New Estimation Techniques for Adaptive Control

Hägglund, Tore

1983

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

New Estimation Techniques For Adaptive Control

Tore Hägglund

Organization
LUND UNIVERSITY

Department of Automatic Control
Box 725
S 220 07 Lund 7 Sweden

Document name
DOCTORAL DISSERTATION

Date of issue December 1983

CODEN: LUTFD2/(TFRT-1025)/1-120/(1983)

Author(s)
Tore Hägglund

Sponsoring organization
The Swedish Board for Technical Development

Title and subtitle

New Estimation Techniques for Adaptive Control

Abstract

An essential part of an adaptive controller is the routines for estimating the process dynamics. When the adaptive controller is applied to time-variable systems, old information must be discounted in favour of new information to track the time-variations.

This thesis treats the problem of adaptive control applied to time-variable systems. The problem of estimating time-variable parameters is discussed from an information handling point of view. The time-variations are assumed to belong to two categories: Large parameter changes, which occur infrequently, and slow parameter changes. These categories are analysed separately in the thesis.

A new fault detection procedure is presented and shown to meet the requirements of the adaptive control problem. The large parameter changes are handled by modifying the estimation scheme whenever the changes are detected by the fault detector.

A new technique to handle slow parameter variations and variations in the excitation is also presented. By the new method, old information is discounted in such a way that a constant desired amount of information is retained in the estimator. This approach is shown to solve the standard problems caused by poor excitation.

The two methods are combined and shown to work well on problems with both abrupt and slow parameter changes.

Key words
Adaptive control, Parameter estimation, Fault detection
Recursive identification, Time-varying systems.

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language English

ISSN and key title

ISBN

Recipient's notes

Number of pages 120

Price

Security classification

Distribution by (name and address)

PREFACE

This work has been carried out at the Department of Automatic Control in Lund. The most stimulating research atmosphere at the department has been of great importance for my work, and now I would like to take the opportunity to thank all my colleagues. The following persons are particularly mentioned for their contributions to my thesis.

Especially, I would like to thank my supervisor Professor Karl Johan Åström. He suggested the problem and provided an excellent guidance throughout the work. His vast knowledge in the area of adaptive control has been of great support.

I am also indebted to Per Hagander for his careful reading of several versions of the present material, and for suggesting many improvements of the results.

I have had many fruitful discussions with Carl Fredrik Mannerfelt, Sven Erik Mattsson, Lars Nielsen, Rolf Johansson and Björn Wittenmark, and I also appreciate their valuable criticism of the manuscript.

It is also a pleasure to thank Leif Andersson for his excellent document preparation programs, and Britt-Marie Carlsson and Doris Nilsson for carefully preparing the figures.

Finally, the financial support given by the Swedish Board for Technical Development (STU), under the contracts 78-3763 and 82-3430, is gratefully acknowledged.

Lund in December 1983
Tore Hägglund

CONTENTS

1. INTRODUCTION

An automatic control problem consists of an analysis of the process to be controlled followed by a design of a controller, based on the analysis. This procedure is mostly carried out repeatedly, i.e. the design is followed by a new analysis of the received closed loop system leading to modifications of the design, and so forth.

In the simplest control problems, the procedure is just a "trial and error" procedure. A controller is inserted in the feedback loop, and the behaviour of the control is analysed. The parameters of the controller are then adjusted and the new behaviour analysed etc. This simple tuning procedure is commonly used for PID control. It is suitable for processes which are rather easy to control combined with controllers with no more than two or three adjustable parameters.

The simple tuning procedure is however not adequate for more difficult control problems. A more advanced analysis is required, and often also a more advanced controller design. Much progress have been made in creating useful analysis tools and control design concepts during the last decades. The analysis may e.g. contain a model building from physical considerations, step-, impulse- or frequency response analysis or parameter estimation. The design is then based on the obtained model. Among the design methods, pole placement, linear quadratic gaussian and minimum variance control, Kalman filtering combined with state feedback and robust control can be mentioned.

The effort made in the analysis and the design of a control loop does not only depend on how difficult the loop is to control. The skill of the engineer, time-schedules and economical realities are often causing serious constraints. A thorough analysis and design is therefore nowadays made only in very complicated control problems or in control loops which are produced in large series.

Most progress in automatic control have been made concerning linear time-invariant systems. Unfortunately, most difficult control problems arise when the system is both nonlinear and time-varying.

These are the main reasons why the concept of adaptive control has met such interest in automatic control research. The idea of adaptive control is to automate the analysis and the design. Hence, an adaptive controller does not only control the process, but it also collects information about the process behaviour from the control signals and the measured output signals. Based on this information, or analysis, the controller parameters are adjusted on line.

The notion of adaptive control originated in the early fifties, but the main research concerning modern types of adaptive controllers have been made in the last decade. Several important theoretical results concerning stability and convergence have been reported in the literature as well as successful implementations. A résumé is given in Åström (1983).

Most of the theoretical work has been devoted to asymptotic properties such as convergence, while much less attention has been paid to the transient behaviours of the adaptive control system. The analysis is mostly carried out under the assumption that the system parameters are constant, while the main purpose of the adaptive controller is the ability to adapt to time-varying systems. The reason is, that the general adaptive control problem, being both nonlinear, time-varying and stochastic, is so complicated. It has not yet been possible to derive theoretical results apart from in rather restricted cases.

Nevertheless, the ability to handle time-varying parameters is a key problem, which often shows up in the applications. It is therefore not surprising, that most attempts to solve the problem is found in the application literature. These solutions are often based on heuristic arguments.

This thesis is devoted to the adaptive control problem of time-variable systems. In Chapter 2, the adaptive control concept is explained in more detail, and basic notations are introduced. Special problems that arise due to the time-variability are also stated. These problems occur in the part of the controller which performs parameter estimation. In Chapter 3, the estimation algorithm is therefore analysed. In Chapter 4, a special type of time-variability is examined, namely large parameter changes, or faults, of the process. A new fault detection procedure, which satisfies the special requirements for adaptive control, is presented. Chapter 5 treats the general

problem of moderate variations of the process parameters as well as changes in the excitation of the process. A new solution is presented and analysed. Together, Chapters 4 and 5 capture a large class of time-variations. Chapter 6 summarizes the new results of Chapters 4 and 5 by a simulation study of an industrial robot. Conclusions and references are given in Chapters 7 and 8 respectively.

2. PRELIMINARIES

In this chapter, the adaptive control concept is further explained. The basic notations and limitations of the problem treated in the thesis are given. The special problems related to time-variable systems are discussed.

2.1 Principles of adaptive control

As was said in Chapter 1, an adaptive controller not only controls the process, but it also collects information about the dynamics of the process and the disturbances, and adjusts the controller parameters based on this information. This can be done in many ways. So far, two schools have dominated the adaptive control literature, namely the model reference approach (MRAC) and the self-tuning regulator (STR). These and other approaches are summarized in Åström (1983). It has been shown in Egardt (1980), that the MRAC and STR are in principle just two ways of looking at the same thing. An adaptive controller designed from an MRAC point of view can be reformulated in STR concepts and vice versa. In this thesis, the STR approach will be followed.

The structure of a process controlled by a self-tuning regulator is shown in Figure 2.1. The adaptive controller consists of three parts. The regulator (R) together with the process forms the ordinary feedback loop. The input and the output of the process are collected in a parameter estimator (E) which performs an on-line estimation of the process parameters. These parameters are used to modify the controller parameters via a design calculation (D). The parameter estimator and the design calculation form the adaptation loop. The different parts of the system will now be explored in more detail.

The process (P)

Throughout the thesis, it will be assumed that the process can be described by the model

$$y(t) = \theta(t-1)^T \varphi(t) + e_n(t) \qquad (2.1)$$

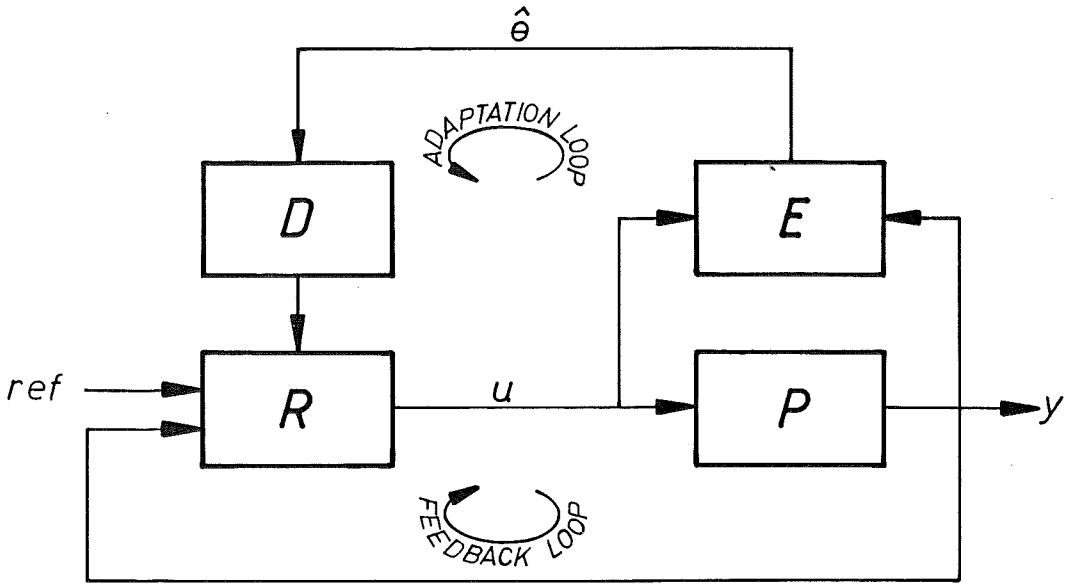


Figure 2.1 - The structure of a process controlled by a self-tuning regulator. The four blocks denote the process (P), the regulator (R), the estimator (E) and the design calculation (D).

where $y(t)$ is the measured output from the process, $\varphi(t)$ is a vector containing old inputs and outputs of the process, $\{e_n(t)\}$ is a disturbance sequence of independent random variables and $\theta(t)$ is a parameter vector. Furthermore, it will be assumed that the disturbances $\{e_n(t)\}$ have a symmetrical probability distribution.

Figure 2.1 makes an implicit assumption about separation of the process variables into two categories, parameters $\theta(t)$ and states $\varphi(t)$. In the area of adaptive control, it is assumed that the parameters fulfil at least one of the following requirements.

1. The parameters $\theta(t)$ are constant but unknown.

2. The parameters $\theta(t)$ vary slowly compared with the states $\varphi(t)$.

3. The parameters $\theta(t)$ are subject to sudden large changes, which are infrequent compared with the time-constants of the system.

If this separation between the parameters and the states cannot be made, the control problem has to be solved using other nonlinear control methods. In Chapter 3, the restrictions on the time-variability of the process parameters $\theta(t)$ will be further discussed.

The model (2.1) assumes that the output is a linear combination of past inputs and outputs. The self-tuning regulator may however be suitable even for processes with a not too strong nonlinearity, since the estimator may produce a linearization of the nonlinear model at the actual operating point. The assumption of white noise disturbances is not very restrictive. It is made just for convenience. If the noise happens to be coloured, a modified estimation procedure can be used as is described below.

The parameter estimator (E)

The parameter vector $\theta(t)$ in Equation (2.1) is to be estimated from past input and output signals. It can be done in numerous ways, and a vast literature has been devoted to this estimation problem. Some common methods are the least squares, the maximum likelihood, the stochastic approximation and the instrumental variables methods. A summary of different estimation methods is given in Ljung and Söderström (1983). In this thesis, the most common method, the recursive least squares algorithm, will be considered. The results are however in no way limited to this method. If e.g. the noise acting on the process is coloured, another method such as the extended least squares, the generalized least squares or the maximum likelihood method may be used.

In the original least squares algorithm, the parameter vector $\theta(t)$ is estimated by the following equations:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\varphi(t)\varepsilon(t) \qquad (2.2a)$$

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{1 + \varphi(t)^T P(t-1)\varphi(t)} \qquad (2.2b)$$

$$\varepsilon(t) = y(t) - \hat{y}(t) = \left[\theta(t-1) - \hat{\theta}(t-1)\right]^T \varphi(t) + e_n(t) \triangleq$$

$$\triangleq \tilde{\theta}(t-1)^T \varphi(t) + e_n(t) \qquad (2.2c)$$

Here $\hat{\theta}(t)$ is the estimate of $\theta(t)$, $\hat{y}(t)$ is the prediction of $y(t)$ made at time $t-1$, and $\tilde{\theta}(t)$ is the estimation error at time t.

In case of time-variable parameters, the above equations are however useless. Equation (2.2b) has to be modified. This problem and other aspects of the least squares algorithm will be discussed further in the next section and especially in Chapter 3.

The design calculation (D)

The self-tuning regulator concept is not limited to any particular design method, but any of the traditional methods may be used. It is desirable to avoid making the design calculation on line, since it often has bad numerical properties. This is possible by reformulating the process model in terms of the controller parameters. The desired controller parameters are then estimated directly. See Aström (1983).

The regulator (R)

No particular regulator structure is assumed in this thesis. The control signal is supposed to be a causal function of the estimated parameters and their covariances, and of the input and output signals, i.e.

$$u(t) = f\left[\hat{\theta}(t), P(t), y(t), y(t-1), \ldots, u(t-1), u(t-2), \ldots\right] \tag{2.3}$$

2.2 Difficulties in time-varying systems

The estimator derives a model of the plant from input-output data. The least squares estimate given by Equation (2.2), can be interpreted as the estimate which minimizes the loss function

$$J(\hat{\theta}) = \sum_{i=1}^{t} \left[y(i) - \hat{\theta}^T \varphi(i)\right]^2 \tag{2.4}$$

It follows from this interpretation that the same weight is put on every measurement. This is reasonable if both the parameters to be estimated and

the variance of the measurement disturbances are constant. If the plant to be identified is time-varying, old input-output pairs may however not be relevant for the actual model. Their influence on the estimate should therefore be reduced. If the variance of the measurement disturbances is varying, a weight inversely proportional to the variance should ideally be put on the measurements.

A common way of discounting old data is to use a forgetting factor (λ). Equation (2.2b) is then modified to

$$P(t) = \frac{1}{\lambda}\left[P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{\lambda + \varphi(t)^T P(t-1)\varphi(t)} \right] \qquad (2.5)$$

This corresponds to exponential weighting of past data, since a measurement received n samples ago has a weight proportional to

$$\lambda^n = e^{n \cdot \ln(\lambda)} \qquad 0 < \lambda \leq 1 \qquad (2.6)$$

The choice of λ is a trade-off between fast adaptation and long term quality of the estimates. This trade-off can sometimes be unsatisfactory. It may be desirable to discount quickly when the model is changing rapidly, and to discounting slowly when the parameters are constant or the excitation is poor. The problem is illustrated in Example 2.1.

Example 2.1: Consider the process model

$$y(t+1) = \alpha(t) \cdot y(t) + e(t+1) \qquad (2.7)$$

where y(t) is the output signal and {e(t)} is a white noise sequence. The parameter $\alpha(t)$ is estimated according to Equations (2.2a), (2.2c) and (2.5). Three different values of the forgetting factor λ are used, namely $\lambda = 0.95$, $\lambda = 0.99$ and $\lambda = 0.995$. The results are shown in Figure 2.2. At time t = 100, the parameter $\alpha(t)$ changes from -0.9 to -0.3. The trade-off in the choice of λ is obvious.

□

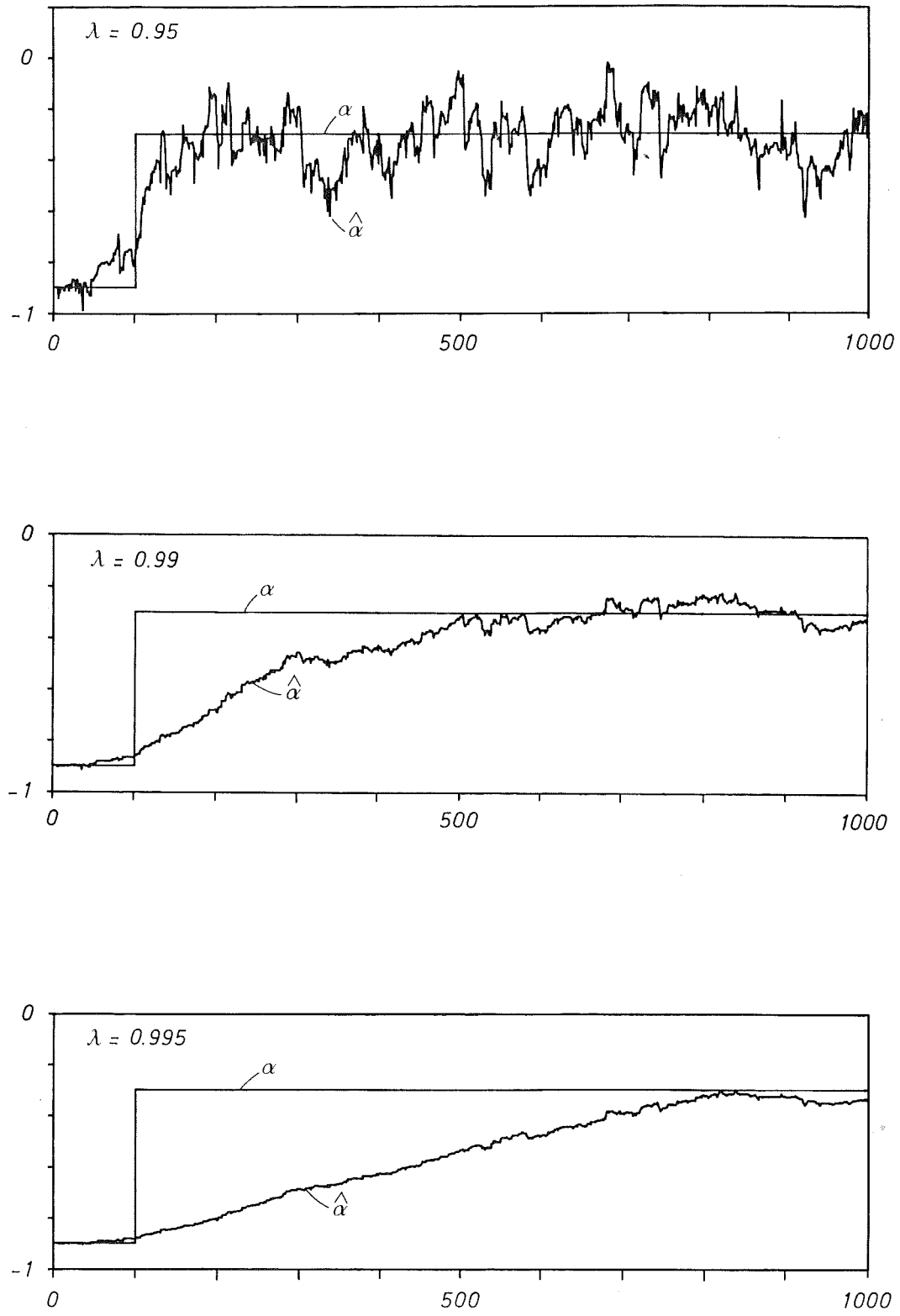


Figure 2.2 - Estimation of the parameter α(t) in Example 2.1. Different forgetting factors are used in the different simulations.

Estimator wind-up is another problem that may occur when a constant forgetting factor less than one is used. The method of exponential weighting of the incoming data may work well if the incoming information is uniformly distributed both in time and in space. Especially in the servo problem, when the major excitation comes from the variations of the command signal, this is seldom the case. (This may be one reason why most adaptive control applications are devoted to the regulator problem.) Discounting during times of poor excitation may then lead to uncertain estimates and numerical difficulties.

In order to avoid these problems, it has been attempted to use time-variable forgetting factors. Fortesque et al (1981) and Wellstead and Sanoff (1981) suggest the use of a forgetting factor which depends on the magnitude of the residuals $\varepsilon(t)$. When the magnitude of the residuals is large, the model is supposed to be changing. The forgetting factor is therefore decreased to discount old data more rapidly. The method is further discussed in the following chapters.

Irving (1979) proposed the use of a forgetting factor that keeps the trace of the P-matrix constant. The estimator wind-up problem is then eliminated, but the other problems caused by nonuniform excitation in the parameter space still remain unsolved.

The estimation can also be restarted repeatedly, instead of using a forgetting factor. This method is successfully practiced by Evans and Betz (1982), where the P-matrix is reset to a large matrix repeatedly.

In all the suggestions above, little is assumed about the nature of the parameter variations. When more a priori information is present, more sophisticated solutions are possible. If the parameters e.g. can be modeled by stochastic difference equations

$$\theta(t) = A\theta(t-1) + v(t) \qquad (2.8)$$

where A and the statistics of $v(t)$ are known, the Extended Kalman filter is suitable. In Åström (1980), the problem of estimating parameters which are a sum of an ARMA signal and a piece-wise deterministic signal is considered. Several papers have also been devoted to the problem when the parameters

switch between a limited number of sets. See e.g. Sorenson and Alspach (1971), Lo (1972), Wittenmark (1979) and Millnert (1982).

Obviously, the major problems originating from the time-variability of an adaptive control system arise in the parameter estimator, (E) in Figure 2.1. The least squares algorithm will be analysed in more detail in the next chapter. This analysis will serve as a basis for a discussion of modified algorithms.

3. ESTIMATION OF TIME-VARYING PARAMETERS

In this chapter, estimation of time-varying parameters is discussed. The recursive least squares method is examined, and earlier suggestions for modifications to treat time-varying parameters are investigated.

3.1 The recursive least squares algorithm

The recursive least squares (LS) algorithm will now be explored in more detail. As mentioned before, the LS algorithm is the most common way of estimating parameters from input and output data. For a thorough description of the LS method, see Kendall and Stuart (1961) and Ljung and Söderström (1983).

In the (weighted) LS estimation procedure, the vector $\hat{\theta}(t)$ which minimizes the loss function

$$J(\hat{\theta}(t)) = \sum_{i=1}^{t} \frac{1}{\omega(t,i)} \left[y(i) - \hat{\theta}(t)^T \varphi(i)\right]^2 \qquad (3.1)$$

is selected. A desirable choice of the weights $\omega(t,i)$ would be the variances of the corresponding measurements. Compare with the minimum variance estimator in case of known regression vectors $\{\varphi(t)\}$. A key problem in identification of time-varying systems is, however, the lack of knowledge about these variances.

At each time instant t, the parameters $\theta(t)$ are estimated based on measurements in the period $[0,t]$. From Equation (2.1), the relation between the measurement received at time i and the parameters $\theta(t)$ can be derived as

$$y(i) = \theta(i-1)^T \varphi(i) + e_n(i) =$$

$$= \theta(t-1)^T \varphi(i) + \left[\theta(i-1) - \theta(t-1)\right]^T \varphi(i) + e_n(i) \triangleq$$

$$\triangleq \theta(t-1)^T\varphi(i) + e_m(t,i) + e_n(i) \qquad (3.2)$$

A comparison between Equations (3.2) and (2.1) shows that the error can be interpreted as being composed of two terms, the measurement noise e_n and the model error e_m. The model error e_m is zero if the parameters are constant. It reflects the nature of the time-variations of the parameters. This model error may be caused by a change of the operating point in a nonlinear system, by changes in the process depending on temperature variations, wear or aging, by failing sensors or actuators etc.

In the LS method, each measurement is weighted depending on its uncertainty, see Equation (3.1). As seen above, this uncertainty can be interpreted as composed of two independent components, namely the measurement noise e_n and the model error e_m with the corresponding variances

$$\sigma(t,i)^2 = \sigma_m(t,i)^2 + \sigma_n(i)^2 \qquad (3.3)$$

Here $\sigma_m(t,i)^2$ is the model error variance, and $\sigma_n(i)^2$ is the noise variance at time i. As mentioned above, it would be desirable to choose the weights $\omega(t,i)$ in the loss function $J(\hat{\theta}(t))$ equal to $\sigma(t,i)^2$.

Some examples of models for the parameter variations will now be elaborated.

Example 3.1: Constant parameters and constant noise level.

When the parameters are constant, it follows from Equation (3.2) that

$$e_m(t,i) = \left[\theta(i-1) - \theta(t-1)\right]^T\varphi(i) = 0 \qquad (3.4)$$

Hence

$$\sigma_m(t,i) = 0 \qquad (3.5)$$

The total error variance therefore becomes equal to the noise variance, i.e.

$$\sigma(t,i)^2 = \sigma_n(i)^2 = \sigma^2 \qquad (3.6)$$

All the measurements then have the same uncertainty. This is the situation presumed in the original least squares procedure given by Equation (2.2).

□

Example 3.2: Exponentially increasing model error variance and constant noise variance.

If the error variances are given by

$$\sigma_m(t,i)^2 = \left[\left(\frac{1}{\lambda}\right)^{t-i} - 1\right]\sigma^2 \qquad \sigma_n(i)^2 = \sigma^2 \tag{3.7}$$

the total error variance becomes

$$\sigma(t,i)^2 = \left(\frac{1}{\lambda}\right)^{t-i}\sigma^2 \tag{3.8}$$

The uncertainty of the measurements thus increases exponentially with time. This case corresponds to discounting with a constant forgetting factor as was described in Chapter 2.

□

The weighted LS estimate of $\theta(t)$ is given by

$$\hat{\theta}(t) = \left[\Phi(t)^T V(t)^{-1}\Phi(t)\right]^{-1}\Phi(t)^T V(t)^{-1}Y(t) \tag{3.9}$$

where

$$\Phi(t) = \begin{bmatrix} \varphi(1)^T \\ \varphi(2)^T \\ \vdots \\ \varphi(t)^T \end{bmatrix}, \quad Y(t) = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(t) \end{bmatrix}, \quad V(t) = \mathrm{diag}(\omega(t,i))$$

See Kendall and Stuart (1961). Here, and throughout the thesis, it is assumed that the columns of $\Phi(t)$ are linearly independent, and that $V(t)$ is invertible. To simplify the writing, the notation

$$v(t) = \omega(t,t) \tag{3.10}$$

will be used. Since $\sigma_m(t,t) = 0$, see Equation (3.2), it is desirable to choose $v(t)$ equal to the noise variance $\sigma_n(t)^2$.

In case of constant parameters, i.e. $\sigma_m(t,i) = 0$ for all i, the weights $\omega(t,i) = v(i)$ are used, and the recursive version of the LS algorithm becomes

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{v(t)} P(t)\varphi(t)\varepsilon(t)$$

$$\varepsilon(t) = y(t) - \varphi(t)^T\hat{\theta}(t-1) \qquad (3.11)$$

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{v(t) + \varphi(t)^T P(t-1)\varphi(t)}$$

where

$$P(t) = \left[\Phi(t)^T V(t)^{-1}\Phi(t)\right]^{-1} \qquad (3.12)$$

$P(t)$ is the covariance matrix of the parameter estimates. In case of normal distribution of the data, $P(t)^{-1}$ is an estimate of Fisher's information matrix, see Goodwin and Payne (1977). Throughout the thesis, $P(t)^{-1}$ will be used as a measure of the information available at time t.

The recursive version of the LS algorithm in case of time-varying parameters is mostly not as simple as Equations (3.11). In some restrictive cases, it is however possible to get fairly compact expressions. A familiar example is given below.

Example 3.3: If $\sigma_m(t,i)^2 = ((1/\lambda)^{t-i}-1)\sigma^2$ and $\sigma_n(i)^2 = \sigma^2$, the recursive version of the LS algorithm becomes

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{\sigma^2} P(t)\varphi(t)\varepsilon(t)$$

$$\varepsilon(t) = y(t) - \varphi(t)^T\hat{\theta}(t-1) \qquad (3.13)$$

$$P(t) = \frac{1}{\lambda}\left[P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{\lambda\sigma^2 + \varphi(t)^T P(t-1)\varphi(t)} \right]$$

The P(t)-matrices are normally scaled with σ^2. It is then not necessary to know the value of σ^2. Compare with Equation (2.5).

□

3.2 The weighting problem

The purpose of LS estimation is to find the parameters $\theta(t)$ which minimize the loss function $J(\hat{\theta}(t))$, where the weighting coefficients are equal to the corresponding variances. In case of known variances σ^2, the solution is simply given by Equation (3.9). The problem is however, that σ is normally not known, neither is σ_m nor σ_n. Therefore, σ is estimated or hypothesized in some way.

First of all, some assumptions on σ_m and σ_n must be stated. The problem is meaningless if σ_m is varying as much as the states φ. The notation of variance also becomes meaningless if the variance is varying as much as the stochastic variable itself. The following assumption is therefore made.

Assumption 3.1: If the parameters to be estimated or the noise level vary, they vary slowly and/or seldom compared with the time constants of the system.

□

The assumption means e.g. that large step changes in the parameters may not occur frequently. The assumption should not be any limitation, but rather a check that the problem is well formulated. As was mentioned in Chapter 2, the adaptive control concept is based on the assumption that the above separation between the parameters and the states can be made. If this is not possible, the control problem has to be solved using other nonlinear control methods.

Some of the heuristic schemes for discounting old information will now be interpreted using the concept of model error variance. The proposed methods for choosing σ can be grouped according to the additional assumptions made on σ_m and σ_n apart from Assumption 3.1. Each method belongs to one of the four cases shown in Table 3.1.

	Constant parameters $\sigma_m(t,i) = 0$	Time-varying parameters
Constant noise level $\sigma_n(i) = \sigma$	1	2
Time-varying noise level	3	4

Table 3.1: Assumptions made in a time-variable adaptive system.

