DTU Library

Robust Self Tuning Controllers

Poulsen, Niels Kjølstad; Holst, Jan

Publication date:
1985

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

[Link back to DTU Orbit](#)

Citation (APA):
Poulsen, N. K., & Holst, J. (1985). *Robust Self Tuning Controllers*. Ph.D Thesis 44

ROBUST SELF TUNING CONTROLLERS

Niels Kjølstad Poulsen

imSor

Preface.

====================

The present thesis concerns robustness properties of adaptive controllers. It is adressed to methods for robustifying self tuning controllers with respect to abrupt changes in the plant parameters.

In the thesis an algorithm for estimating abruptly changing parameters is presented. The estimator has several operation modes and a detector for controlling the mode. A special self tuning controller has been developed to regulate plant with changing time delay.
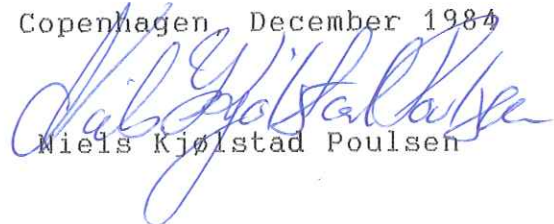
This thesis has been prepared at the Institute of Mathematical Statistics and Operational Research (IMSOR), the Technical University of Denmark, as a part of the requirement for the degree Lic. Tech. (Ph. D. in Engineering).

I would like to express my gratitude to my supervisor, Associate Professor, Tech. Dr. Jan Holst for his valuable suggestions and critical comments.

I would furthermore like to thank Birgit Gade, Elbert Hendricks and Susanne Nørgård Jensen for their help in connection to the manuscript, and finally I want to thank Karin Pehrsson Skettrup for preparing the figures and Svend Aage Jensen for printing the volume.

Copenhagen, December 1984

Niels Kjølstad Poulsen

Contents.

═══════════════════

1

Introduction.

====================================

1.1 Background.

The continuously increasing demand for effectivity and optimality in the production calls for better automatic control. This desire has been actualized by the evolution in the energy - and resource areas during the last two decades. The appearence of microcomputers has made it possible to use more complex control algorithms as e.g. adaptive controllers, which adapt the control to the actual process, meaning that a very high performance can be achieved.

In a conventional control system the determination of the control signal is based on a model of the process which is to be controlled. The type and complexity of the model are determined by the specific control structure. In certain cases it is rather difficult to determine a model which describes the process in the desired detail. This happens e.g. when the process is very complex or when the physics are not fully understood. Modelling the dynamics of the stochastical disturbances from the environment is especially difficult. It is rather important for a controller which is designed to reduce the effect of these disturbances that this part of the model describes the actual situation in reasonable detail.

Today many statistical methods for a'priori determination of a process model exist. These methods are, however, not always applicable, because the properties of the process usually change because of age, external disturbances or because of changes in the working conditions. This problem can be defeated by using a recursive estimation method which performs an identification of the plant on line. These recursive estimation methods can as shown later in this thesis be designed to handle timevarying plants.

It is possible to compensate for an insufficient knowledge of

the plant by applying an adaptive controller. In a self tuning
controller the identification of the process and the
determination of the control signal, u(t), are carried out
simultaneously. Normally, the on line identification is
carried out by using a recursive estimation method. It is
possible, as shown in figure 1, to regard a self tuning
controller as consisting of 3 interconnected parts, an
identification part (I), a design part (D) and a control
part (C). In the identification part the parameters of the
model, i.e. the estimates, are continuously updated in
accordance with the measurements on the plant Based on a
design criterium these estimates are in the design part
transformed into a set of control parameters which are used in
the control part for the determination of the control signal.
Because of the complexity of the self tuning controllers, they
are mainly implemented on computers and are consequently
normally based on a discrete time model of the process.
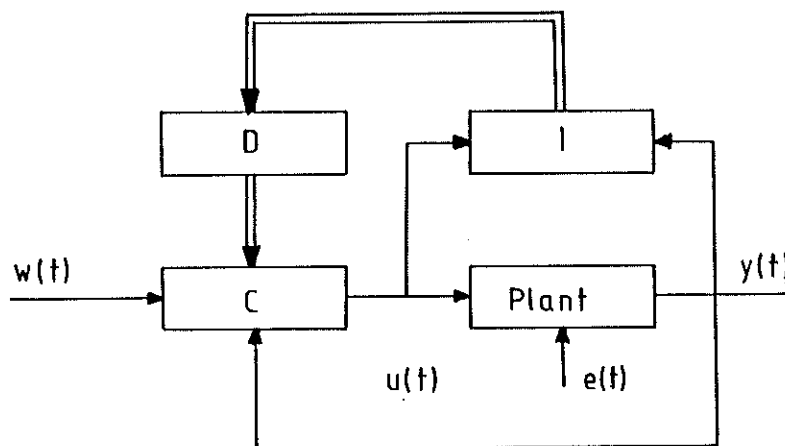
Figure 1: Basic structure of a self tuning controller.

It is possible to parameterize the description of a plant.
This means that the plant characteristics are divided into
some properties which are determined by the plant parameters
and some structual properties as e.g. the system order, number
of inputs etc.

When using a self tuning controller the required a'priori

knowledge is significantly reduced because the plant parameters are estimated recursively. However, the design of a self tuning controller is still founded on some basic assumptions connected to the structure of the plant and the statistical properties of the disturbances. When a self tuning controller is designed, it is normally assumed that the process can be described by a linear model with constant parameters. It is furthermore assumed that the structure of the process is known. This requires in most cases, that the system order and the time delays through the process are be known. No process is, however, truly linear or has a finite system order, but it is believed to be a good local approximation in most cases.

The identification part offers many possible different approaches. If a maximum likelihood approach is chosen, it is customary to assume that the disturbances have a gaussian distribution.

The design assumptions mentioned in the previous paragraph are introduced as mathematical rationalizations, and they are in most practical cases approximately fulfilled. It would, however, be rather convenient if the self tuning controller also in a practical application, is able to handle an exceptional situation, which might be caused by an assumption suddenly failing to be valid. The general intention with the present work is to investigate the properties of self tuning controllers and to modify them in order to improve their robustness properties, i.e. to reduce their sensivity with respect to the design assumptions.

In principle, a self tuning controller is designed to control processes for which the parameters are unknown and constant. However, one motivation for using a self tuning controller instead of an off line identification method combined with control based on the resulting model is a desire of tracking slow variations in the plant parameters. Several methods exist

for modifying the self tuning controllers in order to track such slow variations or drift in the plant parameters. There also exist methods which can handle plants whose parameters have a quick but regular evolution, i.e. without large and sudden jumps or changes.

If the process is nonlinear, a change in the working conditions, caused e.g. by a change in the set point or in the characteristics of the external disturbances, might change the plant parameters in the local linear model in an abrupt manner. Abrupt changes in the plant parameters might also occur because of failure in the actuators, the measurering instruments or in the process itself.

The abrupt changes in the plant parameters will cause a serious problem due to their irregularity. Consider a self tuning controller which is designed to handle a plant with slowly varying parameters. If a forgetting method is applied in the identification part, a large estimation horizon is used in order to obtain a reasonable performance. If an abrupt change occur in some of the plant parameters the quality of the control will be highly reduced for a long period corresponding to the length of the estimation horizon. If now the forgetting algorithm is modified in order to track a possible abrupt change, the estimation horizon must be rather short. Perhaps it is possible to track an abrupt change if such one would occur, but the performance in the normal situation is reduced. Especially if the abrupt changes only occur occasionally this method will give unnecessarily large losses.

The problems which are to be addressed in this thesis are connnected to methods for increasing the robustness properties of self tuning controllers with respect to abruptly changing parameters. The goal is an algorithm which is not sensitive to the assumption of constant or regularly varying plant parameters. In the work for making the self tuning controllers

robust with respect to abrupt changes in the process
parameters, the properties of the identification part have
mainly been brought into focus. The methods which are to be
described in this thesis are constructed for processes with
one output and one control signal. It is, however, possible to
extend the methods in order also to make them applicable to
multivariable processes.

