totally 122 kWs were given in the first pulse. When m_w = 2 kg, 81% of total 280 kWs was given during the first 300 seconds. The input signal is not perfect before 300 seconds either. In this case it would be possible to start-up slightly better.

Second, the computer is too slow. If Skip had kept up with real time the models could have been skipped at earlier times. The input had not been zero for such a long interval during the first 300 seconds. Larger inputs would also have given larger output signals which could have meant that models could have been skipped earlier. This, in turn, would have meant that the output could have been larger. On the other hand ε_{-} would have been larger to some extent counteracting the effect of the larger u.

Third, the second order model is not good enough if HOTCON is to compete with a PD-controller. Better models would have given smaller ε_{-} and faster skipping of wrong models. However, better models can not be used because of the low capacity of the computer.

Careful analysis of Figure 7.11 shows that a model that limits u is skipped at t=260 s but the new model limiting u is not rejected until t=780 s. In spite of that, u=0 during $t \in [280, 300]$ s and u=1 during $t \in [320, 340]$ s. The reason is that the set WorstCases contains two members from t=260 s until t=780. In such a case the calculation of the next input u takes a long time (≈20 s) and Regul does not get the value in time for t=280 s. Consequently u is zero. This would have been a bad means of real time implementation if u=0 were not safe. At the next sampling instance Regul obtains the calculated value u=1. Observe that it is safe to delay the input. Compu starts again to calculate the next input at t=301 s and again it takes too long time. So at t=320 s u is set to zero. At t=340 s u is set to 0.64. The set WorstCases is changed at t=780 s giving a non-zero input u at t=800 s.

A longer sampling period would have improved the controller considerably because in such a case the input u had not been calculated so often. Thus the Skip-process could get more CPU-time and consequently not been so lagged in time. Just the same, it is doubtful if the controller could have beaten the PD-controller when m_w=2 kg. A longer sampling interval would, on the other hand, mean that the need for real-time programming had not been illustrated so well.

When m_w = 0.5 kg HOTCON is nearly as good as the PD-controller. In this case the only problem for the HOTCON-controller is to estimate the loss. The uncertainty in the water content m_w does not hinder the controller, disregarding the need for more computing time. The least m_w in the model set is 0.5 kg.

In order to demonstrate the superiority of model based control over PD-based control another battleground has to be chosen. But the short-comings in speed and memory of the computer prevents it's application in cases where the high-gain PD-controller would have been unsuitable.

CHAPTER 8

DISCUSSION

8.1 THE DOMAIN OF DEFINITION OF A CONTROLLER

When considering the herein used controller HOTCON, it is difficult to observe any feedback. Feedback exits but in a non-conventional way. It is a priori assumed that there are 80 possible types of systems. Such a controller cannot handle disturbances like placing a lid on the kettle. Neither it can take care of slow but large parameter variations. If there are disturbances which need to be taken care of, Σ should be enlarged such that the effect of the disturbances are described by models in Σ.

Computer science recognizes the advantage of strictly defined languages which react if the stipulated rules are not followed. Even in automatic control it would be of great advantage if the domain of definition of the control algorithm were defined.

<u>Example 8.1:</u> Assume that the hot-plate is controlled by the PD-controller in Chapter 6, the reference temperature is 60 ^{o}C, and m_w = 1 kg. The start-up will go smoothly and the temperature will be nearly exactly 60°C after some time. But sooner or later the temperature will go down slowly because of water evaporating away making the temperature sensor partially exposed to air. Due to this, the PD-controller will increase power with the effect of completely exposing

the sensor to air. The effect this has on the PD-controller would be that it gives out maximal power.

□

To modify the PD-controller with an alarm device in the above example is difficult. For the HOTCON-controller the matter is simpler. An alarm is activated when the output is not fulfilling Assumption 3.1.

8.2 MODELLING BY USE OF DRAWINGS AND PARTS LISTS

The use of a physical model when doing control emphasizes the possibility of using drawings and parts lists. The physical laws describing the system are often known. So much of the basic material for making models is known, or should be known, if the control engineer has worked jointly with the construction engineer. No doubt it would be possible to let a computer make models directly from drawings and parts lists, which in the near future will already be located in the computer due to the use of CAD/CAM-systems.

In the case of the hot-plate process, it would mean that at least the model of the plate itself could have been done automatically. The contrast between such a method compared with the one used in Chapter 7 is sharp. Figure 7.2 shows that the model is far from perfect. The start-error is impossible to remedy by a second order model, but it seems to be an error, even in the heat capacity of about 5 %. Certainly, the heat elements are warmer than the surface, but another important reason is that the control engineer (i.e. the author) has considered the plate as 1.5 kg of iron. It is, however, 1.5 kg iron, kanthal and magnesium oxide. (Which is in fact only a good guess by a professional hot-plate engineer at one of the larger stove-making companies). The specific heat of magnesium oxide is nearly double in comparison to iron. So there is good reason behind the model's imperfection.