Case 1. This case corresponds to the parameters being unknown but constant. It will be referred to as the tuning case. Since the noise level also is assumed to be constant, the LS algorithm is independent of the noise level σ. The original LS method without discounting of past data, see Equation (2.2), or e.g. $\lambda = 1 - \exp(-t/T)$ to eliminate erroneous initial values, can be used.

□

Case 2. This is also a common assumption. It is often also assumed that the parameters change slowly, and a forgetting factor less than one is used, i.e. $\sigma_m(t,i)^2 = ((1/\lambda)^{t-i}-1)\sigma^2$. If λ is constant, it is at least implicitly also assumed that the parameters are changing all the time at a regular rate.

Instead of using a forgetting factor, a positive matrix R_1 is sometimes added to the right hand side of the updating equation of the P-matrix, Equation (2.2b). This corresponds to the parameters $\theta(t)$ being corrupted by independent disturbances with the covariance matrix R_1.

Experiments with different forgetting factors in different elements of the P-matrix have been made. The reason is then a priori knowledge about differences concerning rates of changes or amount of incoming information of the different parameters.

Sometimes, a time-varying forgetting factor is used, combined with some technique to estimate variations in σ_m. See e.g. Fortesque et al (1981). It is then no longer assumed that the parameters are changing slowly or at a regular rate.

□

Case 3. This problem is very seldom treated, if treated at all.

□

Case 4. This problem is also seldom treated. This thesis is devoted to case 4.

□

With the additional assumptions given above, the recursive LS algorithm is quite simple, see e.g. Equations (3.13). The corresponding adaptive controllers usually work well, if the assumptions are fulfilled. It is however also well-known that the algorithms can behave badly if the assumptions are violated. The consequences of bad modeling of σ_m and σ_n are exemplified below for the cases 1 and 2 in Table 3.1.

Case 1. Assume that both the parameters and the noise level are constant. If the parameters would vary, a very slow adaptation will result. The controller will behave almost like a constant regulator, since the adaptation ability is gradually turned off.

□

Case 2. If the parameters are varying at a slower rate than assumed, the uncertainty of the estimates will be unnecessarily large. A remarkable situation is the estimator wind-up when the P-matrix "explodes", though the parameters are constant. If on the other hand the parameters are varying faster than assumed, the convergence will be slow. These problems were illustrated in Figure 2.2.

The value of the forgetting factor is chosen as a trade-off between fast adaptation and high stationary accuracy of the estimates. An increasing noise level will mostly cause an increased uncertainty of the parameter estimates. The old value of the forgetting factor is then often a bad choice.

The use of variable forgetting factors is often even more dependent on the assumptions. An increase of the noise level σ_n will in most algorithms be interpreted as a variation of the parameters, i.e. an increase of σ_m. This is a serious mistake. It means that the algorithm "believes" that old measurements are more uncertain than the new ones, while the situation is the opposite. The result is, that old information is forgotten, when the algorithm instead should take extra care of those measurements, bearing in mind the poor information that will come in the future. The situation is exemplified in Chapter 6.

□

The remaining part of the thesis is concerned with the weighting problem discussed above, and new proposals will be given. The different types of time-variability mentioned in Assumption 3.1 will be treated separately. In the next chapter, large parameter changes, which occur infrequently, are discussed and handled in a special way. Slow parameter changes and changes of the excitation are then treated in Chapter 5.

4. LARGE PARAMETER CHANGES

This chapter is concerned with parameter estimation in the case of large parameter changes. The problem can be divided into two parts: Detection of parameter changes and modification of the estimation algorithm. The first part is related to fault detection. The chapter therefore begins with a short review of earlier fault detection methods, and a discussion of requirements on a fault detection procedure which is suitable for adaptive control. A new fault detection approach is then presented in Sections 4.3 and 4.4. How to modify the estimation procedure when a fault is detected is treated in Section 4.5. Finally, the new fault detection procedure is illustrated by a simulation example in Section 4.6 and an application to level estimation in tankers in Section 4.7.

4.1 Earlier work

It should first of all be mentioned that the notation "fault" in this thesis means a change in the process model, more precisely in the parameters $\theta(t)$, which does not necessarily originate from a physical fault in the process. It can just as well be a parameter change due to a shift of the operating point in a nonlinear system.

A great variety of methods for fault detection has appeared in recent years. Some of them are general, while others are devoted to special applications or concerned with voting between some known models. The problem described in the previous chapter requires a general method. The following discussion is therefore restricted to such approaches.

Research on fault detection has taken place in many different disciplines, e.g. automatic control, information theory, signal processing and statistics. The reason is probably the wide applicability of such methods. Successful use of fault detection has been reported in medicine (electrocardiograms), image processing (edge detection), geophysics and speech processing. Surveys of fault detection methods and references to applications are given in Basseville (1982) and in Willsky (1976).

A fault detection procedure consists in forming a test sequence which is sensitive to faults, i.e. which has significantly different properties before and after a fault. This sequence is then analysed and decision theory is applied to decide if and when a fault occurs.

The residual sequence $\{\varepsilon(t)\}$, i.e. the differences between the true output signals and the expected output signals of the system, is the predominantly used test sequence. The expected output signals are mostly derived from a Kalman filter or an estimation algorithm as in Equation (2.2). When the statistics of $\{\varepsilon(t)\}$ differs considerably from the measurement noise sequence $\{e_n(t)\}$, a fault is concluded.

There are two great disadvantages with such tests. First, the statistics of the noise sequence $\{e_n(t)\}$ must be known to enable any decision about faults. This is easily seen in Equation (2.2c), where a registered change of the statistical properties of $\{\varepsilon(t)\}$ obviously can originate from either a fault or a change in the noise sequence $\{e_n(t)\}$. The assumption of known disturbance statistics is further discussed in the next section.

The second disadvantage is that only faults that have a large influence on the output signal can be expected to be detected. In processes with reasonable noise levels, large faults may often occur without any immediate large effects on the output signals. An example is given in Section 4.6. It should be possible to detect such successive effects in the output signal by a suitable nonlinear dynamic manipulation of the measurement sequence.

There is a third drawback when using the magnitude of the residuals as a fault indication in least squares estimation, since this method is based on a minimization of a quadratic loss function. See Equation (3.1). The quadratic loss is motivated by the assumption that the noise acting on the system has a Gaussian distribution. The assumption of Gaussian distribution can to some extent be motivated by the central limit theorem. However, a deviation from the Gaussian distribution may result in drastically changed properties of the estimator. Thus such robustness considerations often suggest to pay less attention to large values of $|\varepsilon(t)|$ compared with the quadratic loss. See e.g. Huber (1964) and Poulsen and Holst (1982).

Here is a conflict. From robustness considerations, the gain in the estimator should be small for large values of $|\varepsilon(t)|$. The methods that use the magnitude of $\varepsilon(t)$ as a measure of parameter changes want to have a high gain for large values of $|\varepsilon(t)|$.

As mentioned above, more advanced fault detection methods can be derived from a more sophisticated filtering of the residual sequence. The least squares estimator is such a filter, and it produces estimates of the parameter vector $\theta(t)$. Since the problem of fault detection is concerned with changes in this vector, it seems natural to use the estimate sequence $\{\hat{\theta}(t)\}$ as a starting point for detection. This has also been suggested for some time in the literature. The second disadvantage of the preceding test sequence is then avoided, but these methods are still based on the assumption of known noise statistics.

In spite of the drawbacks of using the residuals $\varepsilon(t)$ as a test sequence, this use is seldom questioned in the literature. Far more interest is payed to the choice of decision method. All variants from the Sequential Probability Ratio Test, see Wald (1947), to simple cumulative sum tests have been suggested. It would lead too far from the theme of this thesis to discuss these methods in detail, but the reader is refered to the references, Basseville (1982) and Willsky (1976), which give extensive reviews with many references.

4.2 Requirements on the fault detection

Many techniques have been proposed for the detection of faults in dynamic systems. Some of them are general, while others use more a priori information. To facilitate the choice of method, some natural requirements for this special application will be stated here.

(R1) The times when the faults occur are not known.

(R2) The nature of the faults is not known.

Since the transformation between the physical parameters in the process and the parameters in the model (2.1) is usually quite involved, this is a natural requirement.

(R3) It must be possible to repeat the detection from the new modes of operation.

This means e.g. that there does not exist any "normal mode". As soon as a change in $\theta(t)$ is accepted, the old parameters are forgotten. This requirement is considered to give a general method. In some applications it can be relaxed.

(R4) A change in the noise level must not disturb the detection.

The only assumption made on the noise sequence $\{e(t)\}$ is that it consists of independent symmetrically distributed random variables. Therefore, a change in the noise level does not effect the parameters $\theta(t)$. This is an important requirement, since a change in the noise level is often much more likely than a change in the process parameters.

Requirement (R4) is important, not only for the reason given above. In real processes, disturbances are often entering at several points, and not only additively to the input or output signals. In the process model, the different disturbance sources are represented by one equivalent disturbance source entering at one point, see Aström (1970). The characteristics of these equivalent disturbances depend on the process parameters. This means that a change in the parameter vector usually also causes a change in the equivalent output noise level. Under these circumstances, it does not seem very realistic to detect faults under the assumption that the noise level in the output is constant.

According to the previous section, the requirement (R4) unfortunately rules out most of the existing fault detection procedures. A new fault detection procedure which satisfies the above requirements is presented in the next section. It was first described in Hägglund (1982).

4.3 A new fault detection method

A new fault detection method will now be discussed. The least squares parameter estimation method with constant forgetting factor will be used as a starting point. This is done in spite of the fact that the use of a forgetting factor is proposed to be replaced by another discounting principle in the next chapter. There are two reasons for this. First of all, the new fault detection method will not be restricted to any particular estimation scheme, so the conversion to the new discounting principle in Chapter 5 is trivial. Secondly, the least squares method with forgetting factor is still the most common estimation scheme in adaptive control.

The real problem is to detect changes in the parameter vector $\theta(t)$. The vector $\theta(t)$ is not known, and neither is $\tilde{\theta}(t)$. However, the difference between two successive estimation errors $\Delta\tilde{\theta}(t)$ is known for $\theta(t)$ constant, since

$$\Delta\tilde{\theta}(t) \triangleq \tilde{\theta}(t) - \tilde{\theta}(t-1) = \theta(t) - \hat{\theta}(t) - \theta(t-1) + \hat{\theta}(t-1) =$$

$$= - \hat{\theta}(t) + \hat{\theta}(t-1) \triangleq - \Delta\hat{\theta}(t) \qquad (4.1)$$

in this case. These differences will give the information needed for the fault detection. To be able to extract this information, the statistics of $\{\Delta\hat{\theta}(t)\}$ will first be investigated.

From Equation (2.2) the differences between two successive estimates are given by

$$\Delta\hat{\theta}(t) = P(t)\varphi(t)\left[\varphi(t)^T\tilde{\theta}(t-1) + e_n(t)\right] \qquad (4.2)$$

At time t, the estimates are thus updated in the direction of the vector $P(t)\varphi(t)$. The probabilities of positive and negative direction are almost the

same in normal operation when no fault has occurred, i.e. when the estimated parameters are close to the true ones. This is intuitively seen from the following arguments.

When $\lambda = 1$, the estimation procedure is the ordinary recursive least squares algorithm without any discounting of past data. It is known to be the best linear unbiased estimator, see e.g. Goodwin and Payne (1977). This implies that there is no correlation between the increments of the parameter estimates in normal operation. If there were a correlation, it would be possible to modify the algorithm so that a smaller variance of the estimates were obtained. (If there is any information at time t about how the estimates will be changed at time t+1, all information given at time t is not used, and it is possible to derive a better estimate). This contradicts the fact that the least squares algorithm is the best linear unbiased estimator. Hence, when $\lambda = 1$ the probabilities for the estimate increments to have positive and negative $P(t)\varphi(t)$ direction are the same, 0.5.

When $\lambda < 1$, a negative correlation between two successive estimate increments is expected. If a forgetting factor less than one is used, the gain in the parameter estimator is greater than it should be for $\lambda = 1$. Intuitively this means that the algorithm in each updating of the estimates has to compensate for the large step taken previously. Hence the expected correlation is negative. However, from continuity arguments this correlation is small when λ is close to one, and the probabilities of positive and negative $P(t)\varphi(t)$ direction of the estimate increments are approximately the same. This is illustrated in Example 4.1 below.

The arguments above imply that under normal operation

$$P\left[\Delta\hat{\theta}(t)^T\Delta\hat{\theta}(t-1) > 0\right] \approx P\left[\Delta\hat{\theta}(t)^T\Delta\hat{\theta}(t-1) < 0\right] \qquad (4.3)$$

where P denotes the probability measure.

When $\tilde{\theta}(t)$ is not close to its true value, i.e. when a fault has occurred, the approximations used in the heuristic arguments above are no longer valid. Since the estimated parameters then will be driven towards the new values, the following inequality holds

$$P\left[\Delta\hat{\theta}(t)^T\Delta\hat{\theta}(t-1) > 0\right] > P\left[\Delta\hat{\theta}(t)^T\Delta\hat{\theta}(t-1) < 0\right] \qquad (4.4)$$

The intuitive arguments above are illustrated in Figure 4.1, where the trajectories of the estimated parameters of a two parameter model are shown in the parameter plane, both in case of stationarity and when a fault has occurred. Figure 4.1 shows the first 200 points from the simulation given in Figure 6.6a. The difference described by equations (4.3) and (4.4) will be used in the sequel to derive the fault detection method.

The intuitive way of arguing that the correlation between successive estimate increments is small in case of constant parameters may be unappealing to readers familiar with more strict mathematical derivations. As mentioned before, the general adaptive system is so complicated, that an analysis unfortunately mostly is limited to very restrictive cases. In the following example, Equation (4.3) is verified under fairly hard restrictions.

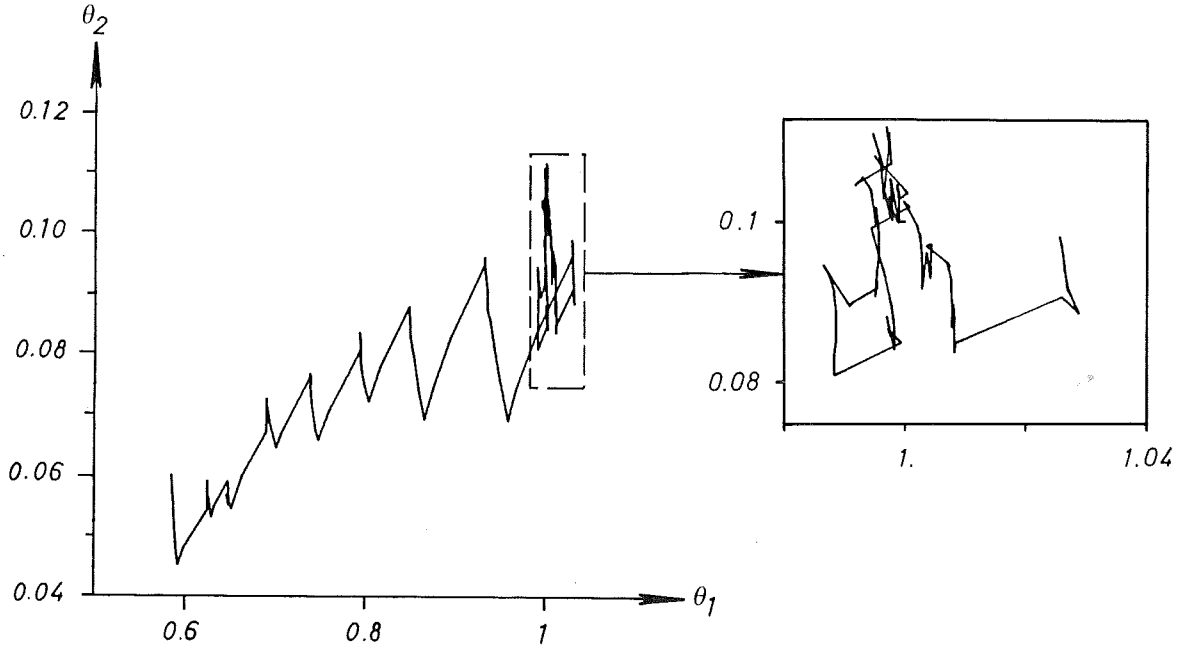


Figure 4.1 - Trajectories of the estimates in Figure 6.6a. The magnified part to the right shows the first 100 points, when the fault has not occured.

Example 4.1. Consider the process model

$$y(t) = \theta \cdot u + e(t) \tag{4.5}$$

where the input u is constant and $\{e(t)\}$ is a sequence of independent Gaussian random variables. If the estimator defined by Equations (2.2a), (2.2c) and (2.5) is applied, the P-matrix converges to the constant scalar

$$P = \frac{1 - \lambda}{u^2} \tag{4.6}$$

The updating formula of the estimate of θ then becomes

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1 - \lambda}{u^2} \left[\tilde{\theta}(t-1) \cdot u + e(t)\right] \tag{4.7}$$

Without loss of generality, the constant input signal u is normalized to 1 in the sequel. The estimate increments then becomes

$$\Delta\hat{\theta}(t) = (1-\lambda)\left[\tilde{\theta}(t-1) + e(t)\right] \tag{4.8}$$

By using Equation (4.7) and the fact that θ is constant, Equation (4.8) can also be written as

$$\Delta\hat{\theta}(t) = -(1-\lambda)^2 \sum_{i=0}^{t-1} \lambda^{t-1-i} e(i) + (1-\lambda)e(t) \tag{4.9}$$

Since $\Delta\hat{\theta}(t)$ is a sum of Gaussian random variables, it is Gaussian itself. The aim is now to compute the following probability

$$P\left[\Delta\hat{\theta}(t) \cdot \Delta\hat{\theta}(t-1) < 0\right] = P\left[\Delta\hat{\theta}(t) > 0,\ \Delta\hat{\theta}(t-1) < 0\right] +$$

$$+ P\left[\Delta\hat{\theta}(t) < 0,\ \Delta\hat{\theta}(t-1) > 0\right] = \int_{\Omega}\int f(x,y)dxdy \tag{4.10}$$

where f(x,y) is the simultaneous density function of $\Delta\hat{\theta}(t)$ and $\Delta\hat{\theta}(t-1)$. The integration area Ω is the second and fourth quadrants. Since both $\Delta\hat{\theta}(t)$ and $\Delta\hat{\theta}(t-1)$ are Gaussian, the joint distribution is also Gaussian, N(0,R), with the covariance matrix

$$R \triangleq \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix} = \begin{bmatrix} E\{\Delta\hat{\theta}(t)^2\} & E\{\Delta\hat{\theta}(t)\Delta\hat{\theta}(t-1)\} \\ E\{\Delta\hat{\theta}(t)\Delta\hat{\theta}(t-1)\} & E\{\Delta\hat{\theta}(t-1)^2\} \end{bmatrix} \tag{4.11}$$

From Cramér (1945), the following expression for the integral in Equation (4.10) is obtained

$$P\left[\ \Delta\hat{\theta}(t)\cdot\Delta\hat{\theta}(t-1) < 0\ \right] = \frac{1}{2} - \frac{1}{\pi}\arcsin(\varrho) \tag{4.12}$$

where ϱ is the correlation coefficient. The correlation coefficient is easily calculated from Equations (4.11) and (4.9). Hence

$$\varrho \triangleq \frac{r_{12}}{\sqrt{r_{11}\cdot r_{22}}} = -\frac{1-\lambda}{2} \tag{4.13}$$

Table 4.1 gives some numerical values of the probability of getting a negative scalar product between two successive estimate increments. For reasonable values of λ, Equation (4.3) is obviously true in this simple example.

□

λ	0.9	0.95	0.98	0.99	0.999
Probability	0.5159	0.5080	0.5032	0.5016	0.5002

Table 4.1 - The probability of getting a negative scalar product between successive estimate increments in Example 4.1.

Implementation

Instead of observing the scalar product

$$\Delta\hat{\theta}(t)^T\Delta\hat{\theta}(t-1)$$

it is often more efficient to study the scalar product between $\Delta\hat{\theta}(t)$ and a sum of the latest estimate increments. To simplify the algorithm, an exponential filtering of the increments of the estimates will be used instead of an ordinary sum. For this purpose, introduce w(t) as

$$w(t) = \gamma_1 w(t-1) + \Delta\hat{\theta}(t) \qquad 0 \leq \gamma_1 < 1 \tag{4.14}$$

In the case when a fault has occurred, w(t) can be viewed as an estimate of the direction of the parameter change. The motivations for the Equations (4.3) and (4.4) are valid even when w(t-1) is substituted for $\Delta\hat{\theta}(t-1)$. The test sequence that will be studied is {s(t)}, where s(t) is defined as

$$s(t) \triangleq \text{sign}\left[\Delta\hat{\theta}(t)^T w(t-1)\right] \tag{4.15}$$

The sign function makes the test sequence insensitive to the noise variance. It is now clear in principle how to carry out the fault detection:

> "Inspect the latest values of s(t). If s(t) is +1 unlikely many times, conclude that a fault has occurred."

The idea to use the signs of the differences between successive estimates to decide whether the estimates has converged or not has been proposed before. Kesten (1958) proposed a method to accelerate a stochastic approximation method by letting the gain of the estimator depend on the frequency of the changes of these signs.

Testing method

As mentioned earlier, most fault detection methods end up with a statistical test of a time sequence, e.g. $\{\varepsilon(t)\}$, $\{\varepsilon(t)^2\}$ or as in this case $\{s(t)\}$. A common way of doing this is to use the Sequential Probability Ratio Test (SPRT), or modifications of this test. See Wald (1947) for an exposé of these ideas. The SPRT is designed to decide between two hypotheses, which in this case means between two different parameter vectors. The SPRT is usually efficient in terms of short times to detect the faults, if the values of the true parameters occurring in practice are close to the hypothesized ones. However, if the hypothesized values are taken merely to obtain a SPRT, and do not represent the most frequently occurring values, the SPRT may not lead to any time saving compared with other methods, see Wetherill (1966). Requirement (R2) implies that no a priori information about the parameter changes is available. Since requirement (R1) furthermore implies that a new sequence to be tested must be introduced every sample instant for the SPRT to reach the expected efficiency, a traditional Bayesian approach will be used here instead.

Under normal operation, i.e. when the parameter estimates are close to their true values, s(t) has approximately a symmetric two point distribution with mass 0.5 each at +1 and -1. When a fault has occurred, the distribution is no longer symmetric, but the mass at +1 is larger than the mass at -1. To add the most recent values of s(t), the stochastic variable r(t) defined as

$$r(t) = \gamma_2 r(t-1) + (1-\gamma_2)s(t) \qquad 0 \leq \gamma_2 < 1 \qquad (4.16)$$

is introduced. The sum of the most recent values of s(t) is replaced by an exponential smoothing in order to obtain a simple algorithm. When the parameter estimates are close to the true ones, r(t) has a mean value close to zero. When a fault has occurred, a positive mean is expected.

The parameter γ_2 determines, roughly speaking, how many s(t) values that should be included. E.g. $\gamma_2 = 0.95$ corresponds to about 20 values, which is a reasonable choice in many applications. A small γ_2 allows a fast fault detection, although at the price of less security against false alarms. This trade-off is typical for all fault detection methods. When the signal to noise

ratio is small, it is not possible to detect the faults as fast as otherwise. It is then necessary to have more information available to decide whether a fault is present. This can be achieved by increasing γ_2.

The stochastic properties of r(t) are investigated in the next section. For values of γ_2 close to one, r(t) will have an approximately Gaussian distribution with variance

$$\sigma^2 = \frac{1 - \gamma_2}{1 + \gamma_2} . \qquad (4.17)$$

Since γ_2 is generally chosen in this region, it will in the sequel be assumed that r(t) has a Gaussian distribution.

If r(t) exceeds a certain threshold r_0, a fault may be concluded with a confidence determined from the value of the threshold. In the present algorithm, the threshold can be computed directly as a function of the rate of false alarms f_f. If a false alarm frequency equal to f_f is acceptable, a fault detection should be given every time r(t) is greater than the threshold r_0, defined by

$$P\left[r(t) \geq r_0\right] = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{r_0}^{\infty} \exp\left[- \frac{x^2}{2\sigma^2} \right] dx = f_f \qquad (4.18)$$

If a small value of the threshold is chosen to make it possible to detect faults quickly, the false detection rate will be high. This is seen in Equation (4.18), where there is an inverse relation between r_0 and f_f. As was said before, this compromise between fast detection and security against false alarms must be made in all fault detection methods. The determination of r_0 in this method has the advantage that it is formulated in terms of the expected frequency of false detections, which may be chosen to suit any particular application. In Figure 4.2, the error frequency f_f versus the threshold r_0 is presented for some different values of γ_2.

The fault detection method described above fulfils the requirements stated in Section 2.

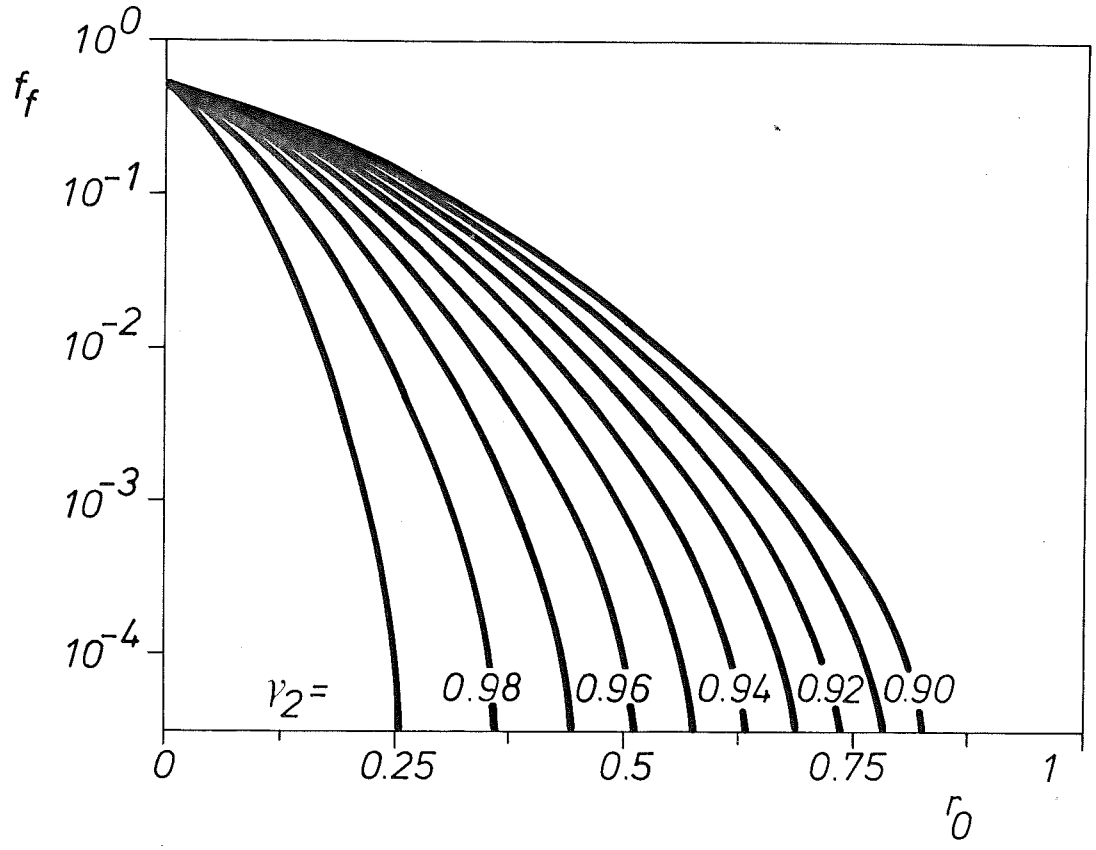


Figure 4.2 - The error frequency f_f versus the threshold r_0.

Finally, it should be remarked that if the requirement (R4) is eliminated, a more powerful fault detection method can probably be derived by studying the quantity $w(t)^T w(t)$. This variable is usually much more sensitive to parameter changes than the generally studied $\varepsilon(t)^2$, as illustrated in Section 4.6.

4.4 A stochastic difference equation

In the previous section, the derivation of the new fault detection method resulted in a proposed test of a sequence, $\{r(t)\}$. To be able to perform this test, the statistical properties of $r(t)$ must first be explored. This is done in the present section.

According to equation (4.16), r(t) is generated by the stochastic difference equation

$$r(t) = \gamma r(t-1) + (1-\gamma)s(t) \qquad 0 \leq \gamma < 1 \qquad (4.19)$$

where the subscript "2" of γ is dropped for the sake of convenience. The sequence {s(t)} consists of independent random variables with the symmetric two point distribution

$$s(t) = \begin{cases} -1 & \text{with probability } 0.5 \\ 1 & \text{with probability } 0.5 \end{cases} \qquad (4.20)$$

The distribution of r(t) is highly dependent on the value of γ. Even if only values of γ close to one are considered in this special application, it is interesting to investigate the behaviour of r(t) in the whole interval $0 \leq \gamma < 1$. This will be done below, starting with $\gamma = 0$.

$\underline{\gamma = 0}$

When $\gamma = 0$, r(t) is equal to s(t) and has consequently a symmetric two point distribution.