Temporarily, focus on the gaussianly distributed disturbances.
It might be a serious problem for conventional estimators if
the distribution for the physical data deviates from the
assumed gaussian. Such a deviation occurs if the disturbance
has outliers or gross errors, i.e. extremely large noise
signals. Then the density function for the innovations has
heavier tails than the gaussian density function. In a
conventional estimator each and every observation has an
unbounded influence on the estimate by which means a single
outlier may destroy the estimate. The problem is, due to the
short length of the estimation horizon, more serious in an
estimator designed to estimate timevarying parameters. The
methods for robustifying the self tuning controllers with
respect to outliers are very simple to implement and they are
from a practical point of view rather obvious. The method
simply consists in restricting or censoring the observations
to an interval including their expected value. A self tuning
controller which is robust with respect to abrupt changes in
the load disturbances and with respect to outliers is
described in Poulsen (1982) and in Poulsen and Holst (1982).
The problem of improving the robustness properties with
respect to outliers will, however, not be discussed further in
this thesis.

1.2 Outline of The Thesis.

In chapter 2 the plant description will be discussed and some particular models for different kinds of parameter variations will be given. In order to simplify the presentation, the methods are described in connection to a simple model structure. The methods are, however, applicable in general.

The basic design principle of two self tuning controllers which will be used throughout the thesis are given in chapter 3. The first, an explicit LQG - controller, can be used for regulation of unstable and nonminimum phase plants. In addition, the time delay through the plant does not enter in the controller. If supplied with an estimation procedure, which is designed, as the detector algorithm, to handle abruptly changing parameters, this controller can be used upon plants which time delay might change. The second, an implicit poleplacement controller, is cancelling the plant zeroes and can accordingly not be used if the plant zeroes are not well damped.

Chapter 4 is devoted to methods for estimating parameters which have a regular evolution. These methods include the variable forgetting method proposed in Fortescue et al. (1981) and the directional forgetting method given in Hägglund (1983).

The estimation methods which are presented in the following chapters are given in connection to the explicit LQG - controller but they are also applicable together with the implicit poleplacement controller. Actually, the different methods described in this thesis can be compared because of the fact that they are used in simulations with the implicit poleplacement controller.

The detector algorithm, which is the main innovation in this thesis, is designed to estimate the plant parameters even if they in their regular evolution have abrupt and irregular changes, is presented in chapter 5. This algorithm consists of an estimator with several operation modes and a detector which is controlling the mode. If an abrupt change has been detected, the estimator change to one of the exceptional modes in order to perform a branch estimation, i.e. an estimation of the parameter change performed on stored data. The LQG – controller can, if combined with this algorithm be used for controlling plants with a varying time delay.

The detector algorithm was first presented in Poulsen (1982) where also the robustness problem with respect to deviation between the datagenerating and the assumed gaussian distribution was included. Here, only abrupt changes in the load disturbances, were discussed. This algorithm was with minor modifications also presented in Poulsen and Holst (1982). In Holst and Poulsen (1984) the algorithm is extended in order to handle abruptly changes in all the plant parameters.

The thesis ends in chapter 6 with a description of related methods for estimating abruptly changing parameters which are proposed by other authors. Here the detector algorithm can be compared to the filterbank algorithm proposed in Andersson (1983) and to a signtest algorithm described in Hägglund (1983). In Millnert (1982) a slightly different problem is treated, where the algorithm is designed to estimate parameters as they change abruptly between a finite number of points in the parameter space, i.e when the plant description is changing between different models.

The formulas and figures are numbered in succession within a chapter. When referring to a formula or a figure from another chapter, the number of the actual chapter followed by the number of the formula will be indicated.

2

PROCESS DESCRIPTION.

In order to make the design work reasonable, it is common practice to assume that the process can be described by a linear model of finite order. This assumption is of course not true, since all plants contain some kinds of nonlinearities. However, if the nonlinearities are not too strong, it is in many cases a reasonably good local approximation to describe the truely nonlinear process by a linear model. This model is then only valid in a region around the operating point. Thus, the coefficients, in this linear model, which in the sequel are referred to as plant parameters, are related to the operating point. This point might change, due to many causes, and methods for identifying time varying processes are accordingly relevant. The main scope of the thesis is not robustness with respect to deviation between the physical process and the possibility of describing it by a finite order linear model it is in the sequel assumed that a description of the process by a finite order linear model is true.

For the sake of simplicity the discussion is here restricted to univariate systems. Since self tuning controllers normally are implemented on computers the plant models are formulated as discrete time models.

2.1 The Process Equation.

Hence, consider the process equation:

$$A(q^{-1})y(t) = \frac{B(q^{-1})}{F(q^{-1})} u(t) + \frac{C(q^{-1})}{D(q^{-1})} e(t) + d \qquad (1)$$

where A, B, C, D and F are polynomials in the backward shift operator, q^{-1}, i.e.

$$q^{-1}y(t) = y(t-1) \qquad (2)$$

and where A, C, D and F are monic, i.e.

$$A(0) = C(0) = D(0) = F(0) = 1 \qquad (3)$$

The order of the polynomials, A, B and C, are n_a, n_b and n_c, respectively. Due to the fact that the control action is taken after a sampling $b_0 = 0$ and:

$$B(q^{-1}) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \qquad (4)$$

The output signal, y(t), is partly determined by the control signal, u(t), and partly by the noise which consists of an innovation part, driven by the innovation sequence, e(t), and the load disturbance, d. The innovation sequence, e(t), is assumed to be white with zero mean. Furthermore, it is normally assumed that e(t) is gaussianly distributed and has a constant variance equal to σ_o^2. The load disturbance, d, has several interpretations, it can e.g. be regarded as a mean of the disturbances or as the level due to the operating point. The plant parameters, i.e. the coefficients in the polynomials $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$, $F(q^{-1})$ and $D(q^{-1})$ and the load disturbance, are in the sequel implicitly assumed to be time-

varying.

In order to make the presentation of the identification part more simple, it is in the sequuel assumed that C, D and E are equal to 1, i.e. that the process can be described by the difference equation:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + e(t) + d \qquad (\ 5\ )$$

where:

$$A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$$
$$B(q^{-1}) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \qquad (\ 6\ )$$

The reason why the identification methods are presented in connection with this simple structure is the fact that the identification problem then can be solved exact. In the general case some approximations have to be introduced. The methods, which are to be described in the following chapters, are derived in connection to processes which have this structure. However, the methods are easily extended in order to handle processes of the more general form, (1), (see e.g. Appendix 1).

The time delay, k, through the process is determined by the index of the first non-zero coefficient in the B – polynomial, i.e. $b_i = 0$ for i = 1, ... ,k-1 and $b_k \neq 0$. Changes in the time delay can consequently be handled by methods designed with respect to abruptly changing plant parameters. This requires, however, that the time delay is not explicitly included in the model used in the identification part (or in the control part).

If the vectors:

$$\phi(t) = (-y(t-1), \ldots , u(t-1), \ldots , 1)^T \qquad (7)$$

and

$$\theta(t) = (a_1(t), \ldots , b_1(t), \ldots , d(t))^T \qquad (8)$$

are introduced the process (5) can be described by the output equation:

$$y(t) = \phi^T(t)\theta(t) + e(t) \qquad (9)$$

where the parameters, $\theta(t)$, are time varying. Notice, the general model (1) can also be written in this form (see Appendix 1). The statistical properties of $\phi(t)$ depents of course on the evolution of the plant dynamics, i.e. the properties of $\phi(t)$ is a function of the unknown sequence:

$$\theta(s), \quad s<t$$

Since the control signal is generated by a self tuning controller the statistical properties are also a function of the known sequence of estimates, i.e. of:

$$\hat{\theta}(s), \quad s<t$$

2.2 The Timevarying Parameters.

Now, focus on the time variation of the plant parameters, $\theta(t)$. The plant parameters might have different types of evolution related to the type of effect which cause the evolution. The four types of evolution which are to be described in the following are characterized by the rate of change.