The main problem is, however, not to make the models, but to estimate the modelling error. The estimate done in Chapter 7 may be regarded as a provisory arrangement. Certainly a more systematic method is needed, especially if it is to be done by computer.

8.3 UTOPIAN IDEA OR NOT

The herein presented control ideas may seem utopian. Even a simple hot-plate could not be controlled efficiently. In a way it is only the future that can tell the answer. The primary reason why the ideas cannot be implemented is that there are presently no computers cheap enough to do the job. But there are other obstacles as well. The software used is easy to use but far from perfect. Among other things, it should be possible to describe mathematical models in a form close to the language used in mathematics and physics.

When it comes to costly processes like furnaces of batch type it may already now be possible to use the presented control ideas. Furnaces of this type should be easier to model than furnaces of continuous type. The cost of the process would admit extensive modelling and the use of powerful computers. There are also few parameters left to adapt to. This implies that the set Ξ could be rather small.

ACKNOWLEDGEMENTS

It is a pleasure for me to thank my advisors in the faculty. Professor Karl Johan Åström encouraged and helped me get started on the subject. Docent Per Hagander continuously criticized and encouraged my efforts as well as contributed to the theory. Docent Björn Wittenmark, in spite of not being particularly fond of all the definitions, pointed out weaknesses and made suggestions. Docent Gustaf Olsson helped me write different parts of the thesis, especially Section 7.1.

I also want to thank former and present colleagues at the Department for a stimulating work environment. Valuable criticism was made by Per-Olof Gutman, Tore Hägglund and Sven Erik Mattsson. The first one also pointed out that (Gutman, 1982) ought to be included among the references. Rolf Braun helped me solve experimental problems and also made the photos. Leif Andersson not only contributed with valuable criticism and helped with experiments. He also was responsible for the text-editing programs TEVE and REPORT. The figures and drawings were done by Britt-Marie Carlsson and Doris Nilsson. Irene Jönsson and Agneta Tuszynski helped me during the last stage by copying and typing.

I have had fruitful discussions with Roland Görtz at the Department of Thermodynamics, Fluid Mechanics and Heat Transfer.

A teaching assistantship at the Department of Automatic Control gave me the opportunity to start the work. Moral and financial support has been given by the Swedish Institute of Applied Mathematics (ITM). The work has also been supported by the Swedish Board for Technical Development (STU) under contract 78-3763.

Last, I want to thank my family Cristina, Frida, and Jonatan for patience as well as valuable suggestions concerning my treatment of time and space problems.

REFERENCES

Åström K.J. (1981): Theory and applications of adaptive control, Plenary lecture at the 8th IFAC Congress in Kyoto.

Åström K.J., Hagander P., Sternby J. (1980): Zeroes of sampled systems. 19th IEEE Conference on Decision and Control, Albuquerque, USA.

Åström K.J., Wittenmark B. (1982): Computer Control Theory. Lecture notes, Lund Institute of Technology, Lund, (To be published by Prentice-Hall).

Bird R. Byron, Stewart W.E., and Lightfoot Edwin N. (1960): Transport Phenomena. See Table 13.1-1. John Wiley & Sons, New York.

Brinch Hansen Per. (1973): Operating system principles. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Elmqvist H. and Mattsson S.E. (1981): A Real-Time Kernel for Pascal. Department of Automatic Control, Lund Institute of Technology, Lund, Sweden. CODEN:LUTFD2/(TFRT-7231)/1-023/(1981).

Fel'dbaum A.A. (1965): Optimal control systems. Academic Press, New York.

Fenech H., Henry J.J. and Rohsenow W.M. (1964): Thermal Contact Resistance. Article in Development in Heat Transfer, editor Rohsenow .W. M. The M.I.T. Press, Cambridge, Massachusetts.

Goodwin G.C., Payne Robert L. (1977): Dynamic System Identification: Experiment Design and Data Analysis. Academic Press, New York.

Gutman Per-Olof (1982): Application of linear programming for on-line control. Part 3 in "Controllers for bilinear and constrained linear systems", Ph.D. dissertation, Dept. Auto. Contr., Lund Inst. of Techn., Lund Sweden.

Gyllensköld H. (1977): Koka, Steka, Blanda. (in swedish) Wahlström & Widstrand, Stockholm, Sweden.

Jacob M. (1949): Heat Transfer. John Wiley & Sons, New York. (Eight printing May 1968).