$\underline{0 < \gamma < 0.5}$

For values of γ in this interval, a distribution of Cantor-type occurs. See e.g. Chung (1968). To see this, it is first noted that asymptotically r(t) can take any of the following values

$$r(t) \in \{ \pm(1-\gamma)(1 \pm \gamma \pm \gamma^2 \pm \ldots) \} \qquad (4.21)$$

Arrange the asymptotic values of r(t) in groups according to

1. $r(t) = \pm(1-\gamma)(1+(\gamma+\gamma^2+\gamma^3+\ldots)) = \pm 1$

   $r(t) = \pm(1-\gamma)(1-(\gamma+\gamma^2+\gamma^3+\ldots)) = \pm(1-2\gamma)$

2. $r(t) = \pm(1-\gamma)(1+\gamma-(\gamma^2+\gamma^3+\gamma^4+ \ldots)) = \pm(1-2\gamma^2)$

   $r(t) = \pm(1-\gamma)(1-\gamma+(\gamma^2+\gamma^3+\gamma^4+ \ldots)) = \pm(1-2\gamma+2\gamma^2)$

3. $r(t) = \pm(1-\gamma)(1+\gamma+\gamma^2-(\gamma^3+\gamma^4+\gamma^5+ \ldots)) = \pm(1-2\gamma^3)$

   $r(t) = \pm(1-\gamma)(1-\gamma-\gamma^2+(\gamma^3+\gamma^4+\gamma^5+ \ldots)) = \pm(1-2\gamma+2\gamma^3)$

   $r(t) = \pm(1-\gamma)(1+\gamma-\gamma^2+(\gamma^3+\gamma^4+\gamma^5+ \ldots)) = \pm(1-2\gamma^2+2\gamma^3)$

   $r(t) = \pm(1-\gamma)(1-\gamma+\gamma^2-(\gamma^3+\gamma^4+\gamma^5+ \ldots)) = \pm(1-2\gamma+2\gamma^2-2\gamma^3)$

and so on.

The groups are formed in the following way. In group number i, the expressions in all the previous groups are rewritten, but with changed sign in front of all terms of order greater or equal to i in the right parenthesis. These values, defining the possible values of r(t), can also be derived in the following way:

1. From the closed interval $[-1,1]$, the open interval $(1-2\gamma)$ times the interval length is removed in the middle. This results in two disjoint closed intervals with endpoints equal to the values in group 1.

2. From the two disjoint intervals obtained under 1, remove $(1-2\gamma)$ times the interval length in each middle. This operation results in four closed disjoint intervals, with end points equal to the values in group 1 and 2 above.

and so on.

Dividing the closed intervals like this into smaller and smaller intervals, a Cantor set is obtained. For values of γ in the interval $0 < \gamma < 0.5$, the distribution is consequently singular. (It can be shown that the distribution function in spite of this is continuous, see Chung (1968)).

$\gamma = 0.5$

When $\gamma = 0.5$, r(t) will asymptotically have a uniform distribution in the interval [-1,1]. This can be shown by investigating the characteristic function. The stochastic variable r(t) can be decomposed into a sum of stochastic variables

$$r(t) = \sum_{n=1}^{t} X_n \; ; \qquad X_n = \begin{cases} 0.5^n & \text{with probability } 0.5 \\ -0.5^n & \text{with probability } 0.5 \end{cases} \tag{4.22}$$

The characteristic function of X_n is

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx}\, 0.5\left[\delta(-0.5^n) + \delta(0.5^n)\right]dx =$$

$$= \cos(0.5^n t) \tag{4.23}$$

The asymptotic characteristic function of r(t) is now given as the product

$$\varphi(t) = \prod_{n=1}^{\infty} \varphi_n(t) = \prod_{n=1}^{\infty} \cos(0.5^n t) = \frac{\sin(t)}{t} \tag{4.24}$$

The last equality can be obtained from ordinary mathematical tables, e.g. Gradshteyn and Ryzhik (1965). The characteristic function $\varphi(t)$ is the characteristic function of a uniformly distributed stochastic variable in the interval [-1,1]. It is therefore proved that r(t) is uniformly distributed when $\gamma = 0.5$.

$0.5 < \gamma < 1$

The distribution of r(t) for values of γ in this interval is not easy to determine. The asymptotic density function f of r(t) can however be determined according to the following theorem.

Theorem 4.1: The asymptotic density function f of r(t) satisfies the following functional equation for $0 < \gamma < 1$.

$$\begin{cases} f(r) = \frac{1}{2\gamma} f\left(\frac{r - 1 + \gamma}{\gamma} \right) + \frac{1}{2\gamma} f\left(\frac{r + 1 - \gamma}{\gamma} \right) \\ f(r) = 0 \qquad |r| \geq 1 \end{cases} \tag{4.25}$$

Proof: Denote the distribution function of r(t) by $F_t(r)$. From Equation (4.19) the distribution function is given by the equations

$$\begin{cases} F_t(r) = \frac{1}{2} F_{t-1}\left(\frac{r - 1 + \gamma}{\gamma} \right) + \frac{1}{2} F_{t-1}\left(\frac{r + 1 - \gamma}{\gamma} \right) \\ F_t(r) = 0 \qquad r \leq -1 \\ F_t(r) = 1 \qquad r \geq 1 \end{cases} \tag{4.26}$$

Furthermore, denote the asymptotic distribution function corresponding to the density function in Equation (4.25) by F(r). The theorem is proved if it is shown that the difference

$$\Delta F_t(r) = F_t(r) - F(r) \tag{4.27}$$

converges to zero. $\Delta F_t(r)$ fulfils the following inequalities

$$\begin{aligned} \left|\Delta F_t(r)\right| &\leq \frac{1}{2} \left|\Delta F_{t-1}\left(\frac{r-1+\gamma}{\gamma} \right)\right| + \frac{1}{2} \left|\Delta F_{t-1}\left(\frac{r+1-\gamma}{\gamma} \right)\right| \triangleq \\ &\triangleq \frac{1}{2} \left|\Delta F_{t-1}(r_{11})\right| + \frac{1}{2} \left|\Delta F_{t-1}(r_{12})\right| \leq \\ &\leq \frac{1}{4} \left|\Delta F_{t-2}\left(\frac{r_{11}-1+\gamma}{\gamma} \right)\right| + \frac{1}{4} \left|\Delta F_{t-2}\left(\frac{r_{11}+1-\gamma}{\gamma} \right)\right| + \\ &+ \frac{1}{4} \left|\Delta F_{t-2}\left(\frac{r_{12}-1+\gamma}{\gamma} \right)\right| + \frac{1}{4} \left|\Delta F_{t-2}\left(\frac{r_{12}+1-\gamma}{\gamma} \right)\right| \leq \\ &\leq \frac{1}{8} \left|\Delta F_{t-2}\left(\frac{r_{21}-1+\gamma}{\gamma} \right)\right| + \ldots \end{aligned} \tag{4.28}$$

Proceeding until at least one argument gets outside the interval [-1,1] yields the inequality

$|\Delta F_t(r)| <$ The mean value of $|\Delta F_{t-k}(r_i)|$ in a couple of points r_i at time t-k.

Hence the difference between the functions $F_t(r)$ and $F(r)$ converges to zero.

□

Equation (4.25) has been solved numerically for some values of γ, see Figure 4.3. For values of γ close to one, r(t) will have an approximately Gaussian distribution with variance

$$\sigma^2 = \frac{1 - \gamma}{1 + \gamma} \qquad (4.29)$$

since r(t) in this case is a sum of almost equally distributed random variables.

It has been shown above that the distribution of r(t) varies considerably with γ. To illustrate this, the density functions for some values of γ are shown in Figure 4.3. For values of γ less than 0.5, the density function is singular. Therefore the peaks in the corresponding diagrams represent Dirac-impulses with appropriate areas.

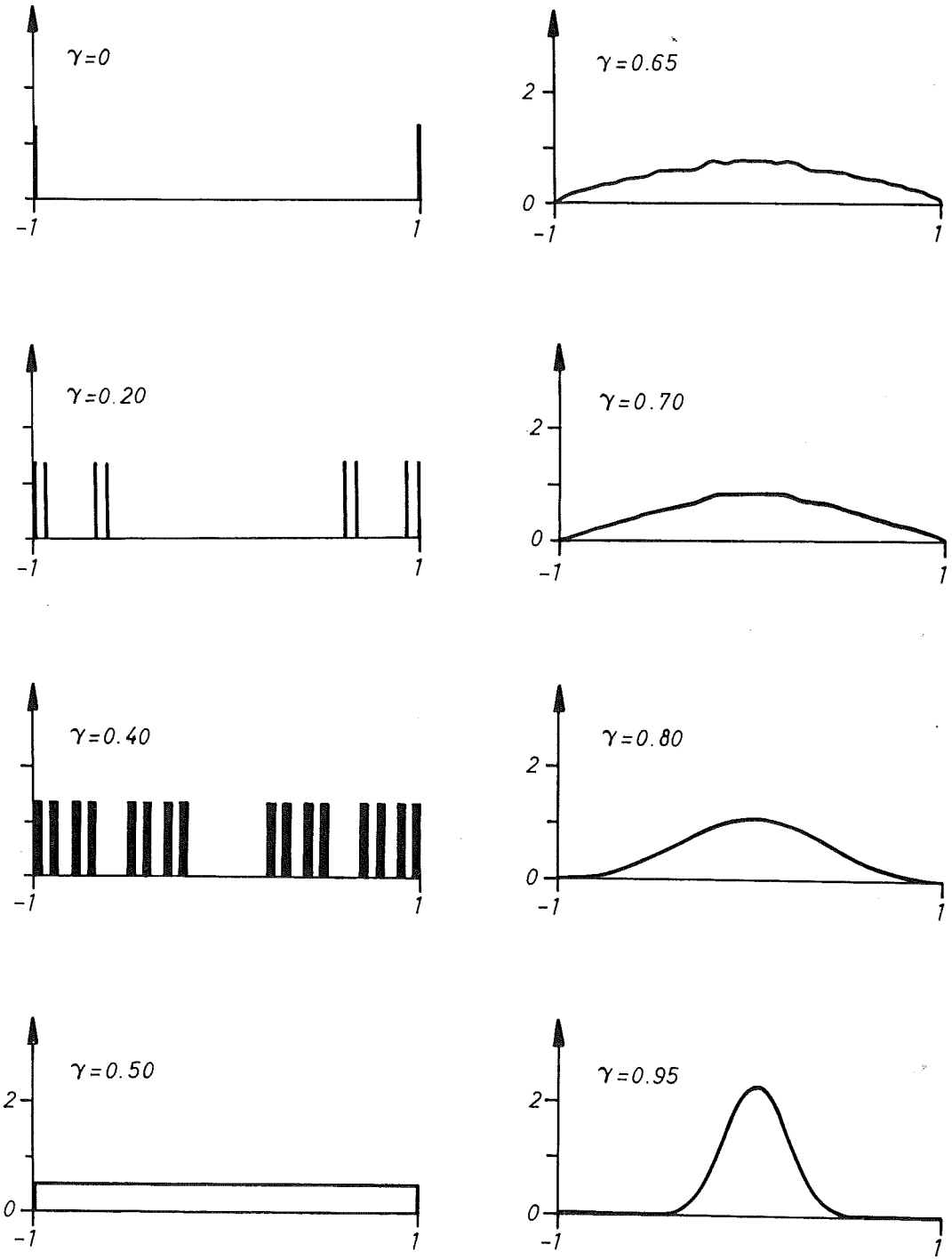


Figure 4.3 - Graphical illustration of the asymptotic density function of r(t) for different values of the parameter γ. Notice that the densities are distributions for $\gamma < 0.50$.

4.5 Modification of the estimation algorithm

The first part of a method to handle large parameter changes was given by the fault detection procedure derived in the previous sections. To complete the method, a procedure to increase the gain in the estimator, i.e. the P-matrix in Equation (2.5), must also be established. From the information handling point of view, given in Chapter 3, the increase of the gain in the estimator can also be seen as a reduction of the information content in the estimator. The inverse P-matrix denotes the information content. When a fault has occured, P^{-1} indicates a too large information content. By decreasing P^{-1} when a fault is detected, the performance of the estimator can be improved considerably.

To motivate the modification of the estimation algorithm, this section begins with an investigation of the effects of an updating of the estimates.

According to Chapter 3, the least squares algorithm gives, at each sample instant, a solution to the minimization problem

$$\min_{\hat{\theta}} J(\hat{\theta}(t)) \triangleq \min_{\hat{\theta}} \sum_{i=0}^{t} \lambda^{t-i} \varepsilon(i)^2 \qquad (4.30)$$

Figure 4.4 shows an example of contours of constant values of $J(\hat{\theta}(t))$ in the parameter space, in the noise-free case. At the point $\hat{\theta}(t-1)$, $\varphi(t)$ is orthogonal to the contour. In the least squares method, the estimate updating is not done in the $\varphi(t)$ direction, but in the $P(t)\varphi(t)$ direction. Near the correct parameter values, $P(t)$ is mostly approximately proportional to the inverse of the Hessian

$$H(t) = \nabla^2 J(\hat{\theta}(t))\Big|_{\hat{\theta}(t)=\theta} \qquad (4.31)$$

By updating in the $H(t)^{-1}\varphi(t)$ direction, the order of convergence is two instead of one, which would be the case if the updating were made along the $\varphi(t)$ vector, see Luenberger (1973). If the updating is done in the $P(t)\varphi(t)$ direction, the order of convergence is supposed to be somewhere between one and two, and close to two near the correct solution.

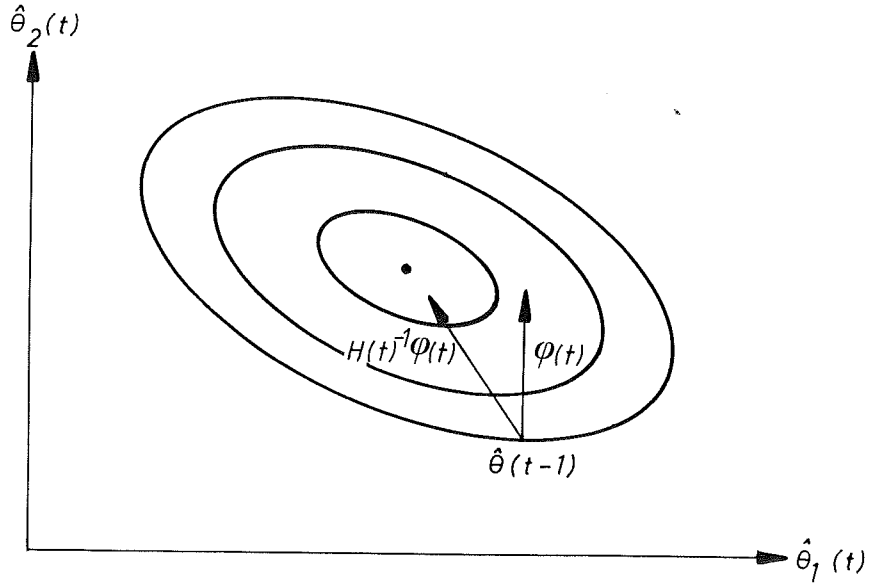


Figure 4.4 - Contours of constant values of $J(\hat{\theta}(t))$ in the parameter space.

When a fault is detected, the gain in the estimation algorithm should be increased. It means that P(t) should be increased. This can be achieved in many ways, but there are mainly two methods that have been used previously. The first one is to decrease the forgetting factor λ. The growth of P(t) is then nearly exponential. The second method is to add a constant times the unity matrix to the P(t)-matrix, in which case P(t) is increased instantaneously.

When a fault has occurred, it is likely that the P(t)-matrix is no longer a good approximation of the inverse Hessian. If the Hessian is unknown, the most reasonable direction of the parameter updating is along the $\varphi(t)$ vector, both from stability and rate of convergence point of view. The gain in the estimation algorithm will therefore be increased according to the second method, and Equation (2.5) will be substituted by

$$P(t) = \frac{1}{\lambda}\left[P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{\lambda + \varphi(t)^T P(t-1)\varphi(t)} \right] + \beta(t)\cdot I \quad (4.32)$$

where $\beta(t)$ is a nonnegative scalar and I is the unity matrix. The variable $\beta(t)$ is zero except when a fault is detected. When a fault is detected, a positive

$\beta(t)$ has the effect that the P(t)-matrix increases and that the parameter updating is made in a direction closer to $\varphi(t)$.

The final problem is to choose a suitable $\beta(t)$. When no fault is detected, $\beta(t)$ is zero. When a fault is detected, it is reasonable to let $\beta(t)$ depend on the actual value of P(t) and on how significant the alarm is, i.e. on the value of r(t). This may of course be done in many ways, and the following proposal is just one possibility.

In the noise-free case, the progress of the estimation error, when $\theta(t)$ is constant, is given by

$$\tilde{\theta}(t) = \tilde{\theta}(t-1) - P(t)\varphi(t)\varepsilon(t) =$$

$$= \left[I - P(t)\varphi(t)\varphi(t)^T\right]\tilde{\theta}(t-1) \triangleq U(t)\tilde{\theta}(t-1) \qquad (4.33)$$

All eigenvalues of U(t) are one, except the one corresponding to the eigenvector $P(t)\varphi(t)$. This eigenvalue determines the step length in the algorithm. A small eigenvalue causes large steps, while an eigenvalue close to one means that the step length in the algorithm is small. Using Equation (4.32), the eigenvalue can be written as

$$1 - \varphi(t)^T P(t)\varphi(t) = 1 - \frac{\varphi(t)^T P(t-1)\varphi(t)}{\lambda + \varphi(t)^T P(t-1)\varphi(t)} -$$

$$- \beta(t)\varphi(t)^T\varphi(t) = \frac{\lambda}{\lambda + \varphi(t)^T P(t-1)\varphi(t)} - \beta(t)\varphi(t)^T\varphi(t)$$

(4.34)

When $\beta(t) = 0$, the eigenvalue is thus

$$\nu_0(t) = \frac{\lambda}{\lambda + \varphi(t)^T P(t-1)\varphi(t)} \qquad (4.35)$$

The eigenvalue is obviously between zero and one as long as $P > 0$. Suppose now, that an eigenvalue equal to $\nu(t)$ is desired when a fault is detected. Then $\beta(t)$ has to be chosen as

$$\beta(t) = \frac{1}{\varphi(t)^T\varphi(t)} \left[\nu_0(t) - \nu(t)\right] \qquad (4.36)$$

The eigenvalue $\nu(t)$ should lie in the interval

$$0 < \nu(t) \leq \nu_0(t) \qquad (4.37)$$

in order to keep the P(t)-matrix positive definite. In practice, this choice of $\beta(t)$ must also be combined with a test for nonsingularity of $\varphi(t)^T\varphi(t)$.

It remains to determine a suitable $\nu(t)$. This can be done in many ways. In the example presented in the next section, $\nu(t)$ is a piecewise linear function of the significance of the fault alarm, see Figure 4.5.

Combining the fault detection procedure in Section 4.3 with the modification of the estimation algorithm proposed in this section, a method to increase the gain in the estimation algorithm in case of large parameter changes is derived. The method is summarized in a block diagram in Figure 4.6. The input to the algorithm is the sequence of parameter estimates $\hat{\theta}(t)$. When a fault is detected, the P-matrix is increased according to Equation (4.32). This corresponds to a decrease of the weights on old data, see Equation (3.12), or in other words, a decrease of the information content in the estimator.

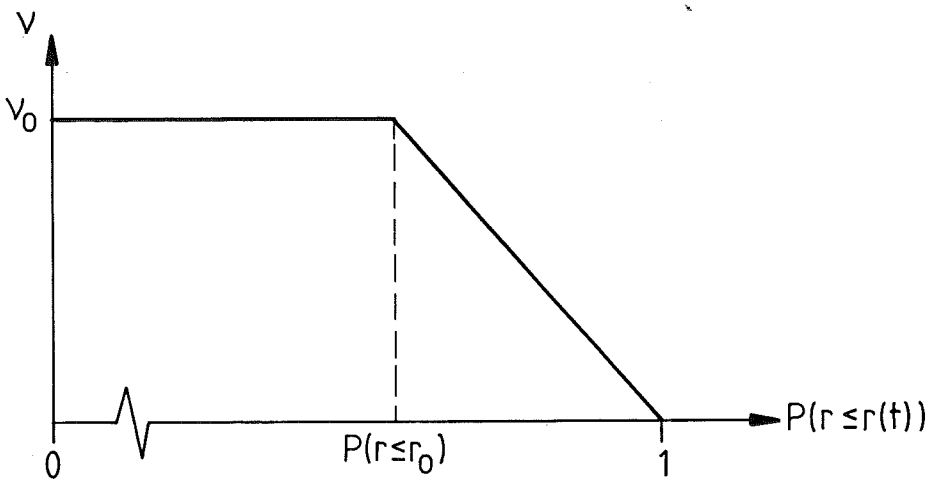


Figure 4.5 - An example of a choice of $\nu(t)$. r_0 is defined in Section 4.3.

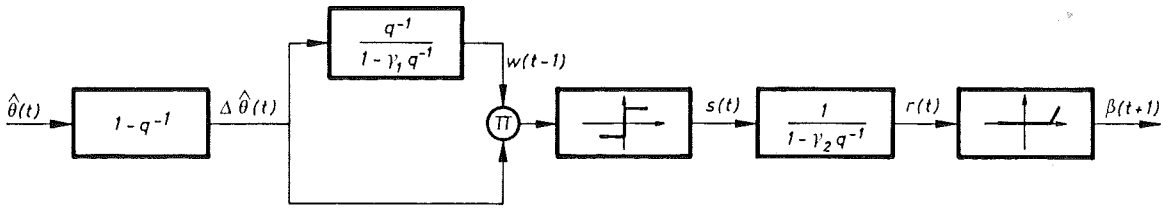


Figure 4.6 - A block diagram describing the fault detection method.

4.6 A simulation example

To illustrate the new fault detection method and the modified estimation algorithm, a simulation study is presented in this section. The simulations are performed using the simulation package SIMNON, see Elmqvist (1975).

The system considered is shown in Figure 4.7. The purpose of control is to keep the level in the tank constant. This is done by measuring the tank level and controlling the inlet valve. The dynamics of the tank is described by the equations

$$\frac{dh(t)}{dt} = \frac{1}{10}\left[q_{in}(t) - q_{out}(t)\right] + 0.005\, e(t) \qquad (4.38a)$$

$$q_{out}(t) = a_{out}\sqrt{2gh(t)} \qquad (4.38b)$$

where $\{e(t)\}$ is a disturbance sequence and a_{out} is the outlet area. The sequence $\{e(t)\}$ is generated as discrete Gaussian N(0,1) random variables

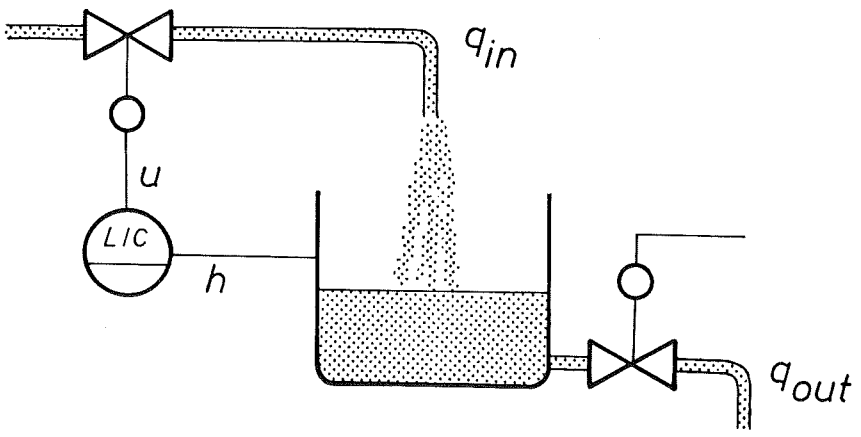


Figure 4.7 - The tank system.

with a sampling period equal to 1/10:th of the controller sampling period. The stochastic part of the equations can be viewed as originating from irregularities in the flow.

The model of the tank used in the estimation algorithm is

$$h(t+1) = a(t)\cdot h(t) + u(t) + \xi(t) \qquad (4.39)$$

where u(t) is the control signal and $\{\xi(t)\}$ is a sequence of independent random variables. The parameter a(t) is estimated by the recursive least squares method according to equations (2.2a), (2.2c) and (4.32). The equations become

$$\begin{aligned} \hat{a}(t) &= \hat{a}(t-1) + P(t)h(t-1)\varepsilon(t) \\ \varepsilon(t) &= h(t) - \hat{a}(t-1)\cdot h(t-1) - u(t-1) \\ P(t) &= \frac{P(t-1)}{\lambda + P(t-1)h(t-1)^2} + \beta(t) \end{aligned} \qquad (4.40)$$

The forgetting factor λ is chosen to 0.995. The equations of the fault detection procedure become

$$\begin{aligned} w(t) &= \gamma_1 w(t-1) + \left[\hat{a}(t) - \hat{a}(t-1)\right] \\ s(t) &= \text{sign}\left(\left[\hat{a}(t) - \hat{a}(t-1)\right]w(t-1)\right) \\ r(t) &= \gamma_2 r(t-1) + (1-\gamma_2)s(t) \\ \nu_0(t) &= \frac{\lambda}{\lambda + P(t-1)h(t-1)^2} \\ \beta(t) &= \begin{cases} 0 & \text{if } r(t-1) < r_0 \\ \dfrac{1}{h(t-1)^2}\left[\nu_0(t) - \nu(t)\right] & \text{if } r(t-1) \geq r_0 \end{cases} \end{aligned} \qquad (4.41)$$

where the two discounting factors γ_1 and γ_2 are 0.85 and 0.95 respectively. The choice of $\nu(t)$ was presented in Figure 4.5. The value of the threshold is

$r_0 = 0.5$, which corresponds to an expected false alarm every 1000:th sample instant. The tank is controlled by a minimum variance regulator with set-point

$$u(t) = h_{ref} - \hat{a}(t) \cdot h(t) \qquad (4.42)$$

For comparison, the problem is first simulated without any fault detection. The result is shown in Figures 4.8 and 4.10. At t=500, the outlet area is increased from 0.01 to 0.011, corresponding to a sudden increase in the outlet flow or a small leak in the tank. This fault is hard to see directly in the output-, input-, or residual sequences. However, looking at the estimated parameter $\hat{a}(t)$, the fault is obvious. For comparison, the sequence $\{w(t)^T w(t)\}$ is also included. This sequence is very sensitive to the fault. In Figure 4.10, the test sequence r(t) is shown. The values of the highest peaks are very unlikely in normal operation, and a fault would have been detected. Note that r(t) has an approximately Gaussian distribution with a standard deviation of 0.16 in case of no fault.

In Figures 4.9 and 4.11, the result of the simulation is given when the fault detection and the modified estimation algorithm are applied. A detection is made after about 30 samples. The increased convergence rate is obvious.

In Figure 4.12, the estimates from the two simulations are compared. Finally, the loss functions in the two simulations are also compared in Figure 4.13. Here the optimal loss function, i.e. the loss function obtained under control with known parameters, is also given.

This simple example has shown that it is possible to improve the estimator by including a device for fault detection. It has also been shown, that the proposed fault detection method is able to detect faults, which have a very small influence on the output signal and the residuals.

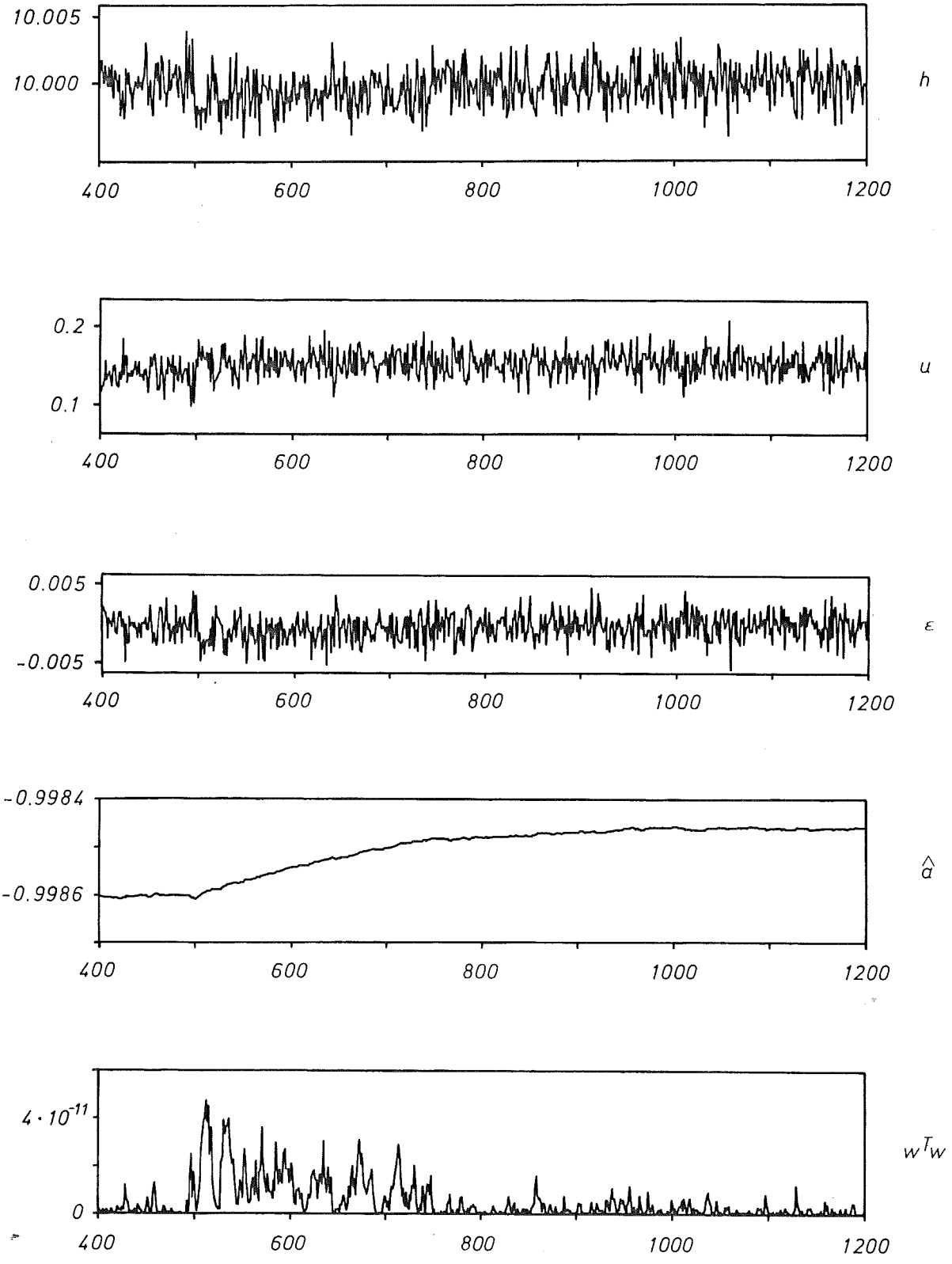


Figure 4.8 - The result of the simulation without fault detection.