In most practical cases the plant parameters have a slow evolution which e.g. is caused by age and wear. This kind of variations are normally very slow compared to the sampling interval or the time constants of the process. The algorithms which are designed to handle this kind of time variations are founded on the assumption that the plant parameters ap-proximately can be regarded as constant within a time interval with a reasonable length.

However, the time variations in the plant parameters might also be faster. Especially when the linear model is obtained as a local approximation to a nonlinear process dynamic, the plant parameters may have a quick evolution. If the parameter variations are regular, i.e. without large jumps or abrupt changes, it is common practice to assume that this kind of variation can be modelled by a simpel Markov equation, i.e.

$$\theta(t+1) = \theta(t) + v_o(t+1) \qquad\qquad (\ 10\ )$$

where the vector, $v_o(t)$, is a white gaussian noise sequence, which is independent of the innovation sequence, $e(t)$. Notice that the output equation (9) forms together with the equation (10) for the parameter variations, a state space model of the plant.

The subject of the thesis is adaptive control of plants which

parameters in addition to their regular evolution have abrupt changes, i.e. irregular and large changes.

The load disturbance, $d(t)$, which here is interpreted as a plant parameter and is included in the parameter vector, $\theta(t)$, has a special status. Since it also might contain external forces it is likely that this parameter might change more frequently than the other parameters. The rest of the plant parameters are determining the dynamic properties of process and are accordingly denoted as dynamic parameters. Hence, it is assumed that a change in the load disturbance can occur independently as well as simultaneously with a change in a subset of the dynamical parameters. This subset might occasionally include all the dynamical plant parameters. Due to the possible difference in frequency and effect on the control quality it is required that the load disturbance and the dynamic parameters are handled seperately. Consequently, at an arbitrary instant of time, t, there are 4 possibilities concerning the abrupt behaviuor of the plant parameters. Firstly, the parameters can remain unchanged. Denote this normal situation by $S_0(t)$. Secondly, there can be a change in the load disturbance only, denoted as $S_1(t)$. Thirdly, let $S_2(t)$ denote the situation that all or a subset of the dynamical parameters, have changed and, finally, let $S_3(t)$ denote a total change. Notice, that the exceptional situation has been seperated into tree subsituations, $S_1(t)$, $S_2(t)$ and $S_3(t)$, and that the two disjoint situations, $S_1(t)$ and $S_2(t)$ are special cases of $S_3(t)$.

Now, one specific model of abruptly changing parameter in a random direction is to be stated. The model for this kind of parameter variations will embody the seperation of the exceptional situation mentioned above. Assume that $\Delta(t)$ is a stochastic variable which takes the values 0, 1, 2 and 3. Let $p_0(t)$, $p_1(t)$, $p_2(t)$ and $p_3(t)$ be the probability of the occurance for $S_0(t)$, $S_1(t)$, $S_2(t)$ and $S_3(t)$, respectively, i.e

$$P\{ \ \Delta(t)=i \ \} \ = \ p_i(t) \tag{11}$$

Then it is possible to model the abruptly changing parameters by the equation:

$$\theta(t+1) \ = \ \theta(t) \ + \ v(t+1)$$

$$v(t) \ = \sum_{i=0}^{3} \delta_{\Delta,i} v_i(t) \tag{12}$$

where

$$v_i(t) \ \varepsilon \ N(0,R_i\sigma_o^2) \tag{13}$$

and is a white noise sequence. Here σ_o^2 is the variance of the innovation sequence and δ is the Kronecker delta. Furthermore let $R_i(t)$ reflect the properties of the situation $S_i(t)$, i.e.:

$$R_1 = \begin{Bmatrix} 0 & \cdots & 0 & | & 0 \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ 0 & \cdots & 0 & | & 0 \\ \hline 0 & \cdots & 0 & | & \times \end{Bmatrix} \qquad R_2 = \begin{Bmatrix} \times & \cdots & \times & | & 0 \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ \times & \cdots & \times & | & 0 \\ \hline 0 & \cdots & 0 & | & 0 \end{Bmatrix}$$

$$\tag{14}$$

$$R_o = \begin{Bmatrix} \times & \cdots & \times & | & \times \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ \times & \cdots & \times & | & \times \\ \hline \times & \cdots & \times & | & \times \end{Bmatrix} \qquad R_3 = \begin{Bmatrix} \times & \cdots & \times & | & \times \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ \cdot & & \cdot & | & \cdot \\ \times & \cdots & \times & | & \times \\ \hline \times & \cdots & \times & | & \times \end{Bmatrix}$$

where $\times$ is a non negative number. The point is that the abrupt changes are sudden, big and not very frequent, i.e. that:

$$R_o \ll R_i, \quad i=1,2,3 \quad \text{and} \quad p_o(t) \gg \sum_{i=1}^{3} p_i(t) \quad (\ 15\ )$$

There exist, of course, other models of abruptly changing parameters. Especially the jump mechanism exhibits a variety of models. Whatever choise of jump mechanisme is assumed it is very important that the resulting algorithm is robust with respect to a mismatch between the actual changes and the assumed model.

Throughout the thesis the properties of the described methods will be illustrated by simulations which results are given in examples.

Example 1: Plant 1.

The properties of the presented detector algorithm will be displayed in simulations with the following plant.

The plant is given by:

$$A(q^{-1}) = 1 - 1.5q^{-1}$$

$$B(q^{-1}) = \begin{cases} q^{-1}(0.01 + 0.8q^{-1}) & t < 65 \\ q^{-1}(1.2 + 0.8q^{-1}) & t \geqslant 65 \end{cases}$$

$$d(t) = \begin{cases} 0.5 & t < 140 \\ -0.5 & t \geqslant 140 \end{cases}$$

$$\sigma_o^2 = 0.1$$

This unstable plant cannot be identified without a stabilizing feedback. The properties of the plant are changing at $t = 65$ from a non minimum phase system to a minimum phase system. This change can also be interpreted as a change in the time delay through the plant. The load disturbance has a change at

t = 140.

Example 2: Plant 2.

In the following chapters several algorithms designed to
estimate time varying plant parameters will be described. In
order to compare the properties of these algorithms a single
plant will be used in the simulations. The test object is
given by:

$$A(q^{-1}) = 1 - 1.5q^{-1} + 0.7q^{-2}$$

$$B(q^{-1}) = q^{-1}(1+0.5q^{-1})$$

$$\sigma_o^2 = 0.1$$

$$d(t) = \begin{cases} 0 & t < 100 \\ 1 & t \geqslant 100 \end{cases}$$

This second order plant is characterized by two complex and
stable poles and a zero in the left hand side of the stability
region. The load disturbance has an abrupt change at t=100. In
some of the shown simulations also the zero changes at t=200
i.e.:

$$B(q^{-1}) = \begin{cases} q^{-1}(1+0.5q^{-1}) & t < 200 \\ q^{-1}(1-0.5q^{-1}) & t \geqslant 200 \end{cases}$$

In some cases, a physical plant has a description which is
known to alternate between a finite number of models. This
means, in particular, that one specific model is suitable in
several periods of time. This parameter variation, which also
can be classified as a sequence of abrupt changes, can be
modelled by the equation:

$$\theta(t+1) = \sum_{i=1}^{M} \delta_{\xi(t),i} \theta_i \qquad (16)$$

where $\xi(t)$ is a M state Markov chain. If the parameter variations are given by (16) the plant parameters will have a number of abrupt changes between a finite number, M, of points in the parameter space. Methods designed to handle abrupt changes in random directions are, of course, applicable in this particular situation too, but are inherently discounting to much information. A method, which design takes its origin in a model as (16), will utilize the special structure of the problem. Such a method has been proposed in Millnert (1982).

28

3

Adaptive Control.

==

In this chapter the basic design principles of two self tuning
controllers will be given. Throughout the chapter it is
assumed that the innovation sequence is gaussianly distributed
and that the plant parameters are constant. Methods for
handling timevarying plant parameters will be the subject of
the following chapters.

The structure in a self tuning controller is shown in figure
1:1. In principle the adaption is not, as in a model reference
adaptive controller, gained by a feedback from an error
between the actual and the desired output, but simply by
combining a parametric identification algorithm and a control
part. However, in spite of the different approaches and basic
philosophies, it can be shown that some adaptive controllers
designed as model reference controllers, are similar to some
self tuning controllers c.f. Ljung and Landau (1978), Landau
(1982) and Egardt (1979).