Saridis George N. (1970): Self-Organizing of Stochastic System. Marcel Dekker, Inc. New York, USA.

Sworder David (1966): Optimal Adaptive Control Systems.
Academic Press, New York.

APPENDIX A

LABORATORY EQUIPMENT

In Figure 2.1 the arrangement of the laboratory equipment is shown. Some data and references about the equipment are given here.

Temperature sensors: (Ni 100, SWEMA). Two types are used. They are specially shown in Figure A.1. The sensor to the right, placed in a cylinder, is used when measuring temperatures in water. The other one is used when surface temperatures are measured.

Hot-Plate: (Electro Standard AB, type number 214, 220 V, 1200 W). The weight of the iron-plate is about 1.5 kg. From below a temperature sensor is mounted in the centre of the iron-plate. This sensor is used for the purpose of making

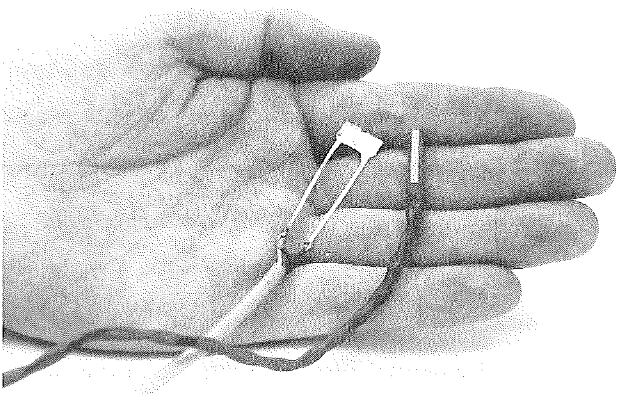

Fig. A.1: The temperature sensors used.

models of the hot-plate. It is not used in the control algorithms.

Control Unit: In the power control a triac is used as a contact which is open or closed the next half sinus period depending on the control signal to the triac.

Computer: (DEC, PDP-11/03). 56 kbytes primary memory and floating point arithmetic in hardware (instruction timing: add/sub≈50μs, mult≈100μs, and div≈200μs).

APPENDIX B

SIMPLE PHYSICAL MODEL FOR THE HOT-PLATE PROCESS

The thesis is illustrated by the use of variants of a second order model of the hot-plate process.

<u>Model B.1:</u>

$$\left.\begin{aligned} & m_w C_w \dot{T}_w = -\Lambda(T_w - T_p) - \alpha(T_w - T_{room}) \\ & m_p C_p \dot{T}_p = \Lambda(T_w - T_p) + u \\ & T_w(0) = T_p(0) = 20°C \\ & u \in [0, u_{max}] \end{aligned}\right\} \qquad (B.1)$$

where

m_w is the mass of the water. The usual range [0.5, 2.0] kg.

m_p is the mass of the plate. Used value: 1.5 kg.

C_w is the heat capacity of water. Used value: 4170 Ws/°C.

C_p is the heat capacity of the plate. Used value: 500 Ws/°C.

T_w is the temperature of the water.

T_p is the temperature of the plate.

T_{room} is the surrounding temperature. Used value is 20°C if nothing else is stated.

Λ is the heat transfer ratio defined by $\Lambda \cdot (T_p - T_w)$ equal to transferred power from plate to water. The value is in the range [1.5, 3.0] W/°C. Used value is 2 W/°C if nothing else is stated.

α is a coefficient. $\alpha \cdot T_w$ is considered as the loss from the process. At 95°C the losses are about 500 W when no lid is used.

u is the input power to the process. $u_{max} \approx 1350$ W.

To get shorter notations the following parameters are often used:

$$c_w = m_w C_w \quad \text{and} \quad c_p = m_p C_p.$$

Variants of Model (B.1) will be used throughout the thesis. Some variants are so simplified that they are not suitable for practical control. Actually the loss term in (B.1) is not possible to use. The loss is non linear, and a linear approximation is not possible to use when doing practical control. Nor is the heat tranfer ratio Λ linear. In practice it is necessary to let it be a non linear function of the temperature difference $T_p - T_w$.

The dynamics of the kettle can be neglected in a simple model. This is in fact possible to do even in practice depending on the low heat capacity (540 Ws) of the kettle and the good heat conductivity of aluminium (210 W/°C). The kettle can also be thought of as part of the water, because the heat transfer ratio between the kettle and the water (about 10 W/°C) is much higher than that between the plate and the kettle (1.5 to 3.0 W/°C).

APPENDIX C

DEFINTION OF RESOLVING POWER

In Section 5.4 the resolving power of Ξ was mentioned without any definition. Here a proposal for such a definition is made. The definition is made in two steps.