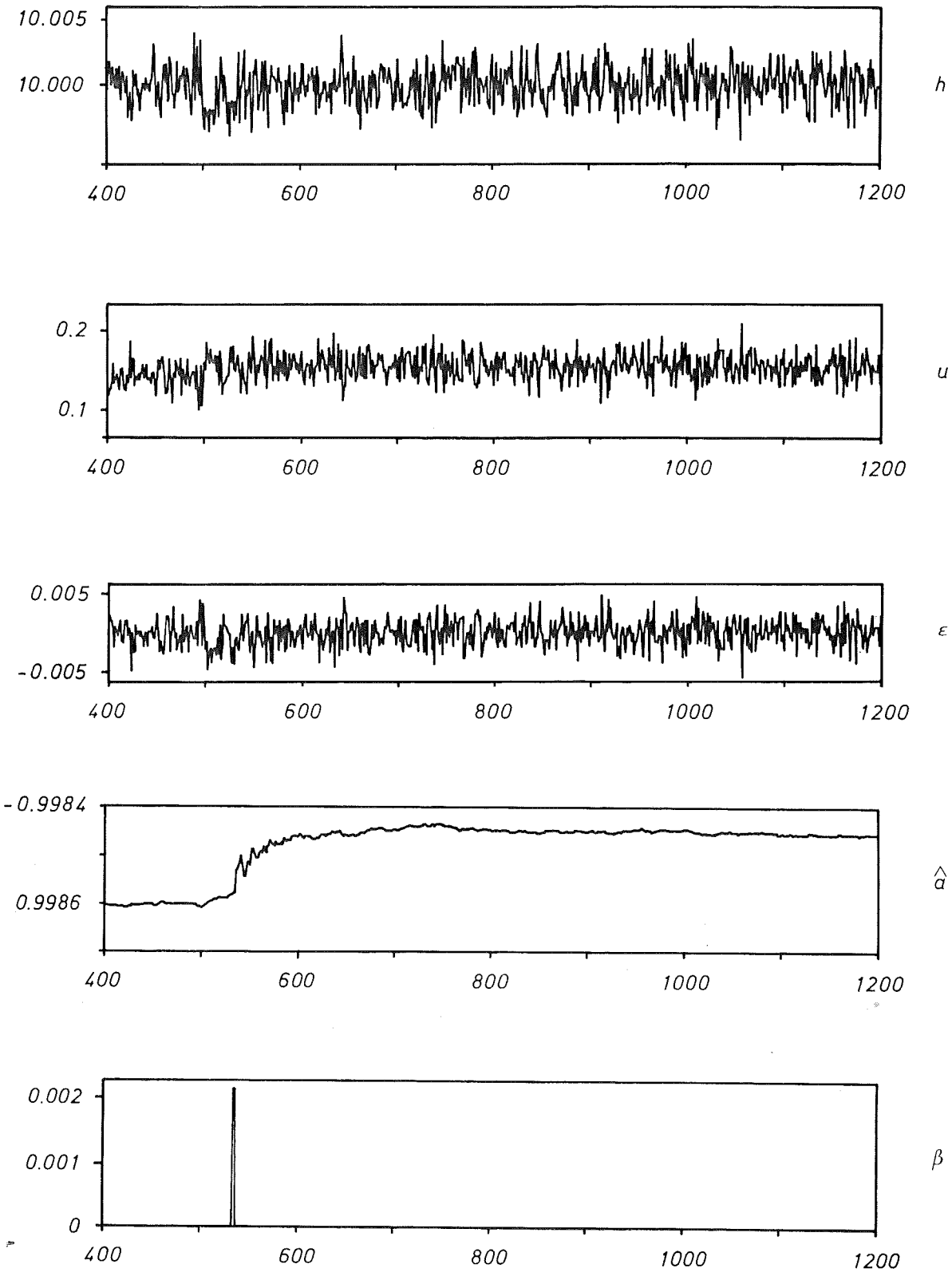


Figure 4.9 - The result of the simulation when the fault detection and the modified estimation algorithm are applied.

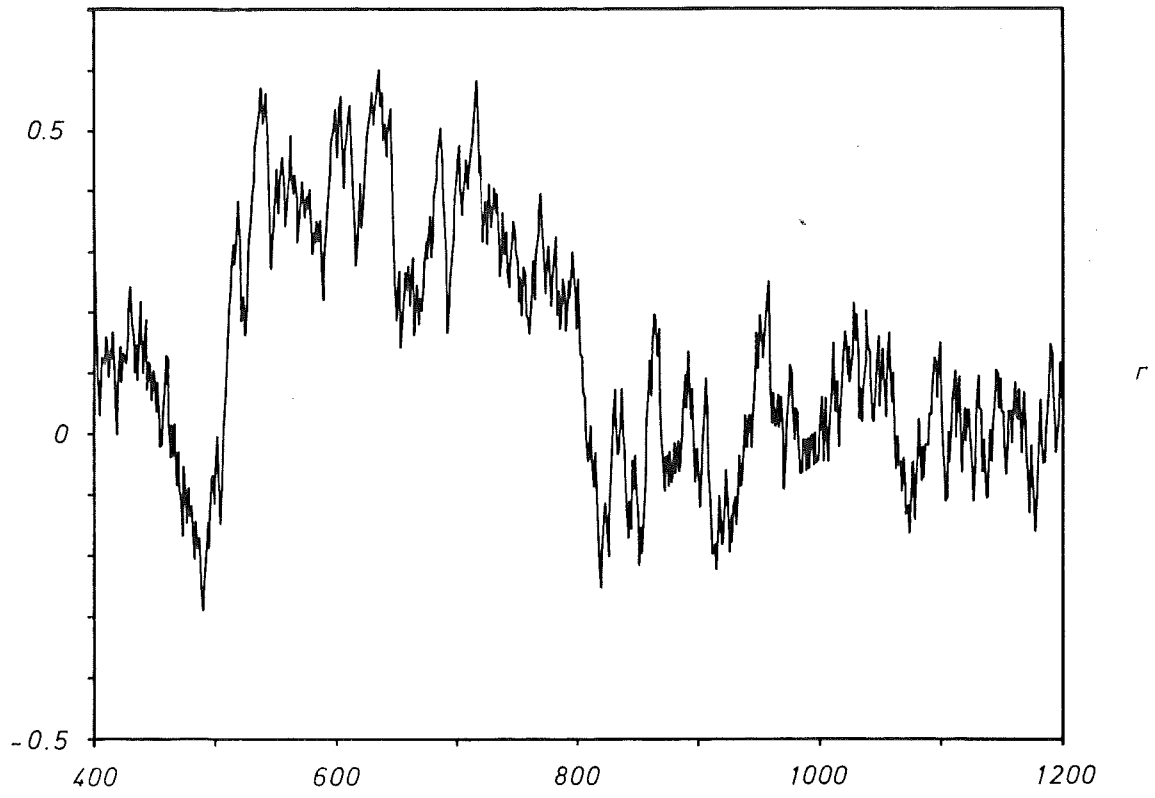


Figure 4.10 - The r(t) sequence when no modification of the estimation algorithm is done.

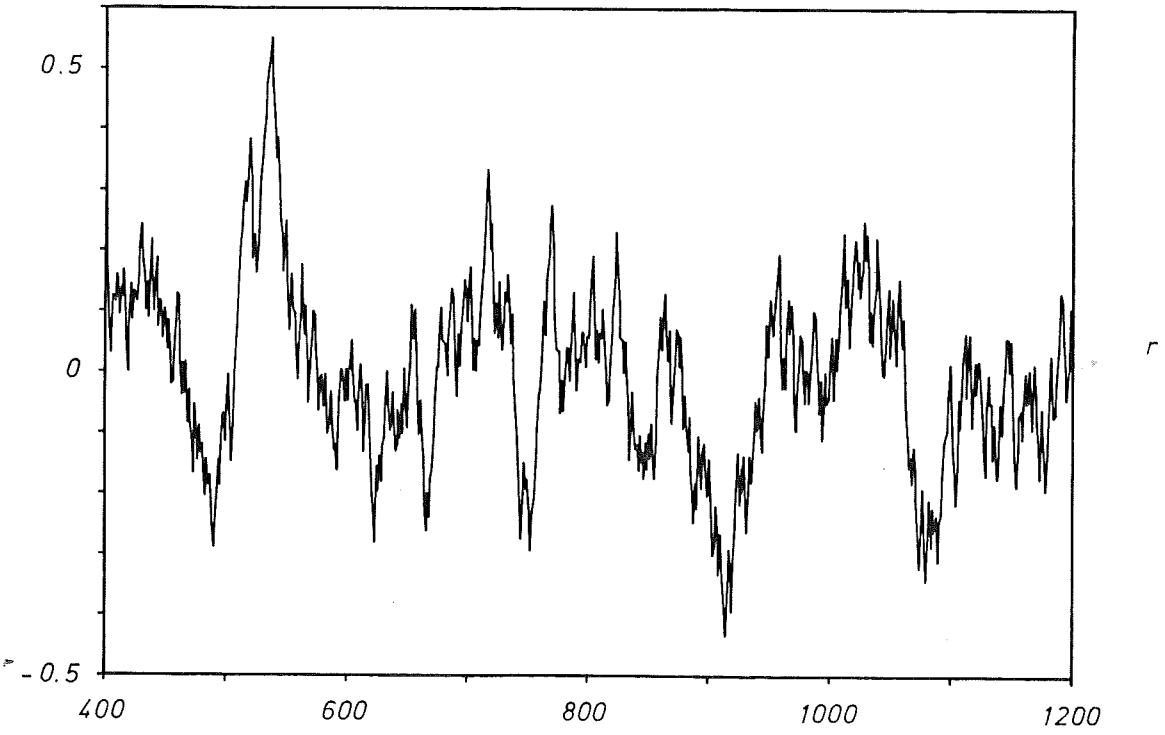


Figure 4.11 - The r(t) sequence when the fault detection and the modified estimation algorithm are applied.

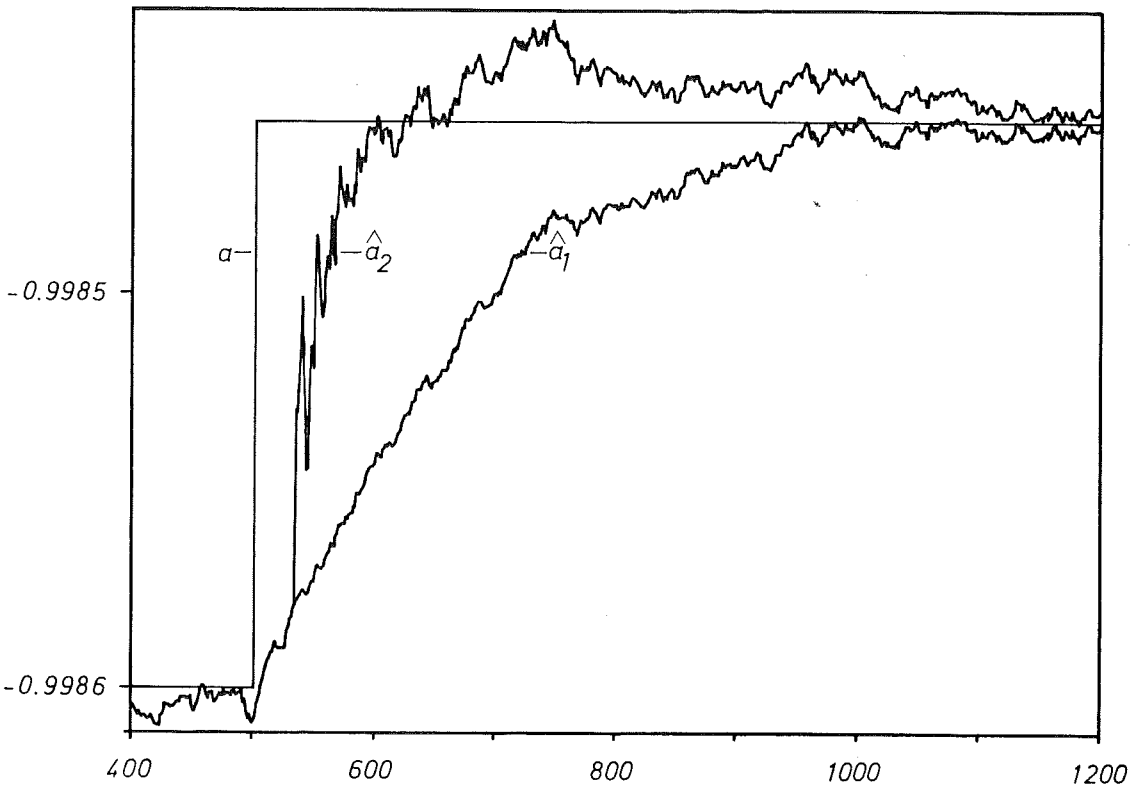


Figure 4.12 - The parameter estimates from the two simulations.

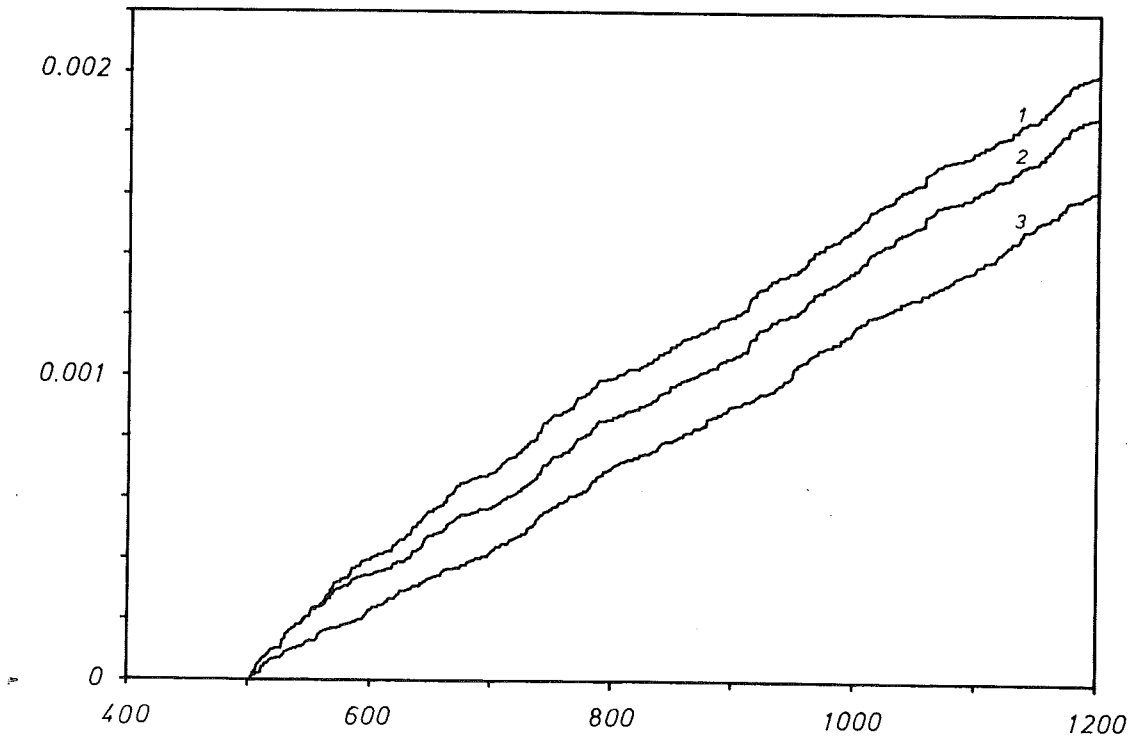


Figure 4.13 - The loss functions in the simulations without fault detection (1), with fault detection (2) and the optimal loss function (3).

4.7 Level estimation in tankers

This section gives another application of the fault detection method. The method is in no way limited to the least squares method. Almost any estimation procedure can be used, since the inputs to the detector are the estimated parameters. To illustrate this, an ordinary Kalman filtering problem is considered.

In tankers carrying fluids, it is important to have accurate estimates of the levels in the different tanks. Loading and unloading must e.g. be made properly in order not to endanger the stability of the tanker. This estimation problem is a typical case where a fault detection procedure is needed.

The noise is highly time-varying, since it mainly originates from the wind and the sea. The noise level in a tank can vary up to a factor of ten. The changes can be very fast, especially when the course of the tanker is changing. Not to bother the operators, it is desirable to filter the level measurements so that the variations do not exceed a few centimetres during times of constant mean level.

It is also important to track changes in the tank levels, caused by filling or emptying, with a small time delay. The low pass filter required for the noise rejection makes the estimator far too slow. Since the start of the filling or emptying of the different tanks mostly is unknown to the operator, and since an automatic feed forward is impossible for security and economical reasons, the filling and emptying can be considered as unknown changes of the system to the estimator. The fault detection procedure described in the previous sections is therefore a feasible way to speed up the rate of convergence of the estimator when a change of the level rate occurs.

The tank model and the Kalman filter

The dynamics of a tank can be described by the following equations

$$\begin{bmatrix} \ell(t+1) \\ \vartheta(t+1) \end{bmatrix} = A \begin{bmatrix} \ell(t) \\ \vartheta(t) \end{bmatrix} + \zeta(t+1)$$

$$y(t) = C \begin{bmatrix} \ell(t) \\ \vartheta(t) \end{bmatrix} + e(t) \qquad (4.43)$$

where $\ell(t)$ [m] denotes the tank level, $\vartheta(t)$ [m/h] the rate of change of the level, $y(t)$ [m] the measurement of the level, and $\zeta(t)$ and $e(t)$ are disturbances. If the sampling period is h [h], the system matrices become

$$A = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \qquad C = (1 \quad 0) \qquad (4.44)$$

In the present application, the value of the sampling period is h = 10.35 [s] = $= 2.875 \cdot 10^{-3}$ [h]. The Kalman filter which minimizes the variance of the estimation error is given by the equations

$$\begin{bmatrix} \hat{\ell}(t+1) \\ \hat{\vartheta}(t+1) \end{bmatrix} = A \begin{bmatrix} \hat{\ell}(t) \\ \hat{\vartheta}(t) \end{bmatrix} + K(t)(y(t) - \hat{\ell}(t))$$

$$K(t) = AP(t)C^T R(t) \qquad (4.45)$$

$$R(t) = \left[CP(t)C^T + r_2(t)\right]^{-1}$$

$$P(t+1) = AP(t)A^T - AP(t)C^T R(t)CP(t)A^T + R_1(t)$$

Here $R_1(t)$ and $r_2(t)$ are the covariance matrices of $\zeta(t)$ and $e(t)$, and $P(t)$ is the covariance matrix of the estimation error. Both $R_1(t)$ and $r_2(t)$ are time-varying and unknown.

The variable $r_2(t)$ corresponds to the measurement noise variance. A constant $r_2(t)$ is used in this example. The value of $r_2(t)$ is chosen to $r_2(t) = 0.01$, corresponding to an average noise variance.

The covariance matrix $R_1(t)$ will be divided into a sum of two matrices

$$R_1(t) = R_1^n(t) + R_1^f(t) \tag{4.46}$$

where $R_1^n(t)$ is the covariance matrix used under normal operation. The matrix $R_1^f(t)$ is nonzero only when a fault is detected.

The matrix $R_1^n(t)$ will be chosen so that the level estimate has a reasonable variance in normal operation. For the sake of simplicity, it is assumed that the uncertainties of the states only enter additively in the velocity equation. It gives a matrix $R_1^n(t)$ with the following structure.

$$R_1^n(t) = \begin{bmatrix} 0 & 0 \\ 0 & r_{22} \end{bmatrix} \tag{4.47}$$

The parameter r_{22} will be chosen as a constant such that the level estimate reaches a desired variance in stationarity. The variance of the level estimate is given by the upper left element of the P(t)-matrix. Combining the Equations (4.44), (4.45), (4.46) and (4.47) gives the following relation between r_{22} and the stationary value of the variance of the level estimate, denoted by p_{11},

$$r_{22} = \frac{p_{11}^4}{h^2(p_{11}+r_2)(p_{11}+2r_2)^2} \tag{4.48}$$

A desired standard deviation of the tank level equal to 1 [cm], i.e. $p_{11} = 1.0 \cdot 10^{-4}$ [m], gives $r_{22} \approx 3.0 \cdot 10^{-6}$.

When a change of the level rate is detected as described previously, the matrix $R_1^f(t)$ is changed from the zero matrix to the diagonal matrix

$$R_1^f(t) = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix} \quad ; \; r(t-1) \geq r_0 \tag{4.49}$$

For the sake of simplicity, the diagonal elements β_1 and β_2 are constant in this example. The value of β_1 is chosen to $\beta_1 = 0.01$, corresponding to an increase of the stationary value of p_{11} by a factor of 100. A reasonable relation between β_1 and β_2 can be obtained from the following arguments. Suppose that a change of the level rate occurs at time t, and that it is detected at time t + nh. Since the level is equal to the integral of the rate of change of the level, the following relation between β_1 and β_2 can be concluded from Equation (4.43)

$$\beta_1 \approx (nh)^2 \beta_2 \tag{4.50}$$

With the parameter choices used in this example, a mean delay time of the detection of about $n \approx 25$ samples occurs. Using $n = 25$ in Equation (4.50) gives $\beta_2 \approx 2$.

The Kalman filter is now defined, except for the details of the detection algorithm. In Section 4.3, the sequence of successive estimate changes $\{\Delta\hat{\theta}(t)\}$ was considered, and shown to have the stochastic features that made the new fault detection procedure possible. In the Kalman filter, the relation between the states is given by the A matrix in Equation (4.45), and not by a unity matrix. The inputs to the detector are therefore given by

$$\Delta_K\hat{\theta}(t) \triangleq \hat{\theta}(t) - A\cdot\hat{\theta}(t-1) \tag{4.51}$$

The complete fault detection procedure is described by the equations

$$\begin{aligned} w(t) &= \gamma_1 w(t-1) + \Delta_K\hat{\theta}(t) \\ s(t) &= \text{sign}\left[\Delta_K\hat{\theta}(t)^T w(t-1)\right] \\ r(t) &= \gamma_2 r(t-1) + (1 - \gamma_2)s(t) \end{aligned} \tag{4.52}$$

where a fault is detected every time the variable r(t) exceeds the threshold

r_0. Since the measurement noise {e(t)} appeared to be slightly coloured, a rather high value of r_0 was used, namely $r_0 = 0.6$. The filter constants used were $\gamma_1 = 0.90$ and $\gamma_2 = 0.95$.

Experimental results

The Kalman filter with fault detection has been applied to several measurement series from different tankers. Results from a tanker filling the tanks with oil in the North Sea is presented below.

Figure 4.14 gives the measurements y(t) of one tank level, recorded during 12 hours. After about 6 hours of constant level, the tank is filled with varying rates. With the precision shown in the graph, the estimates $\hat{x}(t)$ coincides with the measurements.

The estimated rate of change of the level $\hat{s}(t)$ is presented in Figure 4.15. A first glance at the level measurements in Figure 4.14 gives the impression that the filling is made during four distinct periods of constant rate. These four periods are also evident in Figure 4.15. A closer look will however show that the rate is slightly changing even during these periods. See e.g. the small decrease of the filling rate after about 11 hours.

These small changes of the level rate can also explain the relatively high detection frequency shown in Figure 4.16, where the upper left element of the $R_1^f(t)$ matrix is shown. The r(t) sequence is given in Figure 4.17.

To show the behaviour of the estimator in normal operation, the estimated tank level is compared with the measurements of the level between 5 and 6 hours in Figure 4.18. The low pass filtering suppresses the variations to about 1 cm peak to peak.

The rest of the figures demonstrate the behaviour of the estimator between 6 and 7 hours, where a major change of the level rate occurs. Figure 4.19 shows the estimates of the tank level and the level measurements when the fault detection procedure was turned off, i.e. when the matrix $R_1^f(t)$ was zero during the whole experiment. The need for a modification of the Kalman filter like the addition of a fault detection procedure is obvious from this figure.

When the fault detection procedure is used, the level estimate $\hat{x}(t)$ lies totally within the high frequency variations of the measurements y(t). Therefore, the differences between the measurements and the estimates are presented in Figure 4.20. For comparison, the same differences are also shown when the fault detection was turned off, in Figure 4.21.

In Figure 4.22, the estimated rate of change of the level $\hat{\vartheta}(t)$ is shown. Already in Figure 4.14, it is evident that the rate is not increased instantaneously, but the filling rate is gradually increased. From Figure 4.22 it is seen that it took about half an hour to reach the full filling rate.

In Figure 4.23, the times of detection are finally given. Between the 6:th and the 7:th hour, three periods of detection occured.

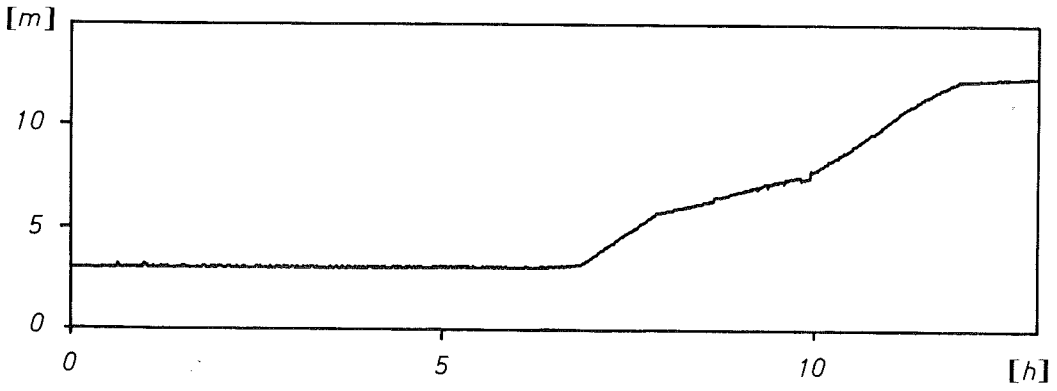


Figure 4.14 - The measurements of the tank level.

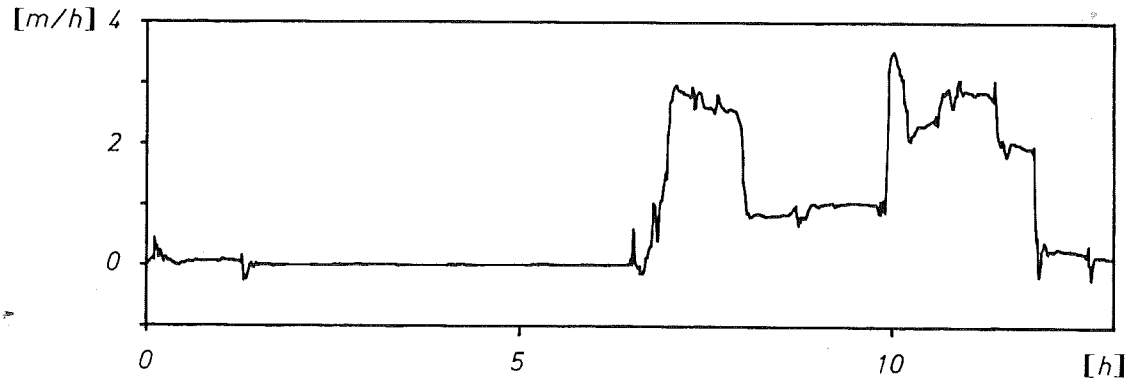


Figure 4.15 - The estimated rate of change of the level.

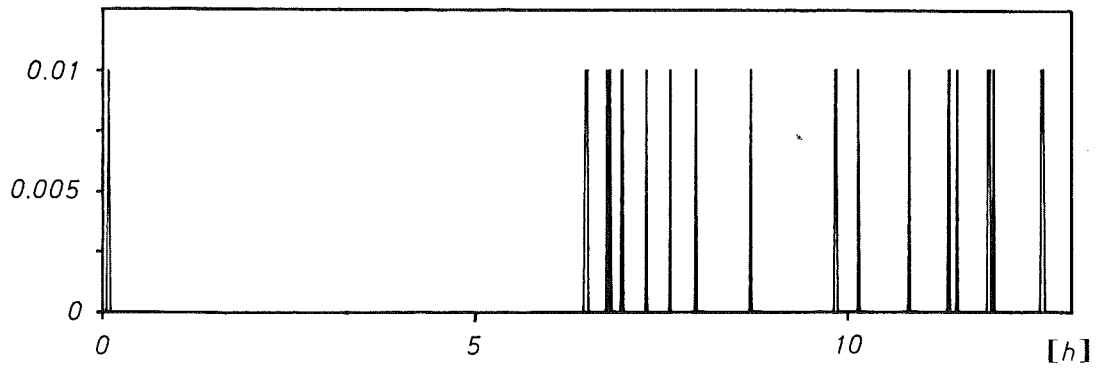


Figure 4.16 - Times of fault detection. The upper left element of $R_1^f(t)$.