When a self tuning controller is designed, it is customary to
use the certainty equivalence principle as an ad hoc design
principle. This means that the control part is designed as if
the plant parameters were known. When the self tuning
controller is in operation, the plant parameters simply are
substituted by their estimates. Consequently, neither the
uncertainties of the estimated plant parameters are included,
nor is the possibility of gaining further information about
the plant by using more exitation than required by the control
concept. The control will consequently be operating on a
rather short horizon.

The self tuning controllers could be of either the implicit or
the explicit type. In an explicit self tuning controller the
design procedure is carried out at every sampling points, i.e
the estimated parameters are at every instants of time in the
design part transformed into a set of control parameters which
are used in the control part for determination of the control

signal. In an implicit controller the model used in the identification part is formulated in the control parameters. Consequently, the control parameters are estimated directly and the design part only consists of transfering the estimated control parameters to the control part.

Two self tuning controllers will be described in this thesis, namely an explicit LQG – controller and an implicit poleplacement controller. Firstly, the design and control parts of the two control algorithms will be described. Secondly, the identification part is to be given.

3.1 The Control
and
The Design Part.

The main weight has been attached to the identification part
when making the self tuning controller robust with respect to
abrupt change in the plant parameters. In principle, any
controller, e.g. a PID or a dead beat controller can be
combined with the identification part. However, the properties
of the control part are of course important, because it might
give problems if the control part can be applied just to a
restricted type of of systems, as e.g. stable or minimum phase
systems. If such a restriction exists there has to be a
corresponding restriction on the directions in which the
parameters might changes. The poleplacement controller which
is to be described in the sequel is cancelling the plant
zeroes and can accordingly not be used for controlling non
minimum phase systems. When using this controller in
connection with a plant which is assumed to be a minimum phase
system it must be required that none of the parameters change
in such a way that the plant gets non minimum phase. Such a
requirement can be fulfilled by e.g. a known physical law, but
is in general not very appropriate to deal with. The same
problem occurs when using a deadbeat controller, as e.g.
proposed in Isermann (1977) (see e.g. Poulsen (1982)). Such
a deadbeat controller is not able to control unstable plants.

Furthermore, it is desireable that the required a'priori
information is reduced as much as possible, in order to avoid
the influence from a false a'priori estimate and in order to
be able to handle time variations. In the poleplacement
controller the time delay, k, through the plant is regarded as
a structural parameter. So is the system order. The time delay
is explicitly entering in the model which is used both in the
identification part and in the control part. Hence, this

parameter, k, has to be determined a'priori and it is not in the methods described in this thesis, possible to use the poleplacement controller if the time delay is varying.

3.1.1 The LQG - Controller.

The controller which is to be described here, can handle unstable non-minimum phase systems and does not require a knowledge of the time delay through the plant. In the sequel it is assumed that the plant is given by the model:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t) + d \qquad (1)$$

with known order of the A, B and C polynomials. Notice, that the general model (2:1) can be written in this form. The parameters are, according to the design principle, assumed to be known. Furthermore, it is assumed that the innovation sequence, e(t), is a zero mean white gaussian noise sequence with constant variance.

Let w(t) denote the reference signal or the set point. The controller which is to be used here is a variant of the LQG - controller proposed in e.g. Aström (1982). The controller has been extended in order also to handle a load disturbance. The extention only consists of an additive quantity in the control law. This quantity is proportional to (the estimate of) the load disturbance. The load disturbance can also be compensated by enforcing an integral action in the controller. Such an extention has been used in Rasmussen and Pedersen (1983). The variant used here and in Rasmussen and Pedersen (1983) is also modified in order to get a DC - amplification equal to one in the transfer function from the reference to the output signal.

If $\underline{u}$ denotes the mean of the control signal, the criterium to be minimized is:

$$J = E\{ (y(t)-w(t))^2 + \rho (u(t)-\underline{u})^2 \} \qquad (2)$$

where the non negative scalar, ρ, is the only tuning parameter. The optimal closed loop pole locations, which are embodied in the polynomial, $P(q^{-1})$,

$$P(q^{-1}) = p_0 + p_1 q^{-1} + \ldots + p_{n_p} q^{-n_p}$$

are determined by the spectral factorization:

$$P(q^{-1})P(q) = \rho A(q^{-1})A(q) + B(q^{-1})B(q) \qquad (3)$$

where $P(q^{-1})$ does not have any zeroes outside the stability area.

The spectral factorization might be solved by several methods. In the presented self tuning controller an iterative algorithm proposed in Kucera (1979) is used with good results.
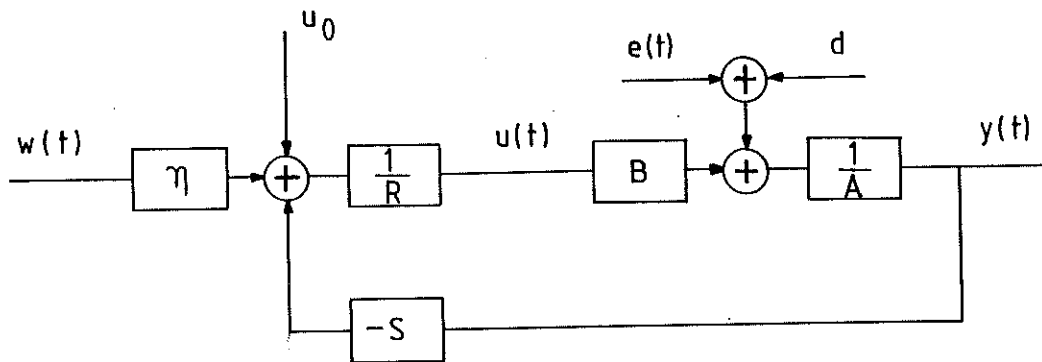
Figure 1: The structure in the LQG - control part for $C(q^{-1})=1$, i.e. the simple structure in (2:5).

The control law is given by:

$$R(q^{-1})u(t) = - S(q^{-1})y(t) + \eta C(q^{-1})w(t) + u_0 \qquad (4)$$

where the order of the control polynomials is:

$$n_r = n_b - 1$$

$$(5)$$

$$n_s = \text{Max}[\ n_a - 1,\ n_p + n_c - n_b\ ]$$

The polynomials:

$$R(q^{-1}) = r_0 + r_1 q^{-1} + \ldots + r_{n_r} q^{-n_r}$$

$$S(q^{-1}) = s_0 + s_1 q^{-1} + \ldots + s_{n_s} q^{-n_s}$$

are determined by the diophantine equation:

$$P(q^{-1})C(q^{-1}) = A(q^{-1})R(q^{-1}) + B(q^{-1})S(q^{-1}) \qquad (6)$$

which in the implemented self tuning controller is numerically solved by the linear equation:

$$
\begin{bmatrix}
1 & & & 0 & & & \\
a_1 & 1 & O & b_1 & 0 & O & \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\
a_{n_a} & \cdot & 1 & b_{n_b} & \cdot & 0 & \\
\cdot & a_1 & \cdot & \cdot & b_1 & \\
& \cdot & \cdot & & \cdot & \cdot & \\
& \cdot & \cdot & & \cdot & \cdot & \\
& O & a_{n_a} & O & b_{n_b} &
\end{bmatrix}
\begin{bmatrix}
r_0 \\ r_1 \\ \cdot \\ \cdot \\ \cdot \\ s_0 \\ s_1 \\ \cdot \\ \cdot \\ \cdot
\end{bmatrix}
=
\begin{bmatrix}
p_0 \\ p_1 \\ \cdot \\ p_{n_p} \\ 0 \\ \cdot \\ \cdot \\
\end{bmatrix}
\qquad (7)
$$

Here p_i are coefficients in the polynomial $P(q^{-1})C(q^{-1})$. Notice that (6) cannot be solved by (7) if $A(q^{-1})$ and $B(q^{-1})$ have common factors. If (1) is obtained from (2:1) with $D(q^{-1}) \neq 1$, the poynomials $A(q^{-1})$ and $B(q^{-1})$ in (1) have a common factor, namely $D(q^{-1})$. Since $D(q^{-1})$ only has zeroes inside the stability area, $D(q^{-1})$ will also be a factor in $P(q^{-1})$. Hence (6) can be solved. In Kucera (1979) an numerical algorithm for solving (6) when $P(q^{-1})$, $A(q^{-1})$ and $B(q^{-1})$ have common factors is proposed.