<u>Definition 5.1:</u> Let α be a physical parameter in the models that generate the members of Ξ. $\Xi(\alpha_1)$ denotes all models of Ξ which have $\alpha=\alpha_1$. α_1 is said to be separable from α_2 if there is an admissible control law F and a $t\geq 0$ such that for all $u \in S_{t-1}$ and $\eta \in S_t$

$$F \in \Phi_{t,u,\eta} \rightarrow \{\Xi_{t,u,\eta} \cap \Xi(\alpha_1) \cap \Xi(\alpha_2)\} \text{ is empty.}$$

□

<u>Definition 5.2:</u> The resolving power of Ξ is the ability to separate physical parameters accociated with the process.

□

PPENDIX D

HE OPTIMALITY OF THE PD-CONTROLLER OF EXAMPLE 6.2

y augmenting the Assumption 3.1 by a requirement on the erivative

$$|y(t) - H(u)(t)| < \varepsilon_0$$
$$|y'(t) - H(u)'(t)| < \varepsilon_1,$$

t is possible to show that the PD-controller of Example 6.2 s completely optimal. Of course differences should be used nstead of derivatives. Ordinary derivatives make it owever, easier to see the essential points. Choose h, ε_0, $_1$, d, and K such that

$$\left.\begin{array}{ll} \text{(i)} & d + 2\varepsilon_0 < \varepsilon_{ref} \\ \text{(ii)} & \varepsilon_0 + T\varepsilon_1 < y_{ref} - a_{max}T \\ \text{(iii)} & 2\varepsilon_0 + 2T\varepsilon_1 + 1/K + a_{max}h < d \end{array}\right\} \tag{D.1}$$

efine t_m as the first time such that $u(t_m) < 1$. Assume first hat $y(t_m) < y_{ref} - d$. A lower estimate of t_m is then given by

$$K[y_{ref} - a\,t_m - T) - \varepsilon_0 - T(a+\varepsilon_1)] < 1. \tag{D.2}$$

onsequently a lower estimate of the output at $t=t_m+T-h$ is

$$(t_m - h) - \varepsilon_0 \geq y_{ref} - 2\varepsilon_0 - T\varepsilon_1 - 1/K - a\cdot h \geq y_{ref} - \varepsilon_{ref}.$$

Let t_u be such that $u(t)=0$ if $t \geq t_u$ and $u(t_u-h)>0$. Assume that $aT \geq d$. The case when $aT<d$ is taken care of below. From (D.2), (D.1), and $aT \geq d$ it follows that

$$a(t_m-h) - \varepsilon_0 + T(a-\varepsilon_1) \geq y_{ref} - 2\varepsilon_0 - 2T\varepsilon_1 - 1/K + aT \geq y_{ref}$$

This implies that

$$t_u \leq t_m + T - h \qquad (D.3)$$

If t_u-T is small it follows from (D.1) that

$$a(t_u-T-h) + \varepsilon_0 + T(a+\varepsilon_1) < y_{ref} \qquad (D.4)$$

From (D.3) and (D.4) it follows that an upper estimate of t_u is given by

$$K[y_{ref} - a(t_u+T-h) + \varepsilon_0 - T(a-\varepsilon_1)] > 0$$

Then an upper estimate of the output is

$$at_u - h + \varepsilon_0 = y_{ref} + 2\varepsilon_0 + T\varepsilon_1 \leq y_{ref} + \varepsilon_{ref}$$

If $y(t_m) \geq y_{ref}-d$ then it follows from (D.1) that for $s \geq t_m$ $y(s)>y_{ref}-\varepsilon_{ref}$. It will now be proved that a possible overshoot is not to large. From (D.1) if follows that

$$y(t_m-h) \geq y(t_m) - 2\varepsilon_0 - ah \geq y_{ref} - d - 2\varepsilon_0 - ah > 0.$$

This and $u(t_m-h)=1$ imply that

$$K[y_{ref} - a(t_m-h-T) - \varepsilon_0 - T(a+\varepsilon_1)] \geq 1$$

Thus

$$y_{ref} - d \leq a(t_m-T) + \varepsilon_0 \leq y_{ref} - ah - T\varepsilon_1 - 1/K - aT$$

Hence $aT+ah \leq d$. In such a case it follows for all s that

$$y(s) \leq H(u)(s) + \varepsilon_0 \leq H(u)(t_u - h) + aT + \varepsilon_0 \leq$$

$$\leq y_{ref} - d + \varepsilon_0 + aT + \varepsilon_0 \leq y_{ref} + \varepsilon_{ref}.$$

Thus the proposed variant of PD-controller is completely optimal.

□