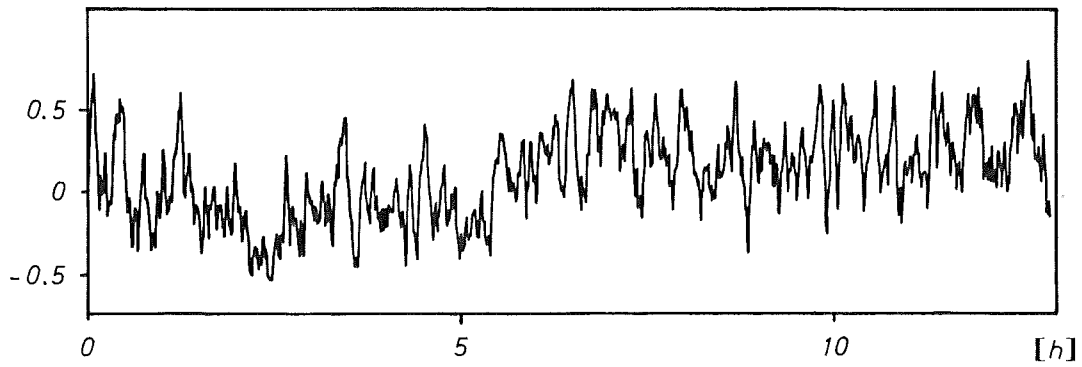


Figure 4.17 - The r(t) sequence.

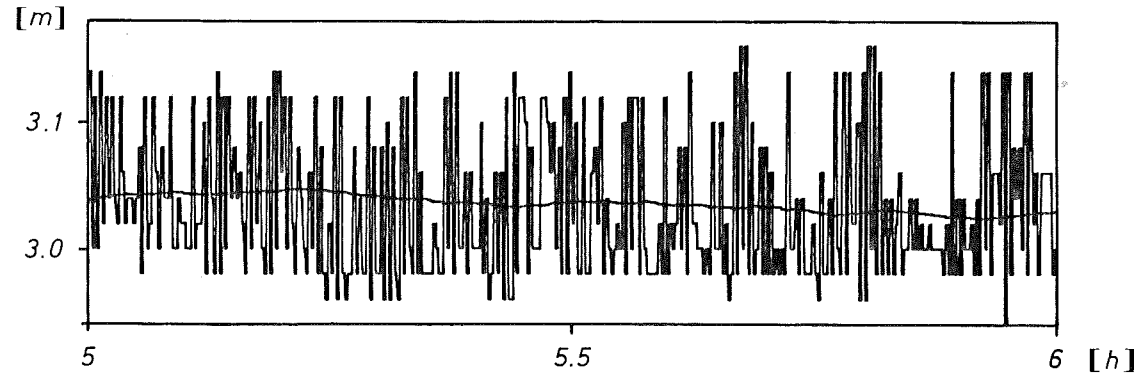


Figure 4.18 - The measurements of the tank level and the estimated tank level during normal operation.

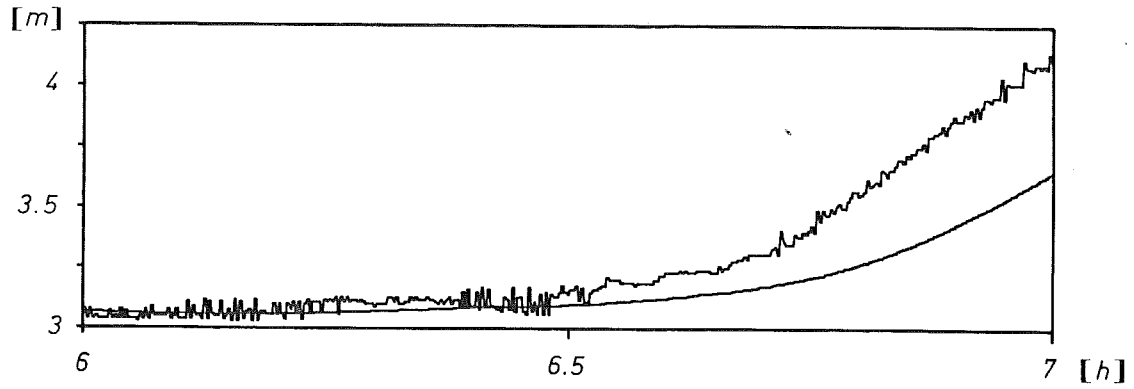


Figure 4.19 - The measurements of the tank level and the estimated tank level, without fault detection.

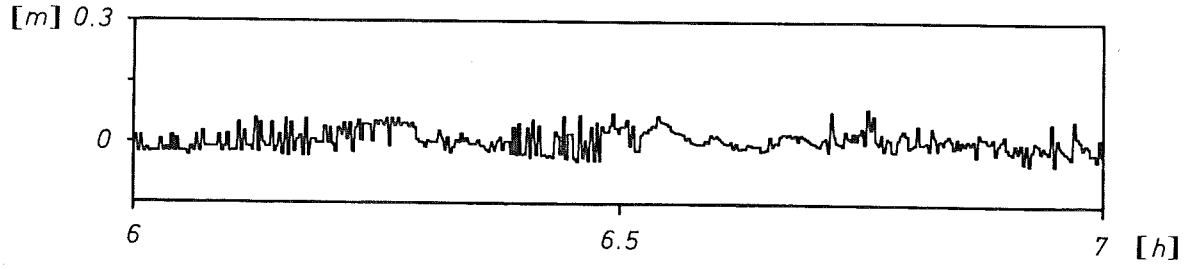


Figure 4.20 - The difference between the measured tank level and the estimated tank level, in case of fault detection.

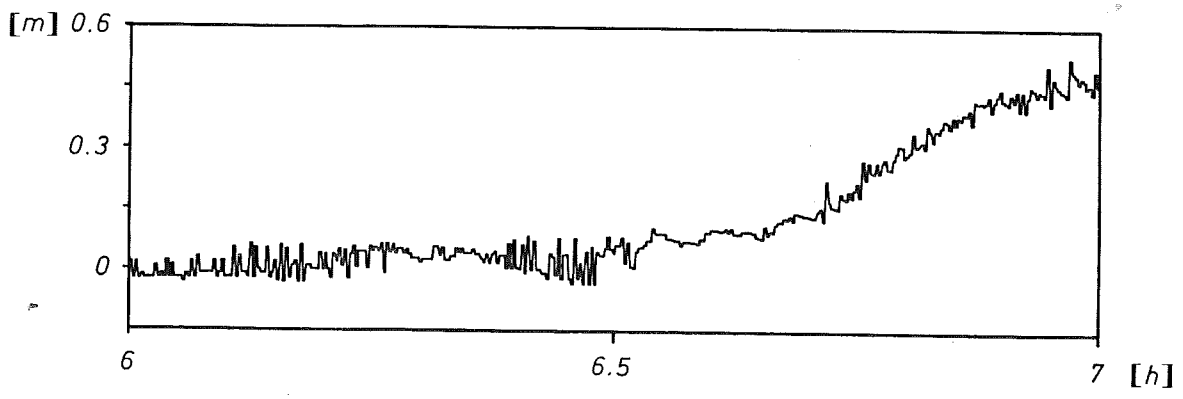


Figure 4.21 - The difference between the measured tank level and the estimated tank level, without fault detection.

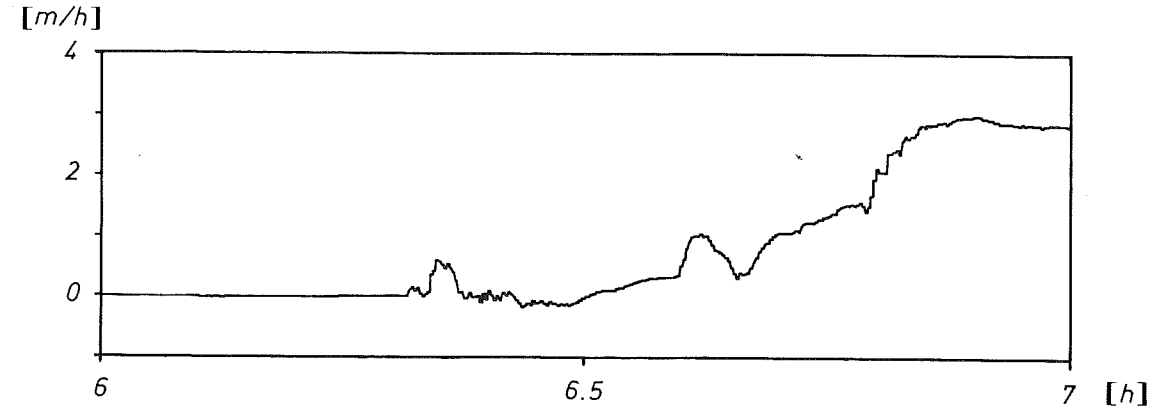


Figure 4.22 - The estimated rate of change of the level.

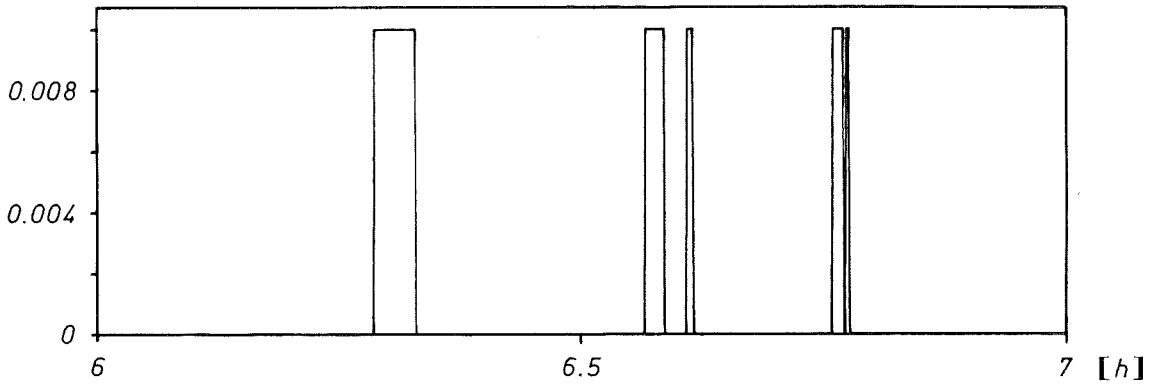


Figure 4.23 - Times of fault detection. The upper left element of $R_1^f(t)$.

5. SLOW PARAMETER CHANGES

In the previous chapter, the estimation problem was discussed for large parameter changes. Special methods could be used, since large parameter changes can be detected easily. This chapter deals with slow parameter changes and variations in the excitation. Based on the weighting problem described in Chapter 3, a new principle of forgetting old measurements is given in Section 5.1. This approach was first presented in Hägglund (1983). An algorithm which implements the principle is presented in the following sections, where the algorithm is also analysed.

5.1 A new solution to the weighting problem

In Section 3.2, it was argued that a reasonable weighting of the measurements could be made if Assumption 3.1 and certain additional conditions hold. It was also shown, however, that severe problems could occur if these assumptions were violated.

In this section, a more general problem will be treated. It is assumed that Assumption 3.1 hold and that the parameter variations are slow. This corresponds to case 4 in Table 3.1.

Information handling in an LS estimator poses two different types of problems. The first one is that the incoming information may be poor, because of bad excitation or large variations of the parameters. This problem can be solved by dual control, which gives active excitation when the incoming information is poor.

The second problem occurs when the incoming information is sufficient, but it is handled incorrectly as described in Section 3.2. In other words, the problem is caused by bad match between real and estimated values of $\sigma_m(t,i)$ and $\sigma_n(i)$. The situation can be improved by better estimates of these variances. The fault detection procedure described in Chapter 4 illustrates one way to provide a better estimate of $\sigma_m(t,i)$. Remember that it is very important to distinguish between $\sigma_m(t,i)$ and $\sigma_n(i)$, since they imply opposite

actions.

The accuracy obtainable in estimating the variances is unfortunately often rather poor. In Chapter 4, it was shown that large increases of $\sigma_m(t,i)$ can be found by fault detection. There will however always be some time delay between the parameter change and the detection, and the estimate may be of the qualitative form "$\sigma_m(t,i)$ has increased". It is possible to get a fairly accurate estimate of $\sigma_n(i)$ under stationary conditions. It will however take some time before fast changes of the noise level are distinguished from parameter changes. The a priori knowledge about $\sigma_m(t,i)$ and $\sigma_n(i)$ is mostly poor.

The purpose of the LS estimator is to provide estimates of the parameter vector $\theta(t)$ with a reasonable accuracy. The weighting problem becomes troublesome because of the lacking knowledge about the uncertainties of the measurements. Instead of using assumptions of variations of the parameters and the noise levels, a different approach will be suggested, where the accuracy of the estimates $\hat{\theta}(t)$ is considered. As will be shown, several problems can be avoided by relating the weighting directly to the accuracy, i.e. the amount of information available, and to the incoming information. The following principle is therefore proposed.

> Discount past data in such a way that a constant desired amount of information is retained, if the parameters are constant.

A quantitative measure of information is defined in Equation (3.12) as the inverse of the P-matrix. More precisely, the algorithm given in the following sections gives a P-matrix which is proportional to the identity matrix. The diagonal elements of P(t) may be interpreted as approximations of the variances of the corresponding parameters. The weights $\omega(t,i)$ are therefore chosen so that these variances get a desired value. The time horizon will consequently vary, depending on the incoming information. If the signals are noisy with a small information content, the time horizon will be long. If no information is coming in at all, nothing will be forgotten. If on the other hand the incoming information content is large, old measurements can be discounted quickly and a fast adaptation to new parameter values is possible.

The method requires an estimate of the noise level if the interpretation of the P-matrix as the covariance matrix shall make sense.

It should be emphasized that the method is not restricted to the case where P converges to a diagonal matrix. Any positive definite matrix could be used as the desired limit of the P-matrix. It will however be shown later that it is often wise to choose the limit P-matrix proportional to the identity matrix.

5.2 Updating the covariance matrix

The goal for the estimator is to weight the incoming data in such a way that the P-matrix becomes proportional to the identity matrix, say $a \cdot I$. The value of a is the desired variance of the parameter estimates. The equation for updating the P-matrix will now be derived.

According to Equations (3.9) and (3.12), the LS estimator is given by

$$\hat{\theta}(t) = \left[\Phi(t)^T V(t)^{-1} \Phi(t)\right]^{-1} \Phi(t)^T V(t)^{-1} Y(t) =$$

$$= P(t)\Phi(t)^T V(t)^{-1} Y(t) \qquad (5.1)$$

Let $V(t;t+k)$ denote the upper left quadratic t-dimensional submatrix of $V(t+k)$. The updating of the information matrix is then given by

$$P(t)^{-1} = \Phi(t)^T V(t)^{-1} \Phi(t) =$$

$$= \left[\Phi(t-1)^T \; \varphi(t)\right] \begin{bmatrix} V(t-1;t)^{-1} & 0 \\ 0 & v(t)^{-1} \end{bmatrix} \begin{bmatrix} \Phi(t-1) \\ \varphi(t)^T \end{bmatrix} =$$

$$= \Phi(t-1)^T V(t-1;t)^{-1} \Phi(t-1) + \varphi(t) v(t)^{-1} \varphi(t)^T \qquad (5.2)$$

The second term in Equation (5.2) represents the new incoming information. The inverse P-matrix is changed by a rank one matrix proportional to $\varphi(t)\varphi(t)^T$. With some abuse of language, it will be said that the inverse P-matrix is "changed in the direction of φ".

If the discounting is made according to the new principle, data are in stationarity only discounted in the direction where the new information is entering. This way of discounting is used even during the transient periods. Therefore, choose V(t-1;t) such that

$$\Phi(t-1)^T V(t-1;t)^{-1}\Phi(t-1) = \Phi(t-1)^T V(t-1)^{-1}\Phi(t-1) - \\ - \alpha(t)\varphi(t)\varphi(t)^T \qquad (5.3)$$

where $\alpha(t)$ is a scalar. The P(t)-matrix is then updated as

$$P(t) = \left[\Phi(t-1)^T V(t-1)^{-1}\Phi(t-1) + \left[v(t)^{-1}-\alpha(t)\right]\varphi(t)\varphi(t)^T\right]^{-1} = \\ = P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{\left[v(t)^{-1}-\alpha(t)\right]^{-1} + \varphi(t)^T P(t-1)\varphi(t)} \qquad (5.4)$$

or equivalently, the information matrix is updated as

$$P(t)^{-1} = P(t-1)^{-1} + \left[v(t)^{-1}-\alpha(t)\right]\varphi(t)\varphi(t)^T \qquad (5.5)$$

Obviously, $\alpha(t)$ must be nonnegative, since a negative $\alpha(t)$ would mean an addition of information instead of a removal. From Equation (5.5) it is seen that $\alpha(t)$ should be equal to $v(t)^{-1}$ in stationarity, i.e. when the P-matrix has converged to its desired value. From the equations above it is also concluded that the P-matrix may become nonpositive if $\alpha(t)$ is chosen too large. The following theorem gives bounds on $\alpha(t)$ such that the P-matrix is positive definite if $\alpha(t)$ is chosen within those bounds.

Theorem 5.1: Given a sequence of matrices {P(t)} which satisfy Equation (5.4). If the initial matrix P(0) is positive definite, then P(t) will be positive definite for all t if and only if $\alpha(t)$ lies within the bounds

$$0 \le \alpha(t) < v(t)^{-1} + \frac{1}{\varphi(t)^T P(t-1)\varphi(t)} \qquad (5.6)$$

Proof: The proof is divided into two parts. First consider the case when $\alpha(t)$ is in the interval

$$0 \leq \alpha(t) \leq v(t)^{-1}$$

The inverse P-matrix is updated according to Equation (5.5). Since the second term is nonnegative, it is clear from recursion that the inverse P-matrices and consequently the P-matrices are positive definite.

If $\alpha(t)$ is chosen in the interval

$$v(t)^{-1} < \alpha(t) < v(t)^{-1} + \frac{1}{\varphi(t)^T P(t-1)\varphi(t)}$$

the second term in Equation (5.4) always becomes positive. By recursion it is therefore concluded that the P-matrices stay positive definite even for $\alpha(t)$ in this interval. It is also easily verified that P(t) becomes nonpositive if $\alpha(t)$ is greater than or equal to the upper bound in the inequality (5.6).

□

Summing up, the new method for updating the P-matrix is given by Equation (5.4) with the bounds on $\alpha(t)$ given in Theorem 5.1. Later on, $\alpha(t)$ will be further restricted so that the desired properties of the estimator are obtained.

An interesting property of the new algorithm can already now be seen, by comparing the new updating formula for the information matrix, Equation (5.5), with the corresponding equation for exponential forgetting. From Equation (3.13), the information matrix is updated as

$$P(t)^{-1} = P(t-1)^{-1} - (1-\lambda)\cdot P(t-1)^{-1} + v(t)^{-1}\varphi(t)\varphi(t)^T \tag{5.7}$$

when a forgetting factor λ is used. Both in Equation (5.5) and in Equation (5.7), the information at time t can be expressed as a sum of three terms: The information at time t-1, the old information which is removed, and the new information. The added information is the same in both cases. It consists of a rank one matrix, i.e. information is only added in one direction. Equation

(5.7), corresponds to information being removed in all directions, since the second term is a full rank matrix. In the new updating formula given by Equation (5.5), information is however only removed in the same direction as the new added information. This property will enable the desired control of the P-matrix.

5.3 Updating the parameter estimates

The new way of updating the P-matrix described above will also influence the updating of the estimates. The new equations will be derived from Equation (5.1).

$$\hat{\theta}(t) = P(t)\Phi(t)^T V(t)^{-1} Y(t) =$$

$$= P(t)\begin{bmatrix}\Phi(t-1)^T & \varphi(t)\end{bmatrix}\begin{bmatrix} V(t-1;t)^{-1} & 0 \\ 0 & v(t)^{-1}\end{bmatrix}\begin{bmatrix} Y(t-1) \\ y(t)\end{bmatrix} =$$

$$= P(t)\left[\Phi(t-1)^T V(t-1;t)^{-1} Y(t-1) + \varphi(t)v(t)^{-1}y(t)\right] \quad (5.8)$$

This expression can generally not be evaluated without a more precise determination of V(t-1;t) than provided by Equation (5.3). Assume however for a moment that no disturbances are acting on the system. Since the parameters θ are assumed to be constant, Y(t-1) is then equal to $\Phi(t-1)\hat{\theta}(t-1)$, and the estimate at time t is given by

$$\hat{\theta}(t) = P(t)\left[\Phi(t-1)^T V(t-1;t)^{-1}\Phi(t-1)\hat{\theta}(t-1) + \varphi(t)v(t)^{-1}y(t)\right] \quad (5.9)$$

This expression can be evaluated using the Equation (5.3).

$$\hat{\theta}(t) = P(t)\left[\Phi(t-1)^T V(t-1)^{-1}\Phi(t-1)\hat{\theta}(t-1) + \right.$$

$$\left. + \left[v(t)^{-1}y(t) - \alpha(t)\varphi(t)^T\hat{\theta}(t-1)\right]\varphi(t)\right] \quad (5.10)$$

By applying the new updating formula for the P-matrix, Equation (5.4), the following equation is obtained after some calculations

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P(t-1)\varphi(t)\varepsilon(t)}{v(t) + \varphi(t)^T P(t-1)\varphi(t)\left[1-\alpha(t)v(t)\right]} \tag{5.11}$$

or

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{v(t)} P(t)\varphi(t)\varepsilon(t) \tag{5.12}$$

where

$$\varepsilon(t) = y(t) - \varphi(t)^T\hat{\theta}(t-1) \tag{5.13}$$

In the general case, when disturbances are acting on the system, Equation (5.8) can not be evaluated without a more detailed determination of V(t-1;t). The full $\Phi(t-1)$, V(t-1) and Y(t-1) matrices are not stored in the recursive estimator, but the information is stored in terms of the parameter estimates $\hat{\theta}(t)$, the corresponding covariance matrix P(t) and the state vector $\varphi(t)$. The remaining freedom to choose V(t-1;t) will be used to retain this appealing structure of the estimator. When y(t) is equal to the predicted output value, i.e. $y(t) = \varphi(t)^T\hat{\theta}(t-1)$, no change of the estimated parameters will be made. Hence, letting $y(t) = \varphi(t)^T\hat{\theta}(t-1)$ in Equation (5.8) should yield $\hat{\theta}(t) = \hat{\theta}(t-1)$. Using this additional requirement together with Equation (5.4) gives

$$\Phi(t-1)^T V(t-1;t)^{-1} Y(t-1) =$$

$$= \Phi(t-1)^T V(t-1)^{-1} Y(t-1) - \alpha(t)\varphi(t)^T\hat{\theta}(t-1)\cdot\varphi(t) \tag{5.14}$$

Inserting this expression in Equation (5.8) gives, after some calculations, the same updating formula as in the noise free case, i.e. Equations (5.11) and (5.12) are true even in the general case.

5.4 Choice of $\alpha(t)$

In the two previous sections, the new updating formulas for the parameter vector $\hat{\theta}(t)$ and the covariance matrix P(t) were derived. They are given by Equations (5.4) and (5.12). To complete the algorithm, the discounting factor $\alpha(t)$ should also be determined. In this section, a choice of $\alpha(t)$ is presented

which makes the P-matrix converge to the desired matrix $a \cdot I$.

First of all, some limitations on $\alpha(t)$ will be introduced. The quantity $\alpha(t)$ is a measure of the amount of information discounted at time t. In Theorem 5.1 it was shown that the P-matrix remained positive definite if $\alpha(t)$ belonged to the interval

$$0 \leq \alpha(t) < v(t)^{-1} + \frac{1}{\varphi(t)^T P(t-1)\varphi(t)} \qquad (5.15)$$

A removal of too much information would give an unstable estimator. It is not within the scope of this thesis to give a complete stability proof for the new estimation scheme in the considered general time-varying case. A stability investigation is however made for the restricted case of constant parameters and no noise.

A crucial part in most stability proofs of adaptive controllers is to show that the function

$$f(t) = \tilde{\theta}(t)^T P(t)^{-1} \tilde{\theta}(t) \qquad (5.16)$$

where

$$\tilde{\theta}(t) = \theta - \hat{\theta}(t) \qquad (5.17)$$

is decreasing. In the following theorem, that requirement is used to give another bound on $\alpha(t)$.

Theorem 5.2: Assume that the parameters $\theta(t)$ are constant and that no disturbances are acting on the system. If $\alpha(t)$ satisfies the requirements of Theorem 5.1, then the function $f(t)$ given by Equation (5.16) decreases if and only if

$$\alpha(t) \leq \frac{1}{\varphi(t)^T P(t-1)\varphi(t)} \left[\frac{1}{2} + \frac{1}{2} \sqrt{1 + \frac{4\varphi(t)^T P(t-1)\varphi(t)}{v(t)}} \right] \qquad (5.18)$$

Proof: In the noise-free case, the residuals $\varepsilon(t)$ are given by

$$\varepsilon(t) = y(t) - \varphi(t)^T\hat{\theta}(t-1) = \varphi(t)^T\tilde{\theta}(t-1) \qquad (5.19)$$

From Equation (5.12), the propagation of the estimation error can therefore be written as

$$\tilde{\theta}(t) = A(t)\cdot\tilde{\theta}(t-1) \qquad (5.20)$$

where

$$A(t) \triangleq I - \frac{P(t-1)\varphi(t)\varphi(t)^T}{v(t) + \varphi(t)^T P(t-1)\varphi(t)\left[1-\alpha(t)v(t)\right]} \qquad (5.21)$$

Hence the function f(t) becomes

$$f(t) = \tilde{\theta}(t)^T P(t)^{-1}\tilde{\theta}(t) = \tilde{\theta}(t-1)^T A(t)^T P(t)^{-1} A(t)\tilde{\theta}(t-1) \qquad (5.22)$$

It follows from Equation (5.5) that the inverse of the P-matrix satisfies

$$P(t)^{-1} = P(t-1)^{-1} + \left[v(t)^{-1}-\alpha(t)\right]\varphi(t)\varphi(t)^T \qquad (5.23)$$

The following equality can be obtained from the Equations (5.21) and (5.23).

$$A(t)^T P(t)^{-1} A(t) = A(t)^T\left[P(t-1)^{-1} - \alpha(t)\varphi(t)\varphi(t)^T\right] =$$

$$= P(t-1)^{-1} - \frac{1 + \alpha(t)v(t)\left[1-\alpha(t)\varphi(t)^T P(t-1)\varphi(t)\right]}{v(t) + \varphi(t)^T P(t-1)\varphi(t)\left[1-\alpha(t)v(t)\right]}\varphi(t)\varphi(t)^T \qquad (5.24)$$

Inserting Equation (5.24) into (5.22) gives

$$f(t) = f(t-1) - \frac{1 + \alpha(t)v(t)\left[1 - \alpha(t)\varphi(t)^T P(t-1)\varphi(t)\right]}{v(t) + \varphi(t)^T P(t-1)\varphi(t)\left[1-\alpha(t)v(t)\right]}\varepsilon(t)^2 \qquad (5.25)$$

The denominator of the second term in Equation (5.25) is always positive if

$\alpha(t)$ satisfies the requirements of Theorem 5.1. Hence $f(t)$ is decreasing if and only if

$$1 + \alpha(t)v(t)\left[1 - \alpha(t)\varphi(t)^T P(t-1)\varphi(t)\right] \geq 0 \qquad (5.26)$$

For values of $\alpha(t)$ within the bounds given in Theorem 5.1 this is achieved if and only if Relation (5.18) holds.

□

Theorem 5.1 and Theorem 5.2 limit the possible choice of $\alpha(t)$. For the sake of convenience, $\alpha(t)$ will in the sequel be chosen such that

$$0 \leq \alpha(t) \leq \frac{1}{\varphi(t)^T P(t-1)\varphi(t)} \qquad (5.27)$$

This condition implies that the requirements of both Theorem 5.1 and Theorem 5.2 are satisfied. Within these bounds, $\alpha(t)$ should be chosen so that information is discounted according to the principle given in Section 5.1. This can be done in many ways, so the method given below is not the only solution.

The goal is to obtain a diagonal P-matrix of the form $a \cdot I$. Equation (5.4) shows that the P-matrix is changed along the $P(t-1)\varphi(t)$ direction in each iteration. Theorem 5.3 below shows that the desired property of the estimator should be obtained if $\alpha(t)$ could be chosen such that

$$\frac{\varphi(t)^T P(t-1)P(t)P(t-1)\varphi(t)}{\varphi(t)^T P(t-1)P(t-1)\varphi(t)} = a \qquad (5.28)$$

Eliminating $P(t)$ in Equation (5.28) using Equation (5.4) gives the following desired value of $\alpha(t)$

$$\alpha_d(t) = v(t)^{-1} + \frac{\delta_d(t)}{\delta_d(t)\varphi(t)^T P(t-1)\varphi(t) - 1} \qquad (5.29)$$

where

$$\delta_d(t) = \frac{1}{\varphi(t)^T P(t-1)^2 \varphi(t)} \left[\frac{\varphi(t)^T P(t-1) P(t-1) P(t-1) \varphi(t)}{\varphi(t)^T P(t-1) P(t-1) \varphi(t)} - a \right] \tag{5.30}$$

From Equation (5.29), the auxiliary variable $\delta_d(t)$ can also be expressed as

$$\delta_d(t) = \frac{1}{\left[v(t)^{-1} - \alpha_d(t)\right]^{-1} + \varphi(t)^T P(t-1) \varphi(t)} \tag{5.31}$$

The value of $\alpha(t)$ given by Equation (5.29) can however not always be used, because of the restrictions on $\alpha(t)$. Before treating these restrictions, the physical interpretation of the variable $\delta(t)$ will be given. Let $\delta(t)$ be defined analogously to $\delta_d(t)$, but with $\alpha_d(t)$ substituted by $\alpha(t)$. From Equation (5.31), Equation (5.4) can then be written as

$$P(t) = P(t-1) - \delta(t) \cdot P(t-1) \varphi(t) \varphi(t)^T P(t-1) \tag{5.32}$$

The variable $\delta(t)$ may thus be interpreted as a gain in the equation for updating the P-matrix.