The factor:

$$\eta = \frac{P(1)}{B(1)} \qquad\qquad (8)$$

is introduced in order to get a DC - amplification equal to 1 in the transfer function from the reference signal, w(t), to the output signal, y(t). Finally, the load disturbance is compensated by the quantity:

$$u_0 = - \frac{R(1)}{B(1)} d \qquad\qquad (9)$$

which is added in the control law. Another method for compensating a load disturbance is to introduce an integrator in the control loop. Such a variant of the LQG - controller has been used successfully by Rasmussen and Pedersen (1983).

With the controller described above the closed loop characteristics are, for correct and constant parameters, given by:

$$P(q^{-1})y(t) = \eta B(q^{-1})w(t) + R(q^{-1})e(t) \qquad\qquad (10)$$

and:

$$P(q^{-1})u(t) = \eta A(q^{-1})w(t) - S(q^{-1})e(t) - \frac{P(q^{-1})}{B(q^{-1})} d \qquad (11)$$

In the minimal variance case ($\rho = 0$), the polynomial $P(q^{-1})$ is equal to the $B(q^{-1})$ polynomial, if it has all its zeroes inside the stability area. If $B(q^{-1})$ contains a zero outside the stability area, then $P(q^{-1})$ contains a corresponding zero which is equal to the reciprocal plant zero (i.e. it is mirroed in the unit circle). This feature makes it possible to use the controller in connection with non minimum phase systems as e.g. plant 1 which is also unstable. For increasing

values of ρ the closed loop poles are moving from the plant
zeroes (or their mirroed versions) towards the plant poles
(or their mirroes versions if unstable).

Example 3: LQG – Controller for plant 1.

For plant 1 which has been presented in example 1 and $\rho = 0.1$
the LQG – control parameters are equal to:

$$R(q^{-1}) = \begin{cases} 0.97 + 1.29q^{-1} & t < 65 \\ 1.45 + 0.84q^{-1} & t \geqslant 65 \end{cases}$$

$$S(q^{-1}) = \begin{cases} 2.41 & t < 65 \\ 1.57 & t \geqslant 65 \end{cases}$$

$$\eta \quad = \begin{cases} 1.02 & t < 65 \\ 1.003 & t \geqslant 65 \end{cases}$$

$$u_0 = \begin{cases} -1.40 & t < 65 \\ -0.57 & 65 \leqslant t < 140 \\ 0.57 & t \geqslant 140 \end{cases}$$

This controller will locate the closed loop poles in 0.15 (
for t < 65) and in – 0.38 (for t $\geqslant$ 65).

In figure 2 the output, y(t), the control, u(t), and the
reference signal, w(t), are obtained in a simulation with this
controller. The plant parameters were estimated with the
detector algorithm described in chapter 5.   ●

Example 4: Dependence of b_k.

In order to illustrate the capability of the control part in
connection with numerically small b_k some properties of the
closed loop system have been calculated for plant 1 when b_1
is varying from -2 to 2. The controller is designed for $\rho = 0$
(minimum variance case). The control polynomials $R(q^{-1})$ and
$S(q^{-1})$ are given by:
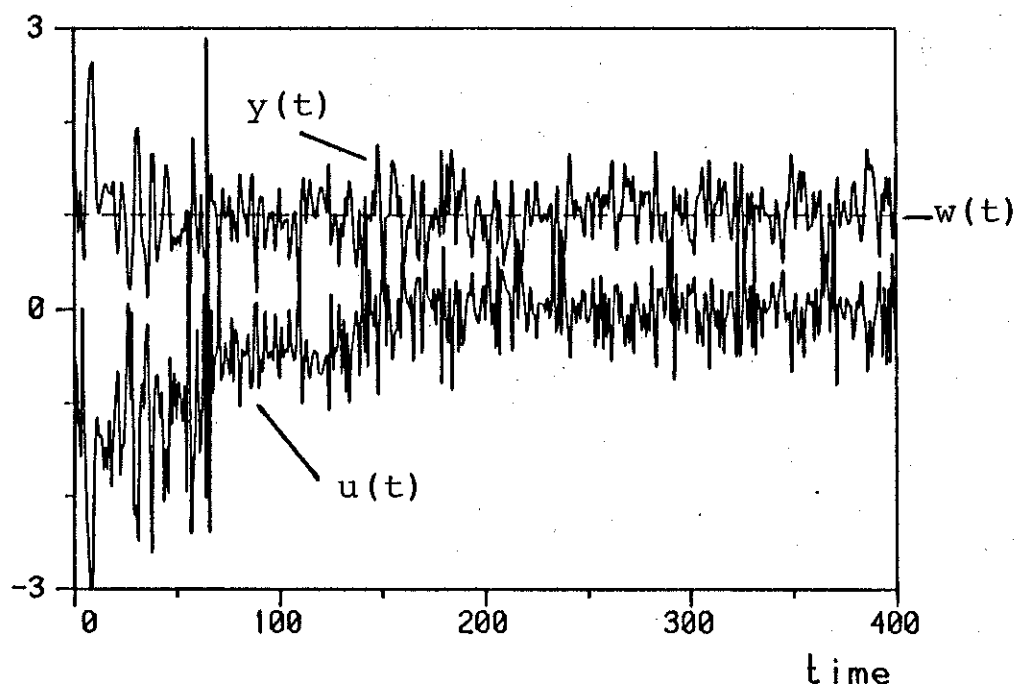
Figure 2: The signals, y(t), u(t) and w(t), obtained in a simulation with the LQG - controller. The plant parameters were estimated (see example 13) with the detector algorithm.

$$R(q^{-1}) = r_o + r_1 q^{-1}$$

$$S(q^{-1}) = s_o$$

In figure 3 the control parameters, r_0, r_1 and s_0, are shown as functions of b_1. In the same figure the optimal closed loop pole is also shown.

The four closed loop AC - amplifications, $A_{y,e}$, $A_{y,w}$, $A_{u,e}$ and $A_{u,w}$ are plotted in figure 4. The AC - amplification, $A_{z,x}$, in a transfer function from the input signal x to the output signal z is defined as the asymptotic variance of the signal, z, if the input signal, x, is a white N(0,1) noise sequence and all other input signals are equal to zero. If $H_{z,x}(q^{-1})$ is the transfer function from x(t) to z(t) the AC - amplification (c.f. Aström (1970)) is given by:
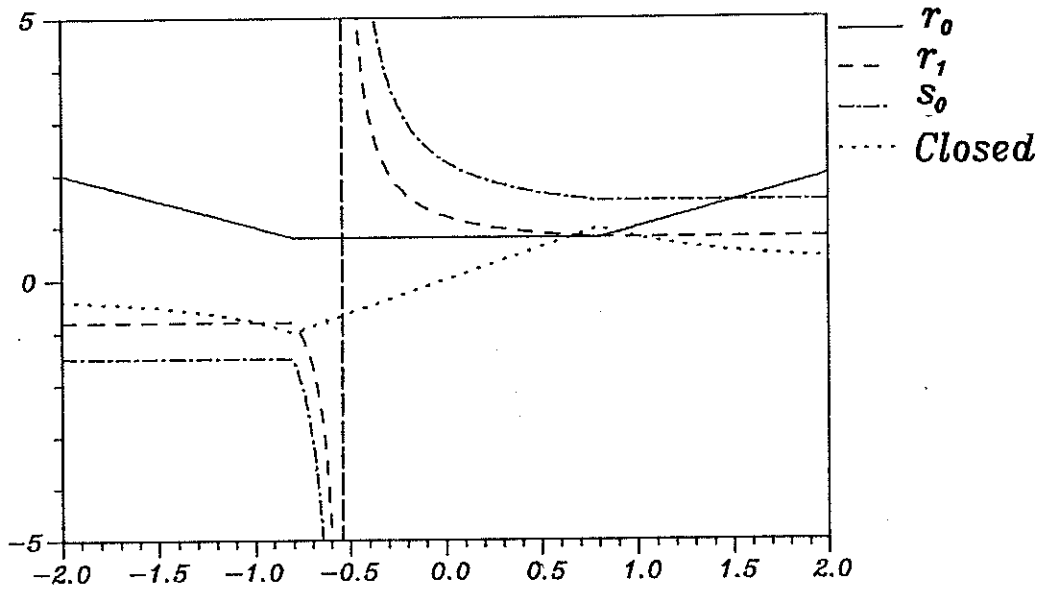
Figure 3: The control parameters and the closed loop pole location versus b_1.

$$A_{z,x} = -\frac{1}{2\pi j} \oint_\Gamma H_{z,x}(q^{-1})H_{z,x}(q)q^{-1} dq$$

where Γ is the unit circle. Accordingly, the AC - amplification is a measure of the amplification averaged over all frequences.

Firstly, it is noted that a minimum variance controller, which for plant 1 (and d = 0) is given by:

$$u(t) = -\frac{1}{b_1} [a_1 y(t) - b_2 u(t-1) + w(t)],$$

or any other controller which is cancelling all the plant zeroes, is not applicable for b_1 within ±0.8. The discontinuities in the overall system properties when b_1 is crossing the stability border are easily seen in both figures. Secondly, for b_1 equal to -0.53 the plant zero coincides with the plant pole. This implies that the unstable plant is

40
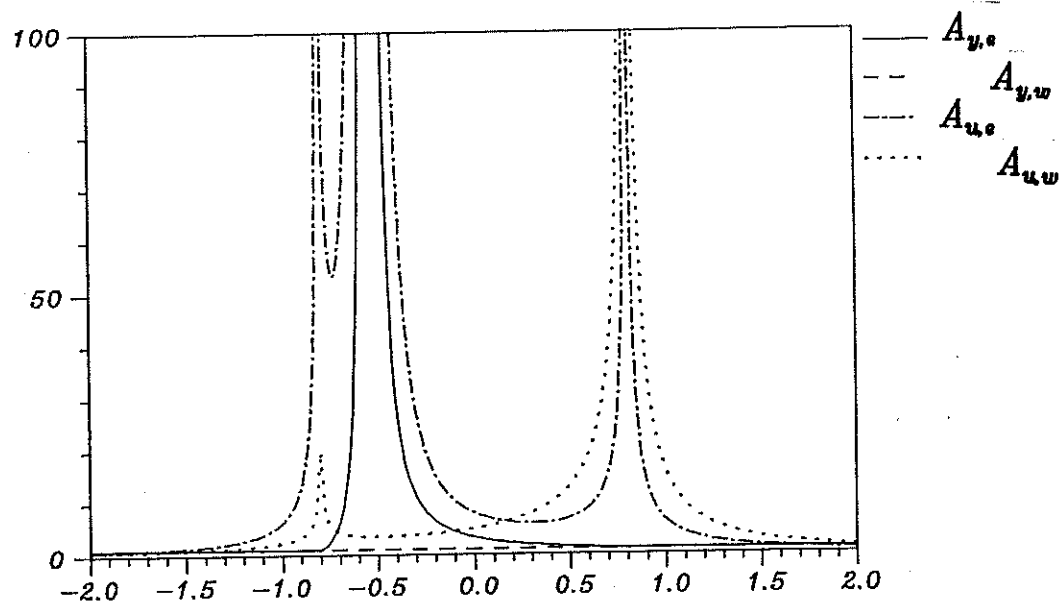
Figure 4: The closed loop AC - amplifications versus b_1.

not both controllable and observable. Such a plant cannot be
controlled by any regulator.

3.1.2 The Poleplacement Controller.

In this section an implicit poleplacement controller is to be described. In Egardt (1979) an unified description of a large class of adaptive controllers including model reference controllers can be found. This description is also used in a comparative study of adaptive controllers in Poulsen (1982).

Assume that the plant is a minimum phase system given by (1) and introduce the ploynomial, $\underline{B}(q^{-1})$, given by:

$$B(q^{-1}) = q^{-k}\underline{B}(q^{-1})$$

$$= q^{-k}[b_k + \ldots + b_{n_b} q^{-n_b}]$$

$$= q^{-k}[\underline{b}_0 + \ldots + \underline{b}_{n_{\underline{b}}} q^{-n_{\underline{b}}}]$$

(12)

where $\underline{B}(q^{-1})$ only has zeroes inside the stability area.

The objective is to make the output signal, y(t), follow the output from a model given by the transfer function:

$$G_m(q^{-1}) = q^{-k} \frac{B_m(q^{-1})}{A_m(q^{-1})}$$

(13)

where A_m is monic of order $\underline{m}$. Normally the DC — amplification, $G_m(1)$, is chosen equal to 1. Introduce the model error, $e_m(t)$, given by:

$$e_m(t) = y(t) - q^{-k} \frac{B_m(q^{-1})}{A_m(q^{-1})} w(t)$$

(14)

The controller has both a deterministic and a stochastic

design which can be found in Egardt (1979) (and in Poulsen (1982)). However, only the stochastic design will be referred here. The criterium to be minimized is:

$$J(t) = E\{ [A_m(q^{-1})e_m(t+k)]^2 \mid Y_t \} \qquad (15)$$

where Y_t denotes the information contained in the signals observed until current time, t. It can be shown (see e.g. Poulsen (1982)) that the minimum for minimum phase system with constant parameters is obtained with the control law:

$$\underline{B}(q^{-1})E(q^{-1})u(t) = C(q^{-1})B_m(q^{-1})w(t) - F(q^{-1})y(t) + u_o \qquad (16)$$

where the control parameters in the polynomials $F(q^{-1})$ and $E(q^{-1})$ are determined by the diophantine equation:

$$A_m(q^{-1})C(q^{-1}) = A(q^{-1})E(q^{-1}) + q^{-k}F(q^{-1}) \qquad (17)$$

The E - polynomial is monic and the orders are equal to:

$$n_e = k - 1$$
$$\qquad (18)$$
$$n_f = Max[\; n_a-1, \; n_c+\underline{m}-k \;]$$

The load disturbance is compensated by u_o which is given by:

$$u_o = - E(1)d \qquad (19)$$

With this control law the closed loop becomes:

$$A_m(q^{-1})y(t) = q^{-k}B_m(q^{-1})w(t) + E(q^{-1})e(t) \qquad (20)$$

Notice, that all plant zeroes are cancelled and are, in the transfer function from w(t) to y(t) substituted by the zeroes embodied in B_m. The control signal is in the closed loop given by:

$$\underline{B}(q^{-1})A_m(q^{-1})u(t) = A(q^{-1})B_m(q^{-1})w(t) - F(q^{-1})e(t) \quad (21)$$

In connection with a minimum phase system this controller can be compared with the LQG - controller for $\rho=0$. Actually, for $A_m = B_m = 1$ they are equal. It can be seen by multiplying the equation (17) by $\underline{B}(q^{-1})$ and compare to (6). For this special case:

$$R(q^{-1}) = \underline{B}(q^{-1})E(q^{-1})$$

$$S(q^{-1}) = F(q^{-1}) \quad\quad\quad\quad\quad\quad (22)$$

$$\eta \quad = 1$$

The next step in the design of the the implicit poleplacement controller is to formulate a model in the control paramters which can be used in the identification part. Then the result from the identification part is the control parameters which only has to be transfered to the control part. It is assumed that $\underline{b}_0$ and the time delay, k, from the control to the output signal, are known.