It is not trivial to handle the bounds on $\alpha(t)$. To get some insight, the variable $\delta_d(t)$ is therefore plotted versus $\alpha_d(t)$ in Figure 5.1, From this figure and from Relation (5.27) the following choice of $\alpha(t)$ is concluded.

$$\alpha = \begin{cases} 0 & \alpha_d \leq 0 \\ \alpha_d & 0 < \alpha_d \leq \dfrac{1}{\varphi^T P \varphi} \\ \dfrac{1}{\varphi^T P \varphi} & \dfrac{1}{\varphi^T P \varphi} < \alpha_d \leq v^{-1} + \dfrac{1}{\varphi^T P \varphi} \\ 0 & \alpha_d > v^{-1} + \dfrac{1}{\varphi^T P \varphi} \end{cases} \tag{5.33}$$

The bounds on α given by Equation (5.33) can also be expressed as bounds on δ according to

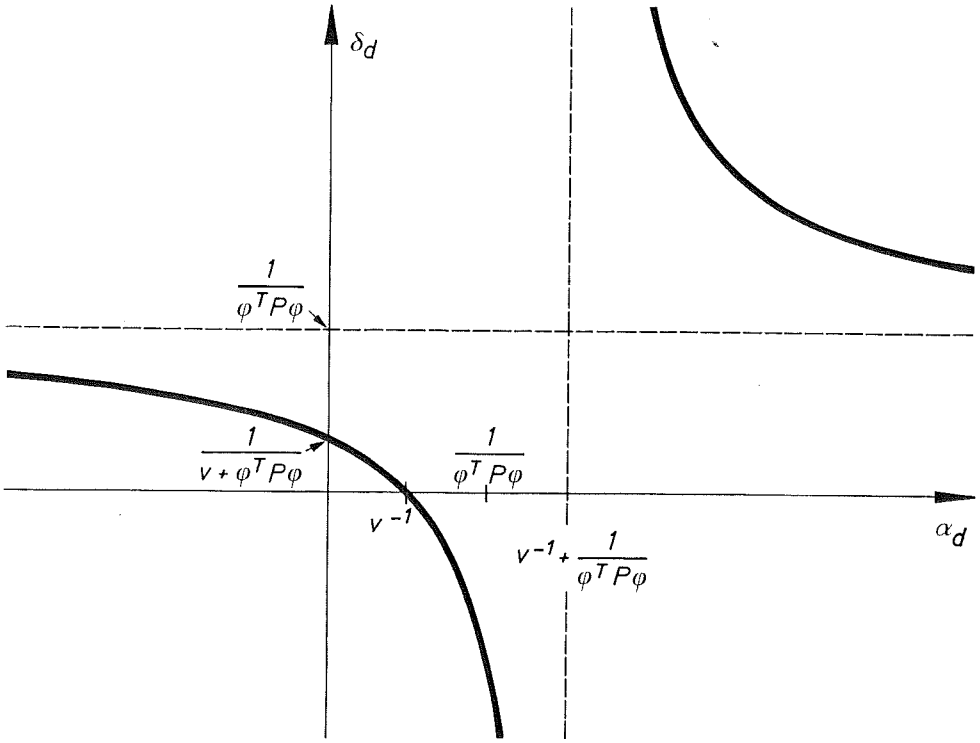


Figure 5.1 - $\delta_d(t)$ versus $\alpha_d(t)$.

$$\delta = \begin{cases} \dfrac{1}{v+\varphi^T P \varphi} & \delta_d \geq \dfrac{1}{v+\varphi^T P \varphi} \\ \delta_d & \dfrac{1}{\varphi^T P \varphi}\left(1 - \dfrac{v}{\varphi^T P \varphi}\right) \leq \delta_d < \dfrac{1}{v+\varphi^T P \varphi} \\ \dfrac{1}{\varphi^T P \varphi}\left(1 - \dfrac{v}{\varphi^T P \varphi}\right) & \delta_d < \dfrac{1}{\varphi^T P \varphi}\left(1 - \dfrac{v}{\varphi^T P \varphi}\right) \end{cases} \tag{5.34}$$

Comparing this equation with Equation (5.33), shows that even if α is a discontinuous function of α_d, the gain δ in the updating formula of the P-matrix is a continuous function of the desired gain δ_d. In a practical implementation, it may be more appealing to introduce δ directly in the updating formula of the P-matrix, as in Equation (5.32), and to use the bounds on $\delta(t)$ given by Equation (5.34).

5.5 Possible convergence points and adaptation rate

This section contains a short discussion of possible convergence points of the P-matrix, and of the adaptation rate of the new estimator.

The P-matrix will not converge to $a \cdot I$, if the parameter a is chosen too large, because the conditions for stability of Theorem 5.2 would then not be satisfied. From Equation (5.32), it is obvious that a stationary solution to the updating equation of the P-matrix may be obtained if and only if $\delta = 0$. To reach the desired stationary solution $P = a \cdot I$, the desired gain δ_d must also be zero. It then follows from Equation (5.34) that

$$v(t)^{-1} \leq \frac{1}{\varphi(t)^T P(t-1)\varphi(t)} \tag{5.35}$$

See also Figure 5.1. This inequality is not valid if the parameter a is chosen too large.

If the inequality (5.35) is not satisfied, the P-matrix will not converge to $P = a \cdot I$, but it will be smaller. This occurs when the desired parameter variance, a, is too large, e.g. due to a high signal to noise ratio. This is rarely any problem in practice, but it will show up in the convergence analysis in the next section. The pleasant solution to the problem is of course to accept a higher parameter accuracy or choose a smaller value of the parameter a when the signal to noise ratio is "too" high.

When discussing the LS algorithm in Section 4.5, the estimation errors were represented by the following equation.

$$\tilde{\theta}(t) = U(t) \cdot \tilde{\theta}(t-1) \tag{4.33}$$

All eigenvalues of U(t) are equal to one, except the one which corresponds to the eigenvector $P(t)\varphi(t)$. This eigenvalue, denoted by ν_0, can be interpreted as a measure of the tracking capability of the algorithm.

The most rapid adaptation rate will be obtained if $\nu_0 = 0$. This means that the projection of the error on the eigenvector $P(t)\varphi(t)$ is brought to zero in one step. Under conditions of persistent excitation, the full estimation error will then be zero in n steps, where n is the number of parameters.

Equation (4.33) also holds for the new algorithm. The eigenvalue ν_0 is easily calculated from Equations (5.20) and (5.21). It is given by

$$\nu_0 = 1 - \frac{\varphi(t)^T P(t-1)\varphi(t)}{v(t) + \left[1 - \alpha(t)v(t)\right]\varphi(t)^T P(t-1)\varphi(t)} \qquad (5.36)$$

In stationarity, when $\alpha(t) = v(t)^{-1}$, the eigenvalue becomes

$$\nu_0 = 1 - \frac{\varphi(t)^T P(t-1)\varphi(t)}{v(t)} \qquad (5.37)$$

This equation shows clearly how the tracking capability, or equivalently the time horizon of the estimator, is influenced by the signal to noise ratio.

It is interesting to see how the restrictions on the possible convergence points of the P-matrix influences the adaptation rate. A smaller P-matrix will normally give a slower adaptation rate. Notice that ν_0 would be negative if $\alpha(t)$ were chosen as $v(t)^{-1}$ when the inequality (5.35) is not satisfied. The bounds on $\alpha(t)$ given by Equation (5.33) guarantee, however, that ν_0 stays nonnegative. In particular when $\alpha(t)$ is equal to the upper bound in Figure 5.1, i.e. $\alpha(t) = 1/\varphi(t)^T P(t-1)\varphi(t)$, the eigenvalue is zero. Therefore, the restrictions above do not limit the rate of adaptation.

5.6 Proof of convergence

It remains to show that the P-matrix actually converges to the constant matrix $a \cdot I$ when $\alpha(t)$ is chosen according to the previous sections. The proof is based on the following two lemmas.

Lemma 5.1: Given a sequence of symmetrical matrices {S(t)} which are updated according to

$$S(t) = S(t-1) - k(t)x(t)x(t)^T \tag{5.38}$$

where {x(t)} are normalized vectors and the scalar k(t) is chosen as

$$k(t) = \varkappa(t)x(t)^T S(t-1)x(t) \triangleq \varkappa(t)\sigma(t-1) \tag{5.39}$$

with

$$0 < \varkappa(t) < 2 \tag{5.40}$$

Let $\lambda_i(t)$, i = 1..n, be the eigenvalues of S(t). Then the function

$$W(t) = \sum_{i=1}^{n} \lambda_i(t)^2 \tag{5.41}$$

is decreasing.

Proof: By using Equation (5.38), the function W(t) can be written as

$$\begin{aligned} W(t) = \mathrm{tr}\, S(t)^2 &= \mathrm{tr}\, \left[S(t-1)^2 - 2k(t)x(t)x(t)^T S(t-1) + \right. \\ &\left. + k(t)^2 x(t)x(t)^T \right] = W(t-1) + \\ &+ \mathrm{tr}\, \left[- 2k(t)x(t)^T S(t-1)x(t) + k(t)^2 x(t)^T x(t) \right] \end{aligned} \tag{5.42}$$

Substituting k(t) using Equation (5.39) gives

$$\begin{aligned} W(t) &= W(t-1) - 2\varkappa(t)\sigma(t-1)^2 + \varkappa(t)^2\sigma(t-1)^2 = \\ &= W(t-1) - \left[1 - (1 - \varkappa(t))^2\right]\sigma(t-1)^2 \end{aligned} \tag{5.43}$$

From (5.40) it is finally concluded that

$$W(t) \leq W(t-1) \tag{5.44}$$

with equality if and only if $\sigma(t-1) = 0$.

□

If there were no constraints on $\alpha(t)$, the quantity $\delta_d(t)$ would be the gain in the updating equation of the P-matrix. See Equation (5.32). The constraints on $\alpha(t)$ given by Relation (5.27) induce corresponding constraints on $\delta(t)$. These are given by the Equation (5.34). The relation between the constrained gain and the desired gain can be written

$$\delta(t) = \varkappa(t) \cdot \delta_d(t) \tag{5.45}$$

where $\varkappa(t)$ will be referred to as the gain reduction factor. Bounds on the gain reduction factor $\varkappa(t)$ will no be given.

Lemma 5.2: If either Relation (5.35) holds or the desired parameter variance is smaller than

$$a < \frac{\varphi^T P^3 \varphi}{\varphi^T P^2 \varphi} - \frac{\varphi^T P^2 \varphi}{2\varphi^T P \varphi} \left[1 - \frac{v}{\varphi^T P \varphi} \right] \tag{5.46}$$

then the gain reduction factor $\varkappa(t)$ is bounded to

$$0 < \varkappa(t) < 2 \tag{5.47}$$

Proof: When Relation (5.35) holds, the lemma is easily verified from Equation (5.34). If Relation (5.35) does not hold, the lower bound in Equation (5.34) may however cause troubles. In this case, $\varkappa(t)$ satisfies Relation (5.47) if and only if $\delta_d(t)$ is greater than

$$\delta_d(t) > \frac{1}{2\varphi(t)^T P(t-1)\varphi(t)} \left[1 - \frac{v(t)}{\varphi(t)^T P(t-1)\varphi(t)} \right] \tag{5.48}$$

Substituting $\delta_d(t)$ by the expression in Equation (5.30) in Relation (5.48) yields the bound on a given by Relation (5.46).

□

Note that the restriction on the parameter a given by Relation (5.46) is imposed by Theorem 5.2. If the weaker restriction on $\alpha(t)$ given by Theorem 5.1 is used, $\varkappa(t)$ fulfils Relation (5.47) for all values of a.

The main result can now be stated.

Theorem 5.3: Let $\lambda_i(t)$, $i = 1..n$, be the eigenvalues of P(t). Assume that the P-matrix is updated according to Equation (5.4) with $\alpha(t)$ chosen as in Equation (5.33), and that a is chosen according to the assumptions in Lemma 5.2. Then the function

$$W(t) = \sum_{i=1}^{n} \left[\lambda_i(t) - a\right]^2 , \tag{5.49}$$

is decreasing.

Proof: According to Equation (5.32), the formula for updating the P-matrix can be written as

$$P(t) = P(t-1) - k(t)x(t)x(t)^T \tag{5.50}$$

where

$$x(t) = \frac{P(t-1)\varphi(t)}{\|P(t-1)\varphi(t)\|} \tag{5.51}$$

and

$$k(t) \triangleq \varkappa(t)\cdot\sigma(t) \tag{5.52}$$

The function $\varkappa(t)$ is defined in Equation (5.45), and $\sigma(t)$ is given by

$$\sigma(t) = \varphi(t)^T P(t-1)^2 \varphi(t)\cdot\delta_d(t) \tag{5.53}$$

The variable $\sigma(t)$ corresponds to the one defined in Lemma 5.1. Since $\varkappa(t)$ satisfies $0 < \varkappa(t) < 2$ according to Lemma 5.2, Lemma 5.2 can now be applied, with S(t) substituted by $P(t) - a\cdot I$.

□

It follows from the proof of Lemma 5.1 that the convergence point will not be reached only if either $\varkappa(t) \to 0$ or 2, or $\sigma(t) \to 0$. There are two cases when $\varkappa(t) \to 0$. From Equations (5.45) and the definitions of $\delta(t)$ and $\delta_d(t)$, or from Figure 5.1, it is seen that $\varkappa(t) \to 0$ if

$$\alpha_d(t) \to v(t)^{-1} + \frac{1}{\varphi(t)^T P(t-1)\varphi(t)} \tag{5.54}$$

This will only occur if $\delta(t) \to \infty$, i.e. if an eigenvalue of the P-matrix goes to infinity, which contradicts Theorem 5.3. The gain reduction factor $\varkappa(t)$ may also approach zero if

$$\frac{\varphi(t)^T P(t-1)\varphi(t)}{v(t)} \to 0 \tag{5.55}$$

or in other words, if the incoming information diminishes.

The gain reduction factor $\varkappa(t)$ is always less than or equal to one as long as Relation (5.35) holds. It approaches 2 only when Relation (5.35) does not hold and the parameter a converges to the bound given by Relation (5.46). Thus, keeping the parameter a away from this bound guarantees that $\varkappa(t)$ never converges to 2.

The other situation, i.e. $\sigma(t) \to 0$ before the P-matrix has converged to $a \cdot I$, may only occur if the vector sequence $\{P(t-1)\varphi(t)\}$ is asymptotically bounded to a certain subspace. Since the P-matrix only is changed in the $P(t-1)\varphi(t)$ directions, this means that the algorithm is unable to make the P-matrix converge to the desired value in directions outside the subspace.

Since a symmetric matrix with equal eigenvalues is diagonal, it is therefore shown that the P-matrix will converge to the diagonal matrix $a \cdot I$, subject to certain excitation conditions. The new discounting principle stated in Section 5.1 has therefore been converted to an algorithm, defined by Equations (5.4), (5.12) and (5.33) or by Equations (5.12), (5.32) and (5.34), which has been proved to meet the desired properties.

It should finally be remarked, that the above equations to control the P-matrix to $a \cdot I$ can give analogous equations to control P^{-1} to $a^{-1} \cdot I$. These equations are conceptually simpler, but involves the unattractive matrix inversion.

5.7 U-D factorization equations

The classical equations describing the least squares method, or modified versions of it, are conceptually appealing. The interpretation of the state of the algorithm as a combination of the estimates and the covariance matrix of the estimates has been of great importance in giving insight into the algorithm. These equations are often called the Kalman equations because of their similarity to the Kalman filtering.

The Kalman equations may unfortunately have bad numerical properties. The main reason is, that the formula for updating the covariance matrix may contain differences between two almost equal terms. The computer roundoff may then deteriorate the estimation. It may even lead to a computed covariance matrix which is not positive definite.

Problems occur most frequently when the P-matrix becomes ill-conditioned. Relating to what is said previously, this situation may arise after an initialization of the P-matrix, if the excitation is not uniform. In other LS algorithms, the P-matrix may also become ill-conditioned after the initialization period if the new information is not uniform for a while. This may not happen with the new algorithm. For numerical reasons, it is thus advantageous to keep the P-matrix close to a diagonal matrix with equal diagonal elements. The effort to make the covariance matrix converge to $a \cdot I$ is therefore also appealing numerically. A different covariance matrix can of course still be obtained by scaling or transforming the original θ and φ vectors.

So called square root algorithms have been important to overcome the numerical problems described above. See Bierman (1977). Instead of updating the P-matrix, a matrix S(t) satisfying

$$P(t) = S(t)S(t)^T \tag{5.56}$$

can e.g. be updated. A rule of thumb is, that these algorithms can use about half the word length required by the Kalman equations.

The use of the true square roots in Equation (5.56) involves square root computations, which are often time consuming compared with other arithmetic operations. The so called U-D factorization of the P-matrix is therefore often used instead. See Bierman (1977). In the U-D factorization method, the P-matrix is factored as

$$P(t) = U(t)D(t)U(t)^T \tag{5.57}$$

where U(t) is a unit upper triangular matrix, i.e. an upper triangular matrix with a unity diagonal, and D(t) is a diagonal matrix. Using the factorization in Equation (5.57), square root computations can be avoided.

Derivation of the U-D factorization formulas

The equations describing the U-D factorization of the new least squares algorithm will now be derived. The equations are analogous to those in Bierman (1977). The modified Kalman equations are first reformulated from Equations (5.4) and (5.12).

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{k_n}{\eta_n}\varepsilon(t) \tag{5.58a}$$

$$k_n = P(t-1)\varphi(t) \tag{5.58b}$$

$$\eta_n = v(t) + \varphi(t)^T P(t-1)\varphi(t)\varrho \tag{5.58c}$$

$$\varrho = 1 - \alpha(t)v(t) \tag{5.58d}$$

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{\eta_n}\varrho \tag{5.58e}$$

The time arguments of the introduced variables are omitted to simplify the

$$P(t) = S(t)S(t)^T \tag{5.56}$$

can e.g. be updated. A rule of thumb is, that these algorithms can use about half the word length required by the Kalman equations.

The use of the true square roots in Equation (5.56) involves square root computations, which are often time consuming compared with other arithmetic operations. The so called U-D factorization of the P-matrix is therefore often used instead. See Bierman (1977). In the U-D factorization method, the P-matrix is factored as

$$P(t) = U(t)D(t)U(t)^T \tag{5.57}$$

where U(t) is a unit upper triangular matrix, i.e. an upper triangular matrix with a unity diagonal, and D(t) is a diagonal matrix. Using the factorization in Equation (5.57), square root computations can be avoided.

Derivation of the U-D factorization formulas

The equations describing the U-D factorization of the new least squares algorithm will now be derived. The equations are analogous to those in Bierman (1977). The modified Kalman equations are first reformulated from Equations (5.4) and (5.12).

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{k_n}{\eta_n}\varepsilon(t) \tag{5.58a}$$

$$k_n = P(t-1)\varphi(t) \tag{5.58b}$$

$$\eta_n = v(t) + \varphi(t)^T P(t-1)\varphi(t)\varrho \tag{5.58c}$$

$$\varrho = 1 - \alpha(t)v(t) \tag{5.58d}$$

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{\eta_n}\varrho \tag{5.58e}$$

The time arguments of the introduced variables are omitted to simplify the

notations. The use of the subscript n is motivated later.

As mentioned above, Equation (5.58e) will be replaced by equations describing the updating of the matrices U and D defined in Equation (5.57). The derivation will be based on the following technical lemma.

Lemma 5.3: Given a unit upper triangular matrix U and an arbitrary matrix X of the same dimensions. Let V denote the matrix

$$V = UXU^T \tag{5.59}$$

Then a lower right submatrix of V is equal to the product of the corresponding lower right submatrices of U, X and U^T respectively.

□

Proof: The lemma is easily verified by dividing the matrices U and X into submatrices and evaluating the product.

□

Let $d(t)_i$ denote the i:th diagonal element of D(t) and $u(t)_i$ the i:th column of U(t). The updating formulas are then given by the following theorem.

Theorem 5.4: The U-D factorization of Equation (5.58) is given by the following algorithm.

$$f = U(t-1)^T \varphi(t) \tag{5.60a}$$

$$g = D(t-1)f \tag{5.60b}$$

$$k_0 = (0 \ \ldots \ 0)^T \tag{5.60c}$$

$$k_i = \sum_{j=1}^{i} u(t-1)_j g_j \qquad i = 1,..,n \tag{5.60d}$$

$$\varrho = 1 - \alpha(t)v(t) \tag{5.60e}$$

$$\eta_0 = v(t) \tag{5.60f}$$

$$\eta_i = v(t) + \varrho \sum_{j=1}^{i} f_j g_j \qquad i = 1,..,n \tag{5.60g}$$

For i = 1,..,n, go through the Equations (5.60h) - (5.60j).

$$d(t)_i = d(t-1)_i \frac{\eta_{i-1}}{\eta_i} \tag{5.60h}$$

$$\lambda_i = -f_i \varrho / \eta_{i-1} \tag{5.60i}$$

$$u(t)_i = u(t-1)_i + \lambda_i k_{i-1} \tag{5.60j}$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{k_n}{\eta_n} \varepsilon(t) \tag{5.60k}$$

Here n is equal to the dimension of the U- and D-matrices, and f_i and g_i are the i:th elements of f and g respectively.

□

Proof: For simplicity, the equation numbers (5.60ν), ν = a, .. ,k, are abbreviated to [ν] in the following proof.

Factorizing the P-matrices in Equation (5.58e) and using the definitions of f and g gives

$$U(t)D(t)U(t)^T = U(t-1)D(t-1)\left[I - \frac{\varrho}{\eta_n} fg^T \right]U(t-1)^T \tag{5.61}$$

From Lemma 5.3, the following set of equations are obtained by taking the determinants of lower right submatrices in Equation (5.61), and using the fact that the determinant of a unit triangular matrix is one.

$$\prod_{i=\ell}^{n} d(t)_i = \left[\prod_{i=\ell}^{n} d(t-1)_i \right]\left[1 - \frac{\varrho}{\eta_n} \sum_{i=\ell}^{n} f_i g_i \right] \qquad \ell = 1,..,n \tag{5.62}$$

It follows from [g] that

$$1 - \frac{\varrho}{\eta_n} \sum_{i=\ell}^{n} f_i g_i = \frac{\eta_{\ell-1}}{\eta_n} \qquad (5.63)$$

Hence

$$d(t)_n = d(t-1)_n \left[1 - \frac{\varrho}{\eta_n} f_n g_n \right] = d(t-1)_n \frac{\eta_{n-1}}{\eta_n} \qquad (5.64)$$

By recursion, [h] is now easily shown, and the updating formula for the D-matrix is thus demonstrated.

It remains to show that the updating of the U-matrix satisfies the Kalman equations. This is done by evaluating the left hand side of Equation (5.61) using the Equations (5.60), and showing that Equation (5.61) is correct.

The left hand side of Equation (5.61) can be written as

$$U(t)D(t)U(t)^T = \sum_{i=1}^{n} u(t)_i d(t)_i u(t)_i^T \qquad (5.65)$$

Using [h] and [j] to write Equation (5.65) in terms of variables known at time t-1 gives

$$U(t)D(t)U(t)^T =$$

$$= \sum_{i=1}^{n} \left[u(t-1)_i + \lambda_i k_{i-1}\right] d(t-1)_i \frac{\eta_{i-1}}{\eta_i} \left[u(t-1)_i + \lambda_i k_{i-1}\right]^T \qquad (5.66)$$

Parts of the right hand side can be evaluated as

$$\frac{\eta_{i-1}}{\eta_i} \left[u(t-1)_i + \lambda_i k_{i-1}\right] = u(t-1)_i + \frac{\eta_{i-1} - \eta_i}{\eta_i} u(t-1)_i +$$

$$+ \frac{\eta_{i-1}}{\eta_i} \lambda_i k_{i-1} = u(t-1)_i - \frac{\varrho f_i g_i}{\eta_i} u(t-1)_i - \frac{\varrho f_i}{\eta_i} k_{i-1} =$$

$$= u(t-1)_i + \frac{\varrho f_i}{\eta_i}\left(k_{i-1} - k_i\right) - \frac{\varrho f_i}{\eta_i} k_{i-1} =$$

$$= u(t-1)_i - \frac{\varrho f_i}{\eta_i} k_i \qquad (5.67)$$

where [g] and [i] are used in the second equality and [d] is used in the third equality. Using Equation (5.67) in Equation (5.66) now gives

$$U(t)D(t)U(t)^T = \sum_{i=1}^{n} \left[u(t-1)_i + \lambda_i k_{i-1}\right] d(t-1)_i \cdot$$

$$\cdot \left[u(t-1)_i - \frac{\varrho f_i}{\eta_i} k_i\right]^T = U(t-1)D(t-1)U(t-1)^T +$$

$$+ \sum_{i=1}^{n} d(t-1)_i \left[\lambda_i k_{i-1} u(t-1)_i^T - \frac{\varrho f_i}{\eta_i} u(t-1)_i k_i^T -\right.$$

$$\left. - \frac{\varrho f_i}{\eta_i} \lambda_i k_{i-1} k_i^T\right] \qquad (5.68)$$

The terms in the sum above can be simplified as follows

$$d(t-1)_i \left[\lambda_i k_{i-1} u(t-1)_i^T - \frac{\varrho f_i}{\eta_i} u(t-1)_i k_i^T - \frac{\varrho f_i}{\eta_i} \lambda_i k_{i-1} k_i^T\right] =$$

$$= - \frac{\varrho g_i}{\eta_{i-1}} k_{i-1} u(t-1)_i^T - \frac{\varrho g_i}{\eta_i} u(t-1)_i k_i^T + \frac{\varrho g_i f_i \varrho}{\eta_i \eta_{i-1}} k_{i-1} k_i^T =$$

$$= \varrho \left[\frac{k_{i-1}}{\eta_{i-1}}\left(k_{i-1} - k_i\right)^T + \left(k_{i-1} - k_i\right)\frac{k_i^T}{\eta_i} +\right.$$

$$\left. + \left(\frac{1}{\eta_{i-1}} - \frac{1}{\eta_i}\right) k_{i-1} k_i^T\right] = \varrho \left[- \frac{k_i k_i^T}{\eta_i} + \frac{k_{i-1} k_{i-1}^T}{\eta_{i-1}}\right] \qquad (5.69)$$

where [b] and [i] are used in the fist equality and [d] and [g] are used in

the second equality. Using Equation (5.69) in Equation (5.68) now finally gives

$$U(t)D(t)U(t)^T = U(t-1)D(t-1)U(t-1)^T +$$

$$+ \varrho \sum_{i=1}^{n} \left[- \frac{k_i k_i^T}{\eta_i} + \frac{k_{i-1} k_{i-1}^T}{\eta_{i-1}} \right] =$$

$$= U(t-1)D(t-1)U(t-1)^T - \frac{\varrho}{\eta_n} k_n k_n^T =$$

$$= U(t-1)D(t-1)\left[I - \frac{\varrho}{\eta_n} f g^T\right]U(t-1)^T \qquad (5.70)$$

which is the right hand side of Equation (5.61). Hence it is shown that Equation (5.60) correctly describes the U-D factorization of Equation (5.58).

□

The U-D factorization equations given above include the variable $\alpha(t)$ explicitly, in Equation (5.60e). In a practical implementation, it may instead be better to introduce the equations involving the variable $\delta(t)$, and to use the bounds on $\delta(t)$ given by Equation (5.34). The corresponding U-D factorization equations are easily obtained by simply expressing ϱ in terms of $\delta(t)$ instead of $\alpha(t)$. Equation (5.60e) is then replaced by

$$\varrho = \frac{\delta(t)v(t)}{1 - \delta(t)\varphi(t)^T P(t-1)\varphi(t)} \qquad (5.60e')$$

Example

To compare the numerical properties of the Kalman equations and the U-D factorization equations, the two computational methods have been simulated using the program package MATLAB, see Moler (1981). These examples may also illustrate the properties of the new recursive estimation algorithm, in a difficult case.

It follows from the discussion at the beginning of this section, that numerical problems are more likely if the information which enters the estimator is nonuniformly distributed. In the present example, a fifth order AR process driven by white noise has been simulated. The AR process has all the five poles at z = 0.3. The compared estimation algorithms are given by Equations (5.58) and (5.60) respectively. The variable $\alpha(t)$ is chosen according to Equation (5.33). For simplicity, only the eigenvalues of the P-matrices are presented in the following examples.

Figures 5.2 and 5.3 show the five eigenvalues of the P-matrices for the two algorithms. The P-matrices are initialized to $500 \cdot I$. MATLAB uses double precision arithmetic, so no differences between the two methods can be seen in Figures 5.2 and 5.3. The slow convergence is a direct consequence of the poor excitation. In Figure 5.4, a longer simulation is presented, which better demonstrates the convergence properties of the algorithm. The desired P-matrix is $0.005 \cdot I$.