In order to formulate a general model which includes the large class of adaptive controllers the two asymptotically stable polynomials

$$\underline{Q}(q^{-1}) = 1 + q_1 q^{-1} + \ldots + q_{n_q} q^{-n_q}$$

$$(23)$$

$$\underline{P}(q^{-1}) = 1 + p_1 q^{-1} + \ldots + q_{n_p} q^{-n_p}$$

$$= P_1(q^{-1})P_2(q^{-1})$$

are introduced. It is assumed that the two polynomials P_1 and P_2 are monic. These two design polynomials, $\underline{P}$ and $\underline{Q}$, assure sufficient flexibility to cover both model reference

and self tuning controllers. The specific choice of polynomials is characteristic for the actual controller. Define the filtered error, $e_f(t)$, by:

$$e_f(t) = \frac{Q(q^{-1})}{P(q^{-1})} e_m(t) \qquad (24)$$

The system (1) and the equation, (17), give the recursion:

$$C(q^{-1})A_m(q^{-1})e_m(t) = q^{-k}\{ F(q^{-1})y(t) + B(q^{-1})E(q^{-1})u(t)$$

$$- C(q^{-1})B(q^{-1})w(t) \} \qquad (25)$$

$$+ E(q^{-1})d + C(q^{-1})E(q^{-1})e(t)$$

Additionally, define the signals:

$$y(t) = \frac{Q(q^{-1})}{A_m(q^{-1})P(q^{-1})} y(t), \quad u(t) = \frac{Q(q^{-1})}{A_m(q^{-1})P(q^{-1})} u(t) \qquad (26)$$

$$w(t) = \frac{B_m(q^{-1})Q(q^{-1})}{A_m(q^{-1})P(q^{-1})} w(t), \quad u_0 = - \frac{Q(q^{-1})E(q^{-1})}{A_m(q^{-1})P(q^{-1})} d$$

$$u_1(t) = P_2(q^{-1}) u(t), \qquad e(t) = - \frac{Q(q^{-1})E(q^{-1})}{A_m(q^{-1})P(q^{-1})} e(t)$$

and the polynomial:

$$G(q^{-1}) = q [B(q^{-1})E(q^{-1}) - b_0 P_2(q^{-1})] \qquad (27)$$

$$n_g = Max[n_b + k - 1 , n_{P_2}]$$

where the result, $n_g = -1$, is interpreted as $G(q^{-1}) = 0$. The

polynomial, $\underline{b}_0 P_2(q^{-1})$, has been separeted from $\underline{B}(q^{-1})E(q^{-1})$ since it is costumary to fixe $\underline{b}_0$ a'priori. Notice, the signals, $\underline{w}$, $\underline{y}$, $\underline{u}$ and e_f are known because the signals y, u, and w and the polynomials $\underline{P}$, $\underline{Q}$, B_m and A_m are known.

Now, the equation, (25), can be formulated in the filtered signals:

$$e_f(t) = \frac{1}{C(q^{-1})} \{ q^{-k}[F(q^{-1})\underline{y}(t) + q^{-1}G(q^{-1})\underline{u}(t)$$
$$+ \underline{b}_0\underline{u}_1(t) - C(q^{-1})\underline{w}(t)] - \underline{u}_0 \} + \underline{e}(t)$$

(28)

$$= \frac{1}{C(q^{-1})} \{ \underline{\phi}^T(t)\underline{\theta} + \underline{b}_0\underline{u}_1(t-k) - \underline{w}(t-k) \} + \underline{e}(t)$$

if the vectors:

$$\underline{\phi}(t+k) = (\underline{y}(t), \dots, \underline{u}(t-1), \dots, -\underline{w}(t-1), \dots, -1)^T$$

(29)

$$\underline{\theta} = (f_0, \dots, g_0, \dots, c_1, \dots, \underline{u}_0)^T$$

are introduced. Formulated in the filtered signals the control law, (16), is:

$$\underline{u}_1(t) = \frac{1}{\underline{b}_0} [C(q^{-1})\underline{w}(t) - F(q^{-1})\underline{y}(t) - q^{-1}G(q^{-1})\underline{u}(t) + \underline{u}_0]$$

(30)

$$= \frac{1}{\underline{b}_0} [\underline{w}(t) - \underline{\phi}^T(t+k)\underline{\theta}]$$

where:

$$u(t) = \frac{A_m(q^{-1})\underline{P}_1(q^{-1})}{\underline{Q}(q^{-1})} u_1(t) \qquad\qquad (\ 31\ )$$

Notice, the controller is cancelling the term:

$$\underline{\phi}^T(t+k)\underline{\theta} + \underline{b}_0\underline{u}_1(t) - \underline{w}(t) \qquad\qquad (\ 32\ )$$

in the model (28). If the signal:

$$\underline{y}_1 = e_f(t) - \underline{b}_0\underline{u}_1(t-k) + \underline{w}(t-k) \qquad\qquad (\ 33\ )$$

is introduced, the model:

$$\underline{y}_1(t) = \underline{\phi}^T(t)\underline{\theta} + \underline{e}(t) \qquad\qquad (\ 34\ )$$

can be used in the identification part for estimating the control parameters in the selftuning poleplacement controller. If the control parameters are to be estimated by RLS it must be required that the regressors in, $\underline{\phi}(t)$, are uncorrelated with $\underline{e}(t)$. This requirement is c.f. (26) fulfilled if:

$$\underline{Q}(q^{-1}) = A_m(q^{-1})\underline{P}(q^{-1}) \qquad\qquad (\ 35\ )$$

Example 5: Poleplacement controller for plant 2.

The minimum variance controller proposed in Åström and Wittenmark (1973) is also included in the general description given above. This controller is obtained by letting:

$$\underline{P}(q^{-1}) = \underline{Q}(q^{-1}) = 1$$

$$A_m(q^{-1}) = B_m(q^{-1}) = 1$$

For plant 2 (see example 2) the control parameters are:

$$F(q^{-1}) = -a_1(t) - a_2(t)q^{-1} = 1.5 - 0.7q^{-1}$$

$$G(q^{-1}) = \underline{b}_1(t) = 0.5$$

$$u_o = d = \begin{cases} 0 & \text{for } t < 100 \\ 1 & \text{for } t \geqslant 100 \end{cases}$$

Additionally, in the simulations where the plant zero is changing sign, the G – polynomial becomes:

$$G(q^{-1}) = \begin{cases} 0.5 & \text{for } t < 200 \\ -0.5 & \text{for } t \geqslant 200 \end{cases}$$

3.2 The Identification Part.

The scope of this chapter is to discuss the identification part of the self tuning control algorithms. In the explicit self tuning LQG - controller the identification is based on a model in the plant parameters which are estimated. These estimates are as if they were correct, transformed in the design part into a set of control parameters, which are used in the control part for determination of the control signal. Throughout the thesis the methods will be developed or described in connection with the explicit LQG - controller. In the implicit poleplacement controller the estimation problem is transformed and a model in the control parameters is used instead. In the end of this section an estimation scheme, which is used in the implicit controller, will be described.

The reason to choose to present the methods in connection with the simple structure embodied in the model (2:5) is that the estimation problem then can be solved in an exact way. If the plant has the general structure, (2:1), its parameters can be estimated by using an extention of the general methods described in Ljung (1982). This extention only consists of introducing the load disturbance in the estimation model and is described in Appendix 1. In these extended methods it is necessary to introduce an approximation of the gradient of the prediction error, because the innovation sequence inherently is unknown. In connection with this approximation several choices are possible and in order to simplify the presentation the methods are derived for systems which have the simple structure given by (2:5). The presented methods, including methods for estimating abruptly changing parameters, can be extended in a manner which make them applicable in connection with plants which have a more general structure. Only another interpretation of $\theta(t)$ and the states $\phi(t)$ has to be introduced.