Figures 5.5 and 5.6 show results of the same computations in single precision arithmetic. The U-D factorization method performs almost like the double precision version. The Kalman equations are however deteriorated by the computer roundoff. The algorithm even fails to give a positive definite P-matrix.

An interesting property of the new least squares algorithm can be seen in Figure 5.6, namely that a computed nonpositive covariance matrix do not have to imply that the estimator fails to work. The original least squares algorithm and the modified version with a constant forgetting factor all become unstable if the P-matrix becomes nonpositive. In the example presented above, the algorithm manages to make the covariance matrix positive definite again after three periods of negative eigenvalues.

The example demonstrates, that though the Kalman equations and the U-D factorization method are algebraically equivalent, their behaviour may be totally different in a computer implementation. The U-D factorization equations are therefore recommended for practical implementations.

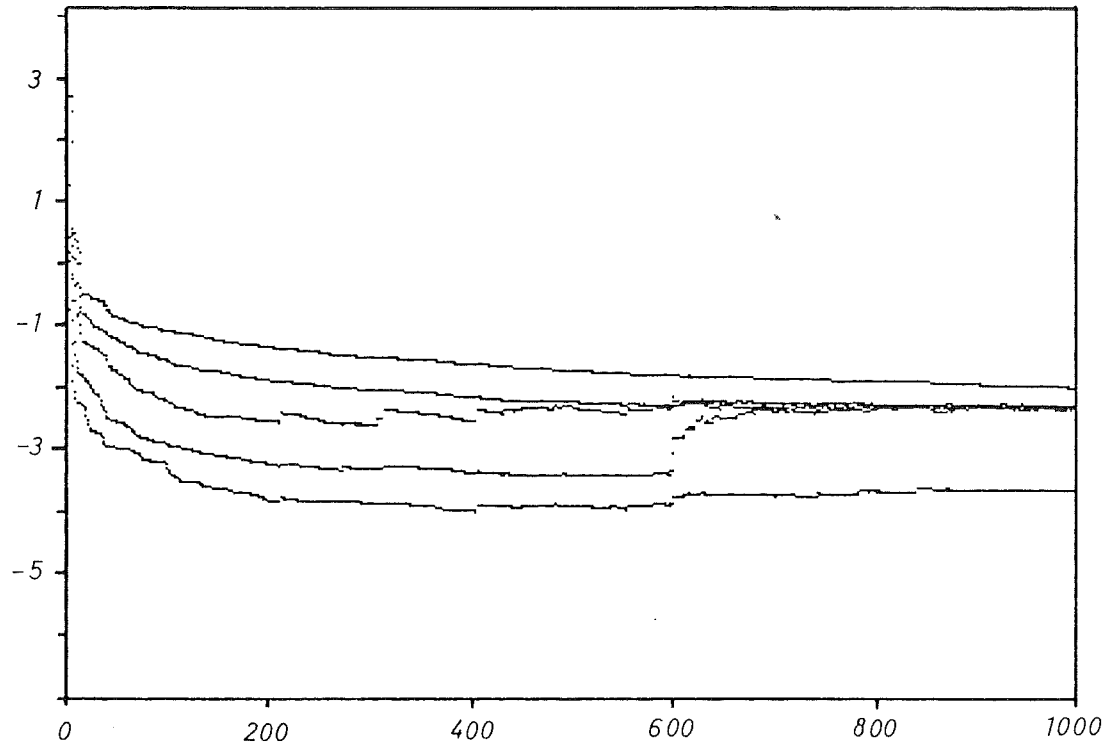


Figure 5.2 - The logarithms of the eigenvalues of the P-matrix when the U-D factorization equations are used. The computations are performed in double precision arithmetic.

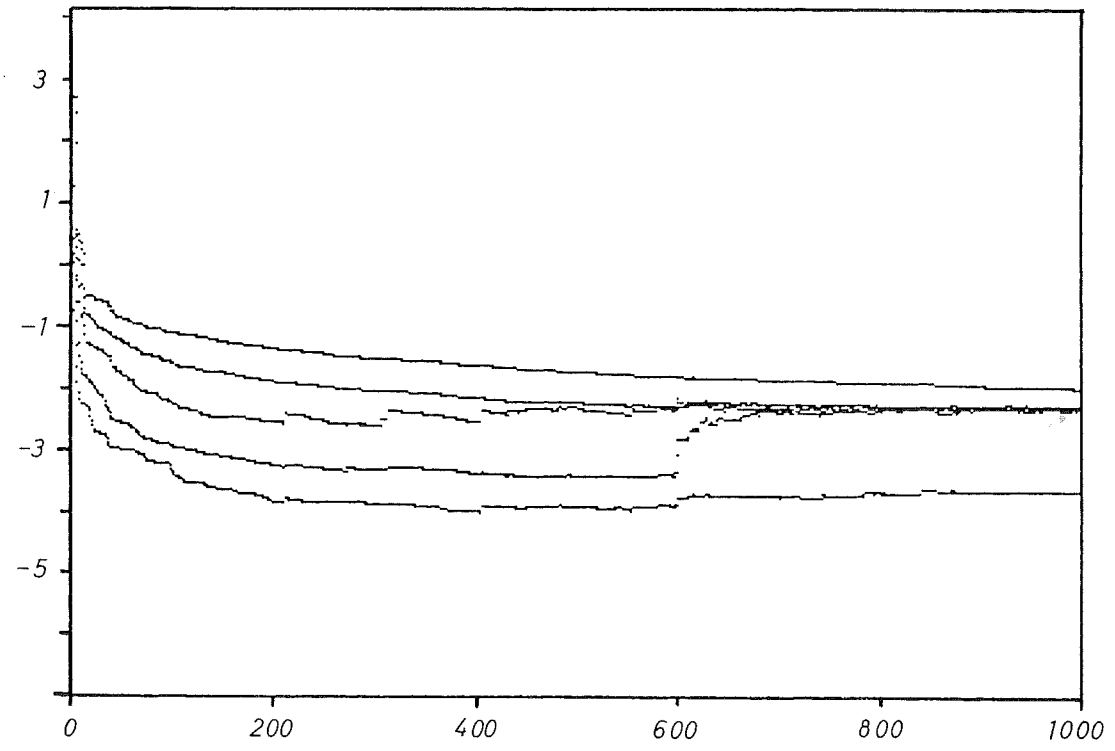


Figure 5.3 - The logarithms of the eigenvalues of the P-matrix when the Kalman equations are used. The computations are performed in double precision arithmetic.

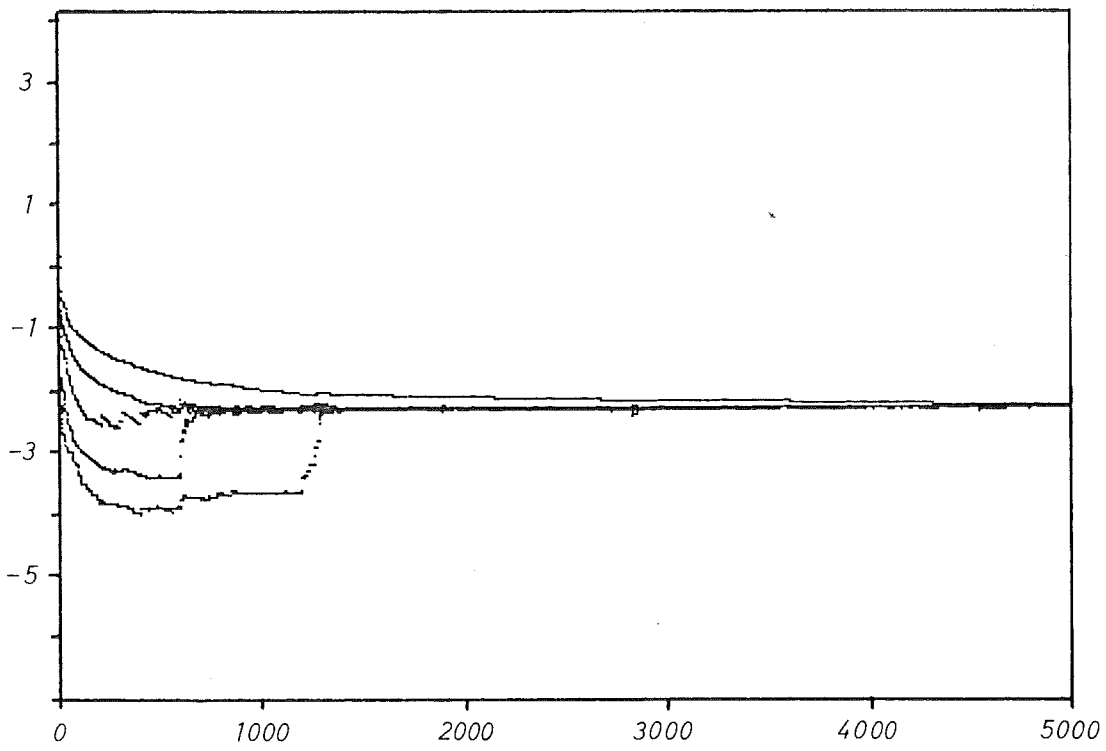


Figure 5.4 - The logarithms of the eigenvalues of the P-matrix when the Kalman equations are used. The computations are performed in double precision arithmetic.

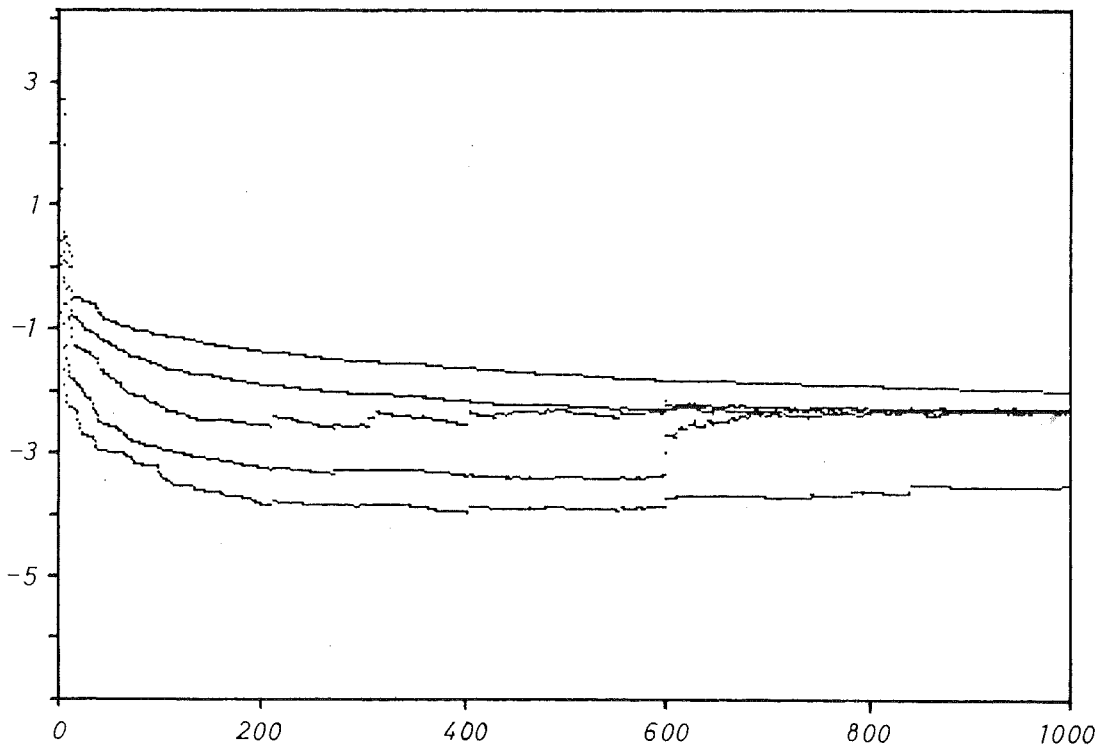


Figure 5.5 - The logarithms of the eigenvalues of the P-matrix when the U-D factorization equations are used. The computations are performed in single precision arithmetic.

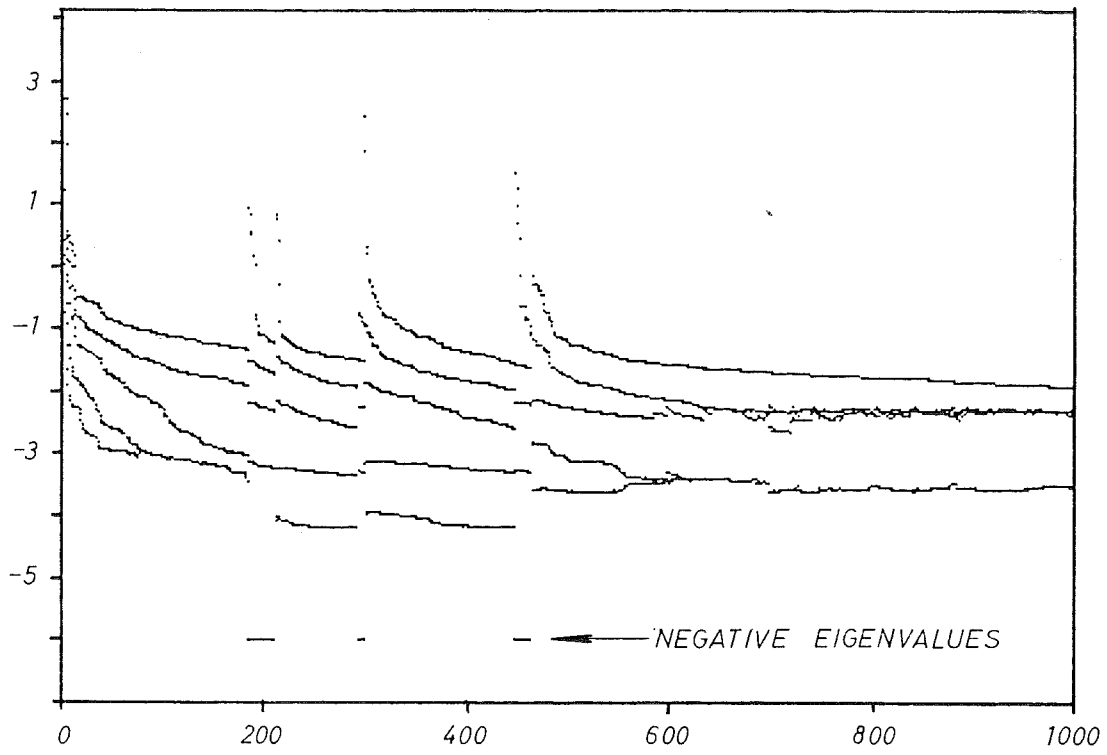


Figure 5.6 - The logarithms of the eigenvalues of the P-matrix when the Kalman equations are used. The computations are performed in single precision arithmetic.

6. EXAMPLE - CONTROL OF AN INDUSTRIAL ROBOT

The properties of the estimation procedure derived in the Chapters 4 and 5 will now be illustrated by an example. The simulations are performed using the simulation package SIMNON, see Elmqvist (1975). A crude model of an industrial robot will be used as a process model.

The complete estimation procedure is first summarized in Section 6.1. The dynamics of the robot is presented in Section 6.2 and the simulation results are finally given in Section 6.3.

6.1 The complete estimation algorithm

The estimation algorithm has been given previously. The complete algorithm is however summarized below for the sake of convenience. There are some remaining comments to make.

Throughout the thesis, the estimate of the noise variance v(t) has been included in the algorithm in order to have the notions of "covariance matrix" and "information matrix" make sense. Low pass filtering of the squares of the residuals $\varepsilon(t)$ is a simple and convenient way of estimating $\sigma_n(t)^2$. From the discussion in Chapter 3, it is easily seen that this method will give a poor estimate when there are large parameter changes. It is therefore wise to avoid estimation of the noise variance when there are large parameter changes. This is conveniently done by considering the variable r(t), see Equation (4.16). Since r(t) is a measure of the probability of a recent parameter change, the noise estimation is to be performed only when r(t) is smaller than a certain threshold r_1. To compensate for the time delay in the detection procedure, it is also useful to introduce a time delay τ in the noise variance estimator. The parameter values $r_1 = 0.1$ and $\tau = 15$ are used in this example. The noise variance estimation is given by Equation (6.1ℓ) below.

The new fault detection method in Chapter 4 was derived for the estimation scheme based on a constant forgetting factor. The determination of the parameter $\beta(t)$ given by Equations (4.33) - (4.36) is easily reformulated for the new method given in Chapter 5. The parameter $\beta(t)$ is given by Equations (6.1j) and (6.1k), where $v(t)$ is chosen according to Figure 4.5.

The complete estimation algorithm is given by the following equations.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{v(t)} P(t)\varphi(t)\varepsilon(t) \tag{6.1a}$$

$$\varepsilon(t) = y(t) - \varphi(t)^T\hat{\theta}(t-1) \tag{6.1b}$$

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{\left[v(t)^{-1}-\alpha(t)\right]^{-1} + \varphi(t)^T P(t-1)\varphi(t)} + \beta(t)I \tag{6.1c}$$

$$\alpha = \begin{cases} 0 & \alpha_d \leq 0 \\ \alpha_d & 0 < \alpha_d \leq \dfrac{1}{\varphi^T P\varphi} \\ \dfrac{1}{\varphi^T P\varphi} & \dfrac{1}{\varphi^T P\varphi} < \alpha_d \leq v^{-1} + \dfrac{1}{\varphi^T P\varphi} \\ 0 & \alpha_d > v^{-1} + \dfrac{1}{\varphi^T P\varphi} \end{cases} \tag{6.1d}$$

$$\alpha_d(t) = v(t)^{-1} + \frac{\delta_d(t)}{\delta_d(t)\varphi(t)^T P(t-1)\varphi(t) - 1} \tag{6.1e}$$

$$\delta_d(t) = \frac{1}{\varphi(t)^T P(t-1)^2\varphi(t)} \left[\frac{\varphi(t)^T P(t-1)^3\varphi(t)}{\varphi(t)^T P(t-1)^2\varphi(t)} - a \right] \tag{6.1f}$$

$$w(t-1) = \gamma_1 w(t-2) + \Delta\hat{\theta}(t-1) \tag{6.1g}$$

$$s(t) = \text{sign}\left[\Delta\hat{\theta}(t)^T w(t-1)\right] \tag{6.1h}$$

$$r(t) = \gamma_2 r(t-1) + (1-\gamma_2)\varepsilon(t) \tag{6.1i}$$

$$\nu_0 = 1 - \frac{\varphi(t)^T P(t-1)\varphi(t)}{v(t) + \left[1 - \alpha(t)v(t)\right]\varphi(t)^T P(t-1)\varphi(t)} \tag{6.1j}$$

$$\beta(t) = \begin{cases} 0 & r(t-1) < r_0 \\ \dfrac{v(t)}{\varphi(t)^T\varphi(t)}\left[\nu_0(t)-\nu(t)\right] & r(t-1) \geq r_0 \end{cases} \tag{6.1k}$$

$$v(t) = \begin{cases} \gamma_3 v(t-1) + (1-\gamma_3)\varepsilon(t-\tau)^2 & r(t-1) < r_1 \\ v(t-1) & r(t-1) \geq r_1 \end{cases} \tag{6.1ℓ}$$

6.2 The robot model

The problem of controlling an industrial robot is an advanced servo problem, which has received considerable attention in recent years. The robot arm is positioned by servos in the different joints. In this example, only circular motions around the vertical axis will be considered. See Figure 6.1. Assuming that the distance between the vertical axis and the tip of the robot arm is constant, and that the moment of inertia J is constant, the torque balance for the vertical axis becomes

$$J\frac{d\omega(t)}{dt} = T_e(t) + T_f(t) + T_d(t) \tag{6.2}$$

where $T_e(t)$ is the torque generated by the motor, $T_f(t)$ a friction torque, $T_d(t)$ a disturbance torque and $\omega(t)$ the angular velocity. The sampled version of Equation (6.2) becomes

$$\omega(t+1) = \omega(t) + \frac{1}{J}\left[\, T_e(t) + T_f(t) + T_d(t) \,\right] \tag{6.3}$$

The disturbance torque $T_d(t)$ will be modeled as a white noise disturbance sequence. The friction torque is modeled as

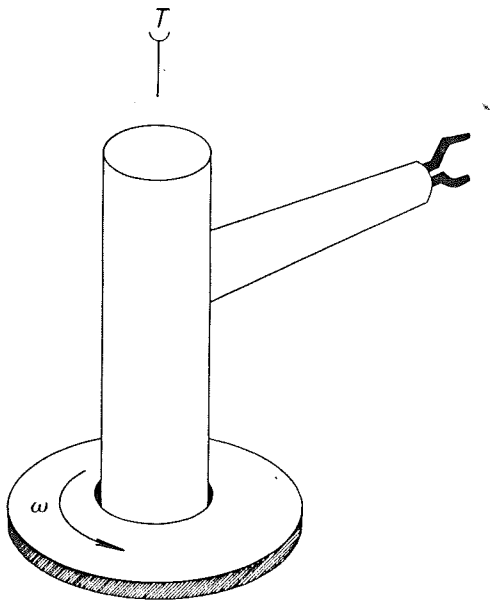


Figure 6.1 - The robot considered in the example.

$$T_f(t) = - k_f \, sign(\omega(t)) \qquad (6.4)$$

where k_f is a constant. This is a good description of friction only for $\omega(t) \neq 0$, but since k_f is small compared to $T_e(t)$ and $T_d(t)$ in the example, the Equation (6.4) is acceptable.

It is assumed that the motor has current feedback. The torque from the motor then becomes

$$T_e(t) = k_I \cdot I(t) \qquad (6.5)$$

where k_I is a constant and I(t) is a function of the angular velocity, defined by the control law. The desired control law in this example is chosen as

$$T_e(t) = 0.5 \cdot J \left[\omega_{ref}(t) - \omega(t)\right] + k_f \, sign(\omega(t)) \qquad (6.6)$$

This corresponds to a pole assignment at 0.5 and a nonlinear compensation for the friction torque. Since it is assumed that the moment of inertia J and the friction constant k_f are unknown, they are estimated using the algorithm described in Section 6.1. From Equation (6.3), the parameter vector θ and the

regression vector φ become

$$\theta(t) = \begin{bmatrix} 1/J & 2k_f/J \end{bmatrix}^T \triangleq \begin{bmatrix} \theta_1(t) & \theta_2(t) \end{bmatrix}^T \qquad (6.7)$$

and

$$\varphi(t) = \begin{bmatrix} T_e(t) & -\frac{1}{2}\,\mathrm{sign}(\omega(t)) \end{bmatrix}^T \qquad (6.8)$$

The applied control law becomes

$$T_e(t) = 0.5\,\frac{1}{\hat{\theta}_1(t)}\left[\omega_{ref}(t) - \omega(t)\right] + \frac{\hat{\theta}_2(t)}{\hat{\theta}_1(t)}\,\mathrm{sign}(\omega(t)) \qquad (6.9)$$

6.3 Simulation results

The process and the regulator defined in the previous section have been simulated together with the adaptation loop given by Equation (6.1). The filter constants were chosen to $\gamma_1 = 0.8$, $\gamma_2 = 0.95$ and $\gamma_3 = 0.95$. The example is included in order to demonstrate several features of the new algorithm. The robot is therefore supposed to perform the following motion scheme.

At the beginning, the robot is supposed to rotate back and forth with a constant velocity. The reference value ω_{ref} of the angular velocity is a square wave with unit amplitude. The variance of the disturbance torque $T_d(t)$ is 0.01, and the friction coefficient is $k_f = 0.05$.

At t = 100, the robot picks up a heavy tool. This is simulated by increasing the moment of inertia J from 1 to 2. The changed moment of inertia causes a change in the parameters by a factor of two, and simultaneously a decrease in the noise variance by a factor of four.

At t = 300, the reference signal becomes zero instead of a square wave, i.e. it is desired to keep the robot arm at rest.

At t = 400, the variance of the disturbance torque increases to 0.1. This may e.g. model the situation that the robot starts to drill.

Figures 6.2 - 6.5 show the results of the simulation. The algorithm is initialized in equilibrium. The parameter estimates, the estimate of the noise variance and the P-matrix have the expected values. The P-matrix is postulated to converge to $0.0005 \cdot I$.

When the moment of inertia changes at time t = 100, the parameter estimates slowly converge towards their new values. At $t \approx 125$, this change is detected with a confidence of more than 99.9% ($r(t) > r_0 = 0.5$). See Figure 6.4. The gain of the estimator is then increased, which shows up as an increased adaptation rate. See Figure 6.2.

When the reference signal becomes zero at t = 300, the moment of inertia J is more difficult to estimate. The time horizon for the estimation of $\theta_1(t)$ is therefore increased, leading to the very slow rate of change of the estimate.

The changes of the noise variance at t = 100 and t = 400 are also detected in the estimates shown in Figure 6.2. Notice also the periods when the estimation of the noise variance is interrupted by large values of r(t). The algorithm compensates for the variations in amount of incoming information so that the parameter variance is kept constant. The slow variations of the parameter estimates in Figure 6.2 at the end of the simulation is a consequence of the increased time horizon.

The variable $\alpha(t) \cdot v(t)$ is plotted in Figure 6.4. At the beginning $\alpha(t) = v(t)^{-1}$, which means that the same amount of information that enters the estimator is discounted. See Equation (5.5). When a parameter change is detected and the diagonal elements of the P-matrix have been increased, the variable $\alpha(t)$ is set to zero which means that all new information is retained. After a while, the P-matrix gets close to the desired value $a \cdot I$, and the value of $\alpha(t)$ varies to perform the fine adjustment. When this is done, $\alpha(t)$ becomes equal to $v(t)^{-1}$ again.

For comparison, the above simulation problem was also solved by two other estimation schemes. Figure 6.6a shows results with an LS estimator with a constant forgetting factor. The forgetting factor is chosen to give the same parameter variance as the former algorithm in the initial phase. Since no parameter changes are detected, the rate of adaptation is slower. If a constant forgetting factor is used, the parameter variance is a function of the noise variance. This is clearly seen in this simulation, where the parameter variances increase drastically when the noise variance increases at t = 400.

Figure 6.6b shows results using an algorithm where the forgetting factor is changed as

$$\lambda(t) = 1 - \frac{\varepsilon(t)^2}{N \cdot \bar{\varepsilon}(t)^2} \qquad (6.10)$$

with

$$\bar{\varepsilon}(t)^2 = 0.95 \cdot \bar{\varepsilon}(t-1)^2 + 0.05 \cdot \varepsilon(t)^2 \qquad (6.11)$$

This is essentially the method of Fortesque et al (1981) and Wellstead and Sanoff (1981) discussed in Chapters 2 and 3. The value of N is chosen to give an adaptation rate approximately as fast as the one in Figure 6.2. The method interprets all increases of $\varepsilon(t)^2$ as parameter changes. The increased noise variance at t = 400 will therefore deteriorate the estimation drastically. Note that the variances of the parameter estimates are much larger than in Figure 6.6a, where a constant forgetting factor is used.

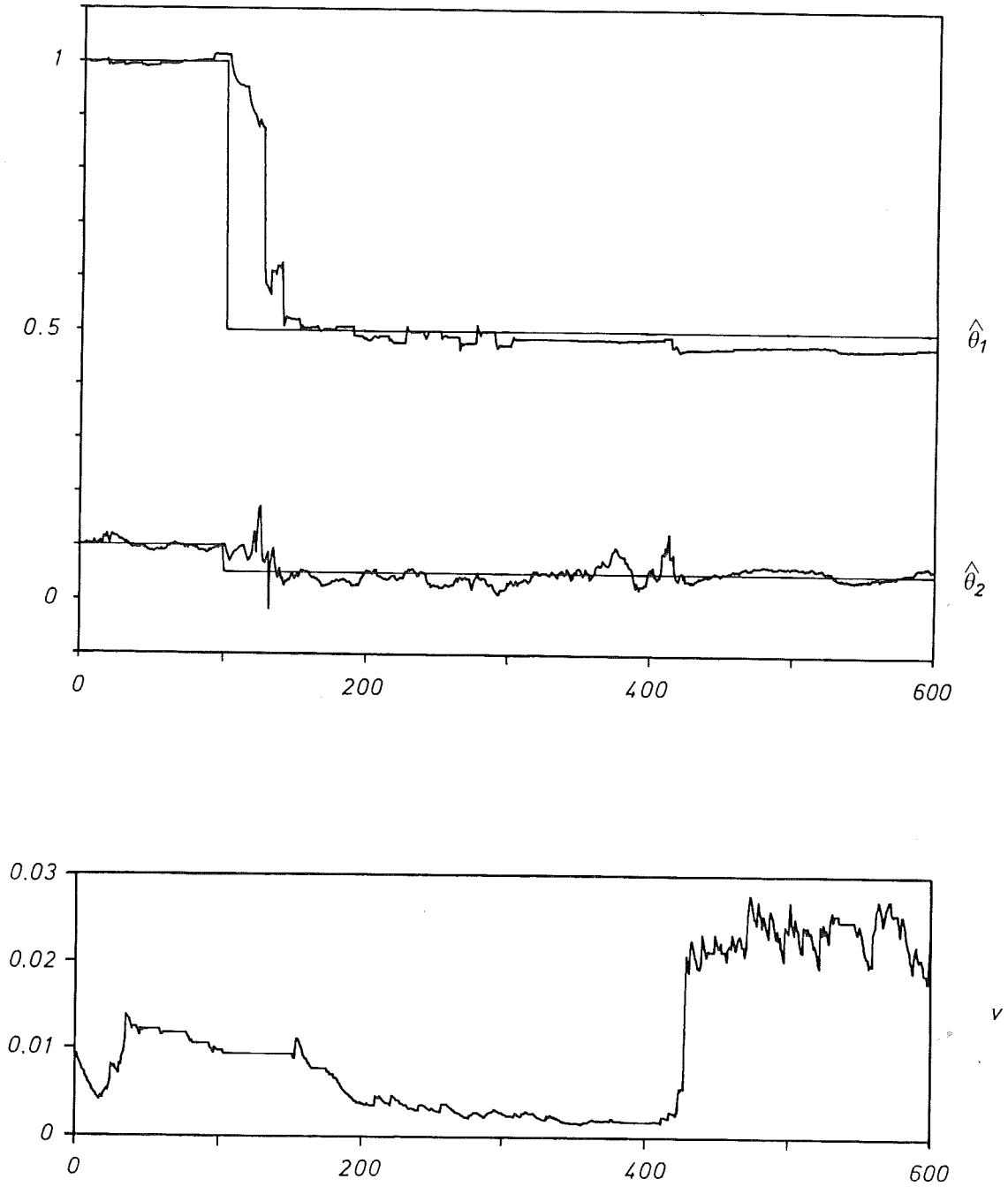


Figure 6.2 - The estimated parameters and the estimated noise variance.

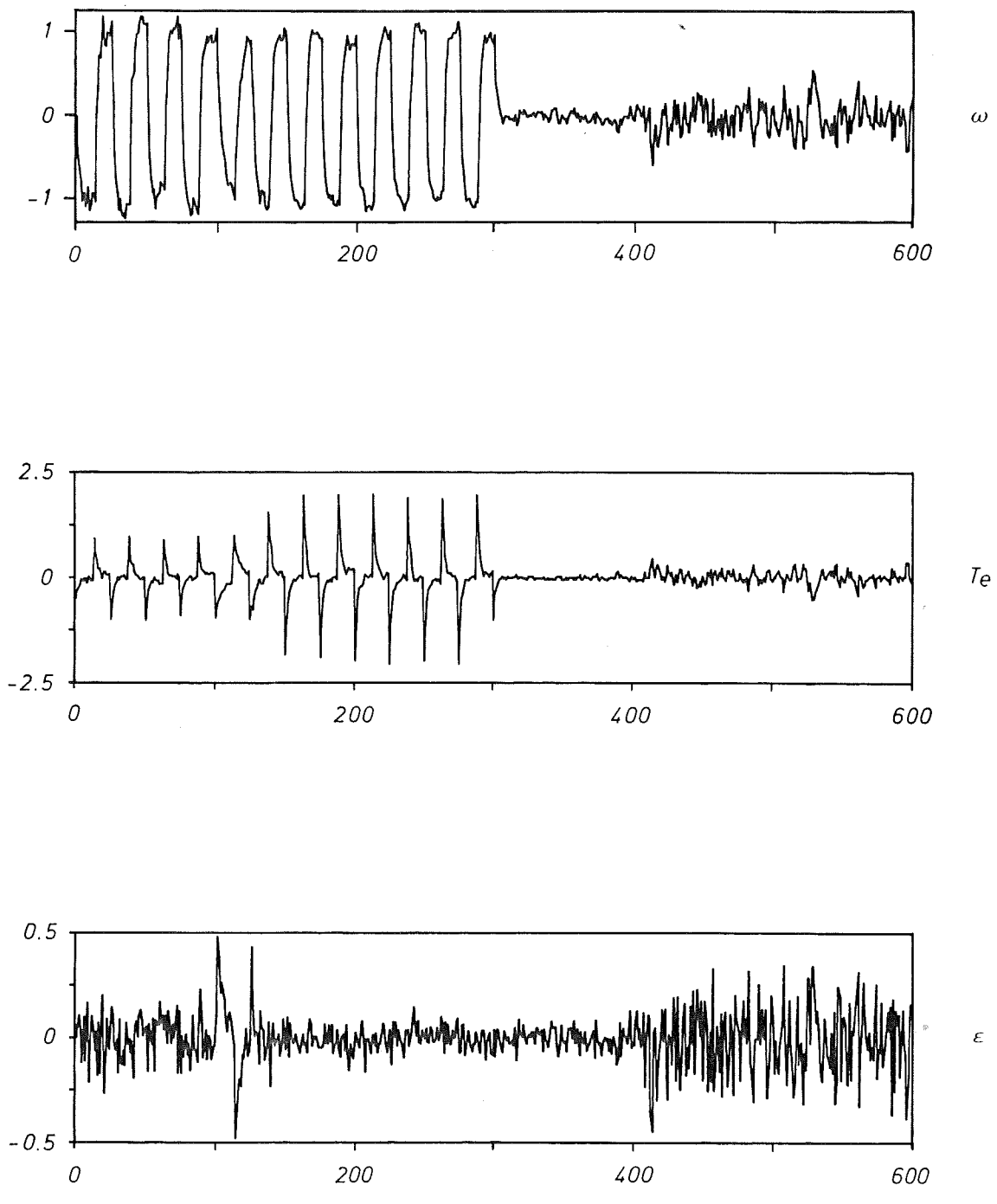


Figure 6.3 - The output- and input-signals of the system, and the residuals $\varepsilon(t)$.

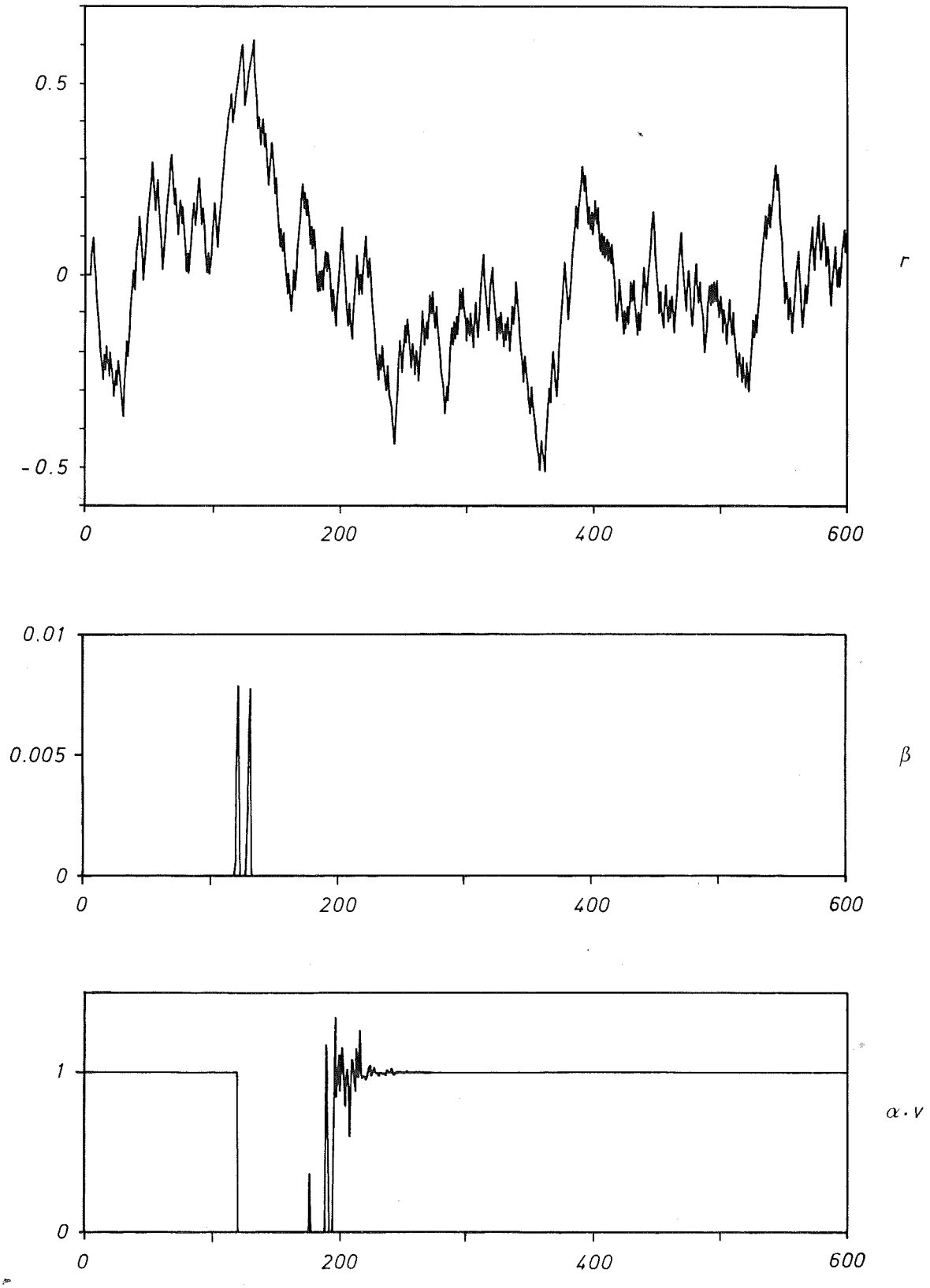


Figure 6.4 - The test signal r(t), the additive gain β(t) and the discounting measure α(t)·v(t).

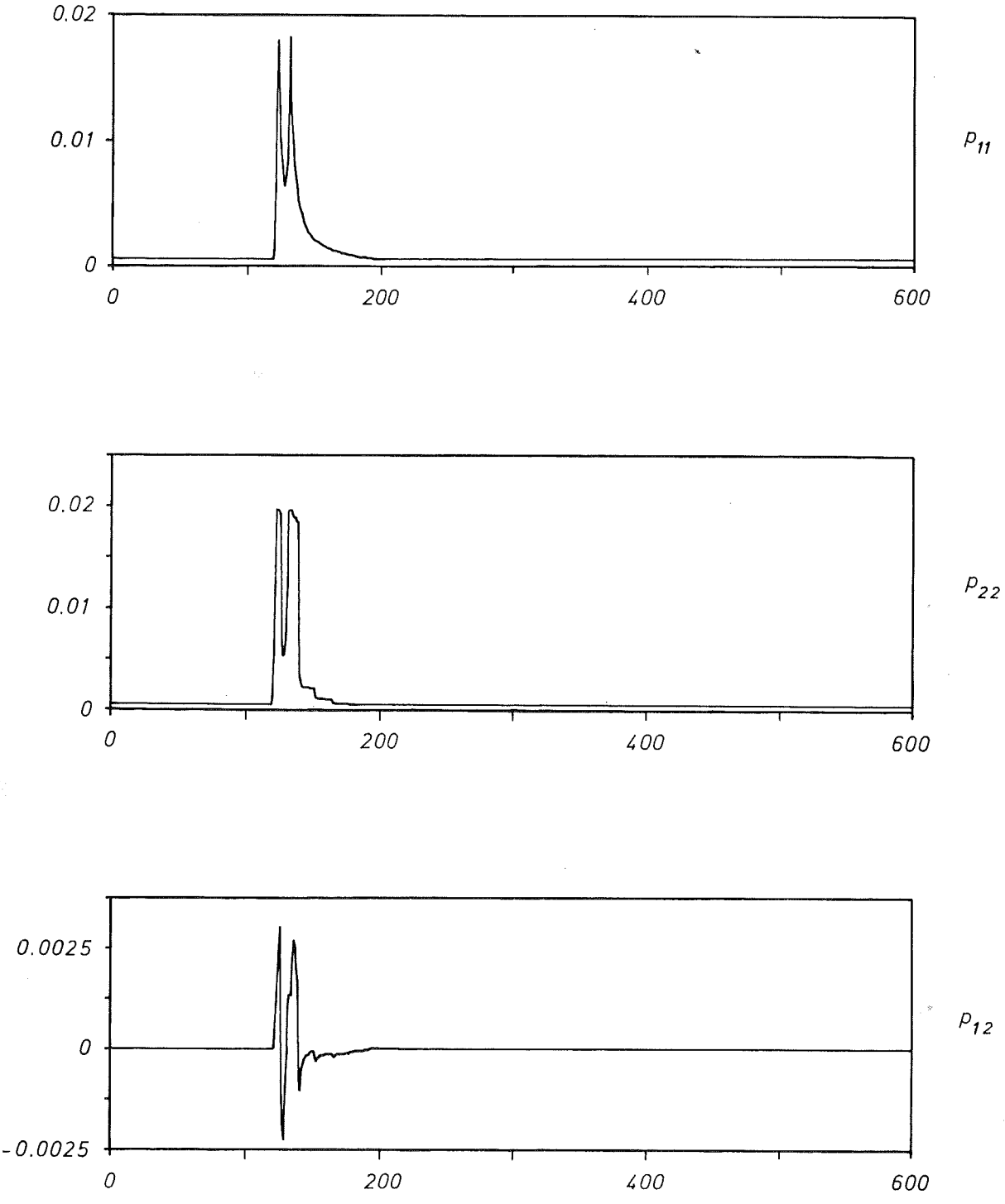


Figure 6.5 - The elements of the P-matrix.

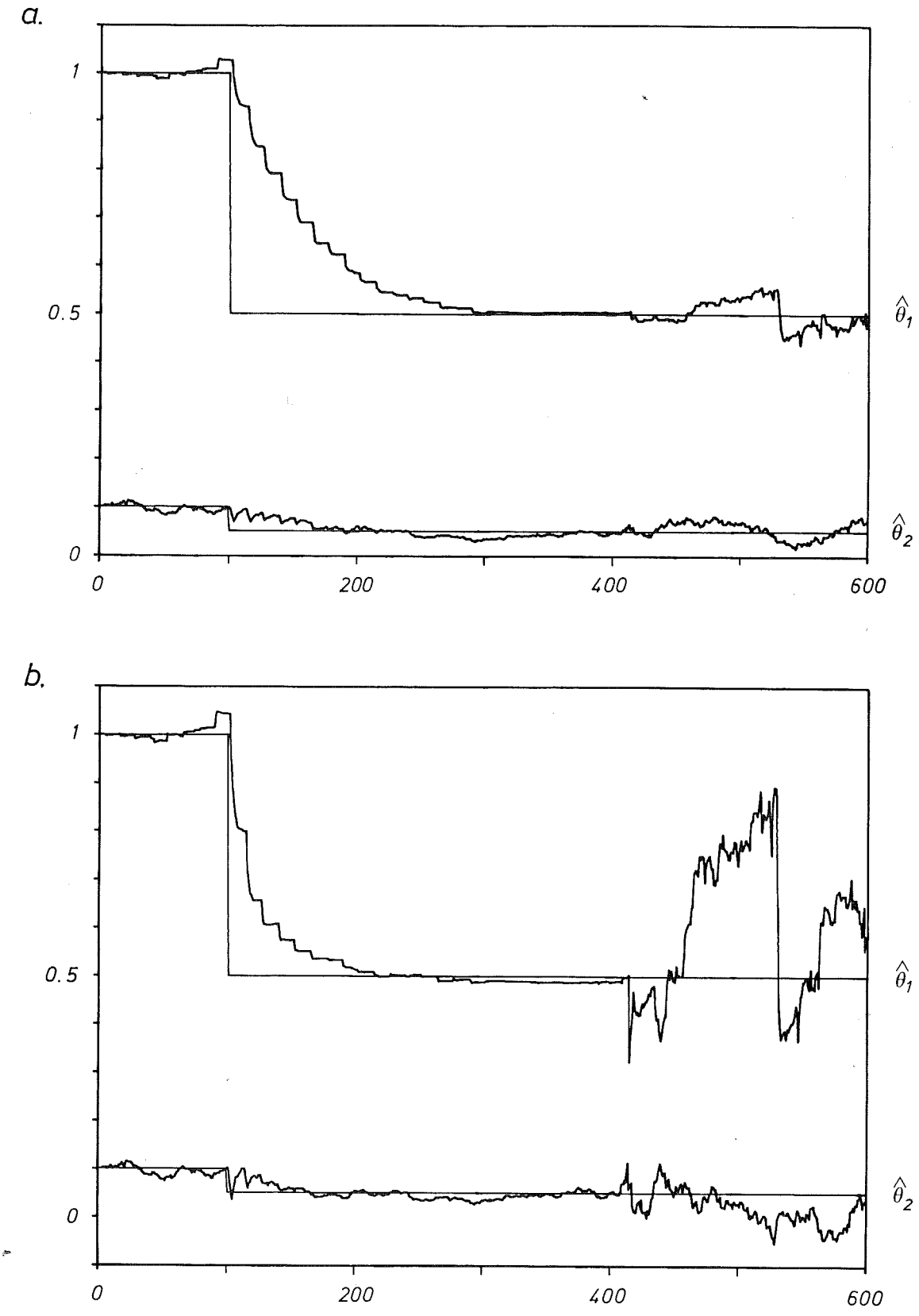


Figure 6.6 - The estimated parameters when a constant forgetting factor (a) and a time-varying forgetting factor (b) was used.

7. CONCLUSIONS

In the Introduction, the adaptive control concept was motivated for two reasons:

1. To facilitate the analysis and the design procedure of the control problem.

2. To extend the applications of the theory for linear time-invariant systems to nonlinear and/or time-variable systems.

The motivation for the thesis has been the second reason, and the main contributions are proposals for new ways to handle time-varying and slightly nonlinear processes by using adaptive control.

The thesis has focused on the estimation part of the problem, since this is where most difficulties due to the time-variability occur. Both the problem and the proposed solutions have been discussed from an information handling point of view. It was shown, in Chapter 3, that most of the difficulties could be deduced from a bad correspondence between true and assumed uncertainty of the measurements, leading to an incorrect relative weighting of the measurements. After a discussion of the basic problem and previous solutions, new solutions to the problem were given in Chapter 4 and Chapter 5.

The way to solve the weighting problem depends on the nature of the time-variations. In Assumption 3.1 of Chapter 3, fundamental assumptions about the time-variations were introduced. The class of time-variations fulfilling Assumption 3.1 was then divided into two categories.

The first one consists of those parameter changes that can be detected by a fault detection procedure. They were called large parameter changes. The way to handle these parameter changes is to detect them and increase the gain in the estimation algorithm whenever they occur. The fault detection performs a nonlinear dynamic filtering of the measurement sequence. This kind of filtering was used in order to get a method which is insensitive to the disturbances.

The second category of the time-variations were called slow parameter changes. They can not be detected easily. The way to handle them is to prevent the gain in the estimator from being to small, by giving more weight to recent information.

A new fault detection procedure was presented in Chapter 4. Methods for fault detection are of course of substantial importance in their own. It is a well-known trend, that the industrial processes are getting more and more sophisticated. The increased performances are often reached at the price of processes that are more sensitive and harder to supervise. Illuminating examples are e.g. aircrafts and power systems. This progress, together with increasing demands on availability and security, has caused an increased interest in the problem of fault detection in dynamic systems. The problem is often solved by redundant hardware combined with some voting technique. Since this is an expensive solution and since computers are getting cheaper and cheaper, software solutions have received much interest in the past few years.

Concerning the adaptive control application, four requirements on the detection method (R1) - (R4) were set up. The new fault detection method was constructed so that these requirements were satisfied. Since these requirements are natural, the new method is believed to be applicable also to other areas of fault detection.

It has been shown in examples that it is possible to detect faults by the proposed method, even if the faults do not influence the magnitude of the residuals $\varepsilon(t)$ much.

In the traditional estimators, the information is usually condensed in the form of an estimate of the parameters and an estimate of the estimation error covariance matrix. It has here been proposed to extend the stored information with variables that keep track of whether a fault in the process model has occurred or not.

In Chapter 5, slow parameter changes and changes in the excitation were treated. The weighting problem is difficult because of the lack of knowledge about the uncertainties of the different measurements. In Chapter 3, the uncertainties were divided into two terms, where one depends on the noise disturbances and the other is due to changes of the model. The importance of distinguishing between these two terms was also stressed. An increase of the noise level means that future measurements are more uncertain than old ones. The estimator gain should therefore be decreased in this case. An increase of the model error means that old measurements are unreliable. The estimator gain should thus be increased.

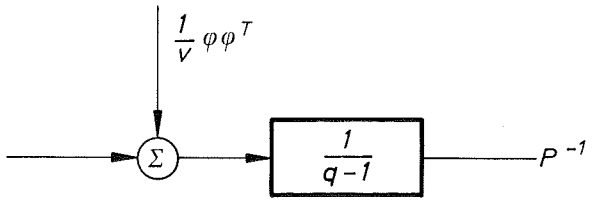
The solution proposed in Chapter 5 results in an estimation scheme which retains a constant desired amount of information in the estimator, when the parameters are constant. It is shown that the method gives a P-matrix which converges to a matrix with desired equal diagonal elements. These diagonal elements can be interpreted as the desired variances of the parameter estimates. (Different variances of the parameter estimates can be obtained by scaling the θ and φ vectors.)

The inverse P-matrix (weighted with the noise variance according to Equation 2.2) is a measure of the information content in the estimator. To be able to track time-varying parameters, it is necessary to keep the inverse P-matrix bounded. This problem could therefore be viewed as a problem of controlling the information used in the estimator. This control problem is described in block diagrams in Figure 7.1.

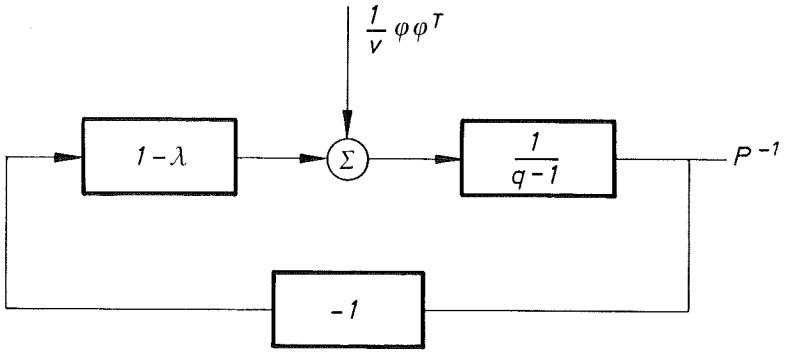
Figure 7.1a shows a block diagram for the updating of $P(t)^{-1}$ in the original LS procedure. The matrix grows boundlessly, since it is the output of an integrator with a positive input signal. The system has to be controlled.

Figure 7.1b shows the updating with a forgetting factor. The system is stabilized, so the inverse P-matrix stays bounded. Since no reference value is specified, $P(t)^{-1}$ fluctuates depending on λ, $\varphi(t)$ and $v(t)$. Some eigenvalues may approach zero, which leads to the problems described in Chapter 3.

a.



b.



c.

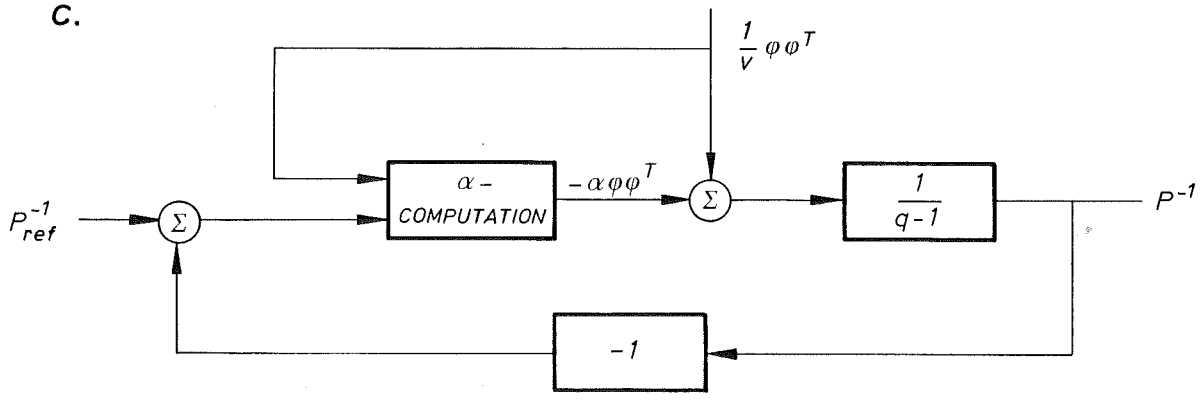


Figure 7.1 - Block diagrams describing the updating of the inverse P-matrix.
a. The original LS procedure.
b. LS with forgetting factor.
c. The new proposed method.

Figure 7.1c shows a simplified version of the scheme described in this thesis. In control systems terminology, the method consists of a feedback from the stored information and a feed-forward from the incoming information.

The problem of "controlling" the P-matrix is of course not as simple as indicated in Figure 7.1. Since the updating of the P-matrix is a part of a complex control system, other demands limit the design possibilities. The bounds on $\alpha(t)$ to obtain stability and the requirement of keeping the P-matrix positive are some examples.

The major advantage of the proposed method is that it solves the problems caused by nonuniform excitation (both in time and in space) of the process. In other methods, the P-matrix is prevented from becoming too large by additional supervisory loops. This supervisory level, or safety net, also contains logic to interrupt the estimation in case of bad excitation. In the new method, this logic can be eliminated, since these problems are handled automatically by the algorithm. A nice feature is also that ad hoc choices of forgetting factors are replaced by the performance related choices of desired parameter variances.

In the final example in Chapter 6, both the new fault detection method and the new way of controlling the P-matrix to a diagonal matrix were used. It should be noted that it is not always necessary to use the total algorithm given by Equation 6.1. The updating of $v(t)$ may be omitted if the noise variance is constant. If the changes of the model are known to be slow, the fault detection procedure can be excluded. What is meant by "slow changes" can be computed automatically, since the values of $\alpha(t)$, $v(t)$, $\varphi(t)$ and $P(t)$ determines the time horizon of the estimator. See Equation (5.36). This knowledge provides also possibilities to adjust the fault detection procedure dynamically.

As mentioned several times before, the methods are not limited to the least squares procedure. The fault detection procedure has in this thesis been applied to three different algorithms, the least squares method with a constant forgetting factor, the Kalman filter and the least squares method with the new discounting method. Since the inputs to the detector are the parameter estimates, see Figure 4.6, the detector can be applied to almost

any estimation scheme.

The new way of discounting past data described in Chapter 5 is concerned with the updating of the covariance matrix. Since most estimation schemes include an updating of a covariance matrix, the new method can also be applied to many other algorithms.

8. REFERENCES

Åström K J (1970): Introduction to Stochastic Control Theory. Academic Press, New York.

Åström K J (1980): Piece-wise deterministic signals. In O D Andersson (ed.) "Time Series", North-Holland Publishing Company.

Åström K J (1983): Theory and Applications of Adaptive Control. Automatica 19, 471 - 486.

Basseville M (1982): Survey of statistical failure detection techniques. In Thèse d'Etat: Contribution à la détection séquentielle de ruptures de modèles statistiques. University of Rennes. France.

Bierman G J (1977): Factorization Methods for Discrete Sequential Estimation. Academic Press, New York.

Chung K L (1968): A Course in Probability Theory. Academic Press, New York.

Cramér H (1945): Mathematical Methods of Statistics. Princeton University Press, New Jersey.

Egardt B (1980): Unification of some discrete-time adaptive control schemes. IEEE Trans. Automatic Control AC-25, 693 - 697.

Elmqvist H (1975): SIMNON, An Interactive Simulation Program For Nonlinear Systems, Report TFRT-3091, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Evans R J and Betz R E (1982): New results and applications of adaptive control to classes of nonlinear systems. Proc. Workshop on Adaptive Control, Florence, Italy.

Fortesque T R, Kershenbaum L S and Ydstie B E (1981): Implementation of Self-tuning Regulators with Variable Forgetting Factors. Automatica 17, 831 - 835.

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

Gradshteyn I S and Ryzhik I M (1965): Table of Integrals, Series and Products. Academic Press, New York.

Huber P J (1964): Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73 - 101.

Hägglund T (1982): Adaptive control with fault detection. Report TFRT-7242, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Hägglund T (1983): Recursive least squares identification with forgetting of old data. Report TFRT-7254, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Irving E (1979): New developments in improving power network stability with adaptive control. Proc. Workshop on Applications of Adaptive Control, Yale University, New Haven.

Kendall M G and Stuart A (1961): The Advanced Theory of Statistics - Volume 2. Charles Griffin & Co, London.

Kesten H (1958): Accelerated Stochastic Approximation. Annals of Mathematical Statistics 29, 41 - 59.

Ljung L and Söderström T (1983): Theory and Practice of Recursive Identification. MIT Press, Cambridge.

Lo J T-H (1972): Finite-Dimensional Sensor Orbits and Optimal Nonlinear Filtering. IEEE Trans. Information Theory IT-18, 583 - 588.

Luenberger D G (1973): Introduction to linear and nonlinear programming. Addison-Wesley, Reading, Massachusetts.

Millnert M (1982): Identification and control of systems subject to abrupt changes. Department of Electrical Engineering, Linköping University, Linköping, Sweden.

Moler C B (1981): MATLAB User's Guide. Report TR CS81-1, Dept. of Comp. Sci, University of New Mexico, Albuquerque, New Mexico.

Poulsen N K and Holst J (1982): Robust Self-tuning Controllers in Nonstationary Situations. Preprints "ISIS Workshop on adaptive control". Florence, Italy.

Sorenson H W and Alspach D L (1971): Recursive Bayesian Estimation Using Gaussian Sums. Automatica 7, 465 - 479.

Wald A (1947): Sequential Analysis. John Wiley, New York.

Wellstead P E and Sanoff S P (1981): Extended self-tuning algorithm. Int. J. Control, 34, 433 - 455.

Wetherill G B (1966): Sequential Methods in Statistics. Chapman and Hall, London.

Willsky A S (1976): A Survey of Design Methods for Failure Detection in Dynamic Systems. Automatica 12, 601 - 611.

Wittenmark B (1979): A Two-Level Estimator for Time Varying Parameters. Automatica 15, 85 - 89.