For constant plant parameters it is possible to apply several approaches in order to estimate the parameters in the model (2:5). Here the the Bayesian approach will be used. Assume that:

$$\theta_o(0) \; \varepsilon \; N(\hat{\theta}_o(0), P_o(0)\sigma_o^2) \tag{36}$$

and let Y_t denote the information, i.e. $y(s)$ and $u(s)$ $s \leqslant t$, which is available at the current time, t. Assume that the model (c.f. (2:7) and (2:9)),

$$\theta_o(t+1) = \theta_o(t)$$

$$y(t) = \phi^T(t)\theta_o(t) + e(t) \tag{37}$$

$$e(t) \; \varepsilon \; N(0, \sigma_o^2)$$

is valid. The innovation, $e(t)$, is a gaussian white noise sequence. The estimation problem is to determine an estimate, $\hat{\theta}_o(t) = \hat{\theta}_o(t|t)$ of $\theta_o(t)$ based on Y_t. As shown in e.g. Jazwinsky (1970) the conditional expectation:

$$\hat{\theta}_o(t|t) = E\{\theta_o(t)|Y_t\} \tag{38}$$

is the optimal solution to a large class of criteria including the quadratic:

$$J(t) = E\{ \; [\; (\theta_o(t) - \hat{\theta}_o(t) \;]^T [\; (\theta_o(t) - \hat{\theta}_o(t) \;] \; \} \tag{39}$$

Let $\phi_N(x; \mu, \sigma^2)$ denote the gaussian density function with mean and variance equal to μ and σ^2, respectively. It is a well known result (see e.g. Jazwinsky (1970)) that, if:

$$p(\theta_o(t)|Y_t) = \phi_N(\theta_o(t); \hat{\theta}_o(t), P_o(t)\sigma_o^2) \tag{40}$$

then

$$p(\theta_o(t+1)|Y_t) = p(\theta_o(t)|Y_t) \qquad\qquad (41)$$

and:

$$p(\theta_o(t+1)|Y_{t+1}) = \frac{p(y(t+1)|\theta_o(t+1),Y_t) \ p(\theta_o(t+1)|Y_t)}{p(y(t+1)|Y_t)}$$

$$(42)$$

$$= \phi_N(\theta_o(t+1);\hat{\theta}_o(t+1),P_o(t+1)\sigma_o^2)$$

where

$$\hat{\theta}_o(t+1) = \hat{\theta}_o(t) + P_o(t+1)\phi(t+1)\varepsilon_o(t+1)$$

$$\varepsilon_o(t+1) = \varepsilon_o(t+1|t) = y(t+1) - \phi(t+1)^T\hat{\theta}_o(t) \qquad (43)$$

$$P_o^{-1}(t+1) = P_o^{-1}(t) + \phi(t+1)\phi^T(t+1)$$

Depending on the applied approach several interesting interpretations of the quantities and recursions in (43) can be given. It is e.g. possible to interpret the estimate, $\hat{\theta}_o(t)$, as obtained by a linear feedback from the prediction error, $\varepsilon_o(t)$, where the amplification:

$$K_o(t) = P_o(t)\phi(t) = \frac{P_o(t-1)\phi(t)}{1 + \phi^T(t)P_o(t-1)\phi(t)} \qquad (44)$$

is time varying. When this feedback is used the estimation error,

$$\tilde{\theta}_o(t) = \theta_o(t) - \hat{\theta}_o(t) \qquad\qquad (45)$$

has a time evolution given by the recursion:

$$\tilde{\theta}_O(t+1) = [\ I \ - \ K_O(t+1)\phi^T(t+1) \ ] \ \tilde{\theta}_O(t)$$

$$- \ K_O(t+1)e(t+1)$$

(46)

If the number of parameters in the vector, θ_O, is equal to n, this recursion is characterized by n-1 eigenvalues equal to 1 and one eigenvalue, $\nu_O(t)$, equal to:

$$\nu_O(t+1) = 1 \ - \ \phi^T(t+1)K_O(t+1)$$

(47)

$$= \ \frac{1}{1 \ + \ \phi^T(t+1)P_O(t)\phi(t+1)}$$

Since the corresponding eigenvector is equal to $K_O(t+1)$, the estimation error, $\tilde{\theta}_O(t)$, is reduced, i.e. multiplied by $\nu_O(t+1)$, in the direction of $K_O(t+1)$.

In the Bayesian approach which has been applied here, the matrix, $P_O(t)\sigma_O^2$, is the conditional variance of the parameters. The quantity:

$$1 \ + \ \phi^T(t+1)P_O(t)\phi(t+1)$$

(48)

which enters in the amplification, $K_O(t+1)$, is the conditional variance of the prediction error, $\varepsilon_O(t+1)$, normalized by σ_O^2.

If the maximum likelihood or the least squares approach is applied then the estimate is obtained as the argument which minimize the criterium, $J(t,\theta_O)$, given by the recursion:

$$J(t+1,\theta_O) = J(t,\theta_O) + \rho(t+1,\theta_O)$$

(49)

$$\rho(t,\theta_0) = \frac{1}{2} [y(t) - \phi^T(t)\theta_0]^2$$

A Newton method (see appendix 2) also results in the recursion, (43). Here the matrix, $P_0^{-1}(t)$, is the hessian matrix to the criterium, $J(t,\theta_0)$, and the quantities:

$$-\phi^T(t+1)\varepsilon_0(t+1) \quad \text{and} \quad \phi(t+1)\phi^T(t+1) \tag{50}$$

are the first and the second derivative, respectively, of $\rho(t+1,\theta)$ evaluated in $\hat{\theta}_0(t)$.

Consider the output or the linear regression equation, (37). When an observation, $y(t)$, is achieved, information about a specific linear combination of plant parameters is obtained. It is possible to regard this situation as if information about the plant is obtained in the direction of the regressor vector, $\phi(t)$, which in the sequel is denoted as the direction of exitation.

If the recursion for the hessian matrix, $P_0^{-1}(t)$, is considered it is possible to regard the matrix, $P_0^{-1}(t)$, as an extended form of counter.

As shown in section 3.1.2 the model, (34), used in the identification part of the implicit poleplacement controller can for constant parameters be formulated as:

$$\underline{\theta}_0(t+1) = \underline{\theta}_0(t)$$
$$\tag{51}$$
$$\underline{y}_1(t) = \underline{\phi}^T(t)\underline{\theta}_0(t) + \underline{e}(t)$$

This model has the same form as (37) except here $\underline{e}(t)$ is a MA(k-1) process. Since $\underline{\phi}(t)$ and $\underline{e}(t)$ are uncorrelated the parameters can be estimated with a recursive least squares algorithm.

For numerical reasons the recursions for $P_o(t)$ should not be implemented as in (43) because it requires a matrix inversion at every time instant. This is an awkward numeriacal problem, which e.g. in a microcomputer with a small machine precision, might give problems. Using the inversion lemma the recursion for $P_o(t)$ becomes:

$$P_o(t+1) = P_o(t) - \frac{P_o(t)\phi(t+1)\phi^T(t+1)P_o(t)}{1 + \phi^T(t+1)P_o(t)\phi(t+1)}$$

or (52)

$$P_o(t+1) = [I - K_o(t+1)\phi^T(t+1)] P_o(t)$$

These forms are frequently used. However, due to the finite machine precision the $P_o(t)$ - matrix might lose its property of being positive definite. It is an experimental fact that this might especially be a serious problem when using exponential forgetting (see. section 4.1.3). In the simulation, shown in this volume, a numerically stabilized kalman algorithm has been used. This algorithm has been described in Bierman (1973). If the numerical problems are serious due to a small machine precision, a square root or a UD - algorithm can be applied. In the UD - algorithm, which also has been proposed in Bierman (1973), the $P_o(t)$ - matrix is factorized:

$$P_o(t) = U(t)D(t)U^T(t) \qquad (53)$$

where $U(t)$ is an upper triangular matrix which diagonal elements are equal to 1. The matrix, $D(t)$, is a diagonal matrix. The recursion for $P_o(t)$ is transformed into a set of numerically well conditioned recursions for $U(t)$ and $D(t)$.

If an a'priori estimate of $\theta_o(0)$ exists, then $\hat{\theta}_o(0)$

and $P_0(0)$ should be chosen in accordance with this estimate. If such an estimate is not avaible, it is costumary to chose:

$$\hat{\theta}_0(0) = 0, \qquad P_0(0) = \underline{\kappa}I \qquad \underline{\kappa} >> 1 \qquad\qquad (54)$$

where $\underline{\kappa}$ is a large number and I is the unit matrix.

Example 6: Initial values for plant 1.

All the shown simulations for plant 1 (see example 1 and 3) are started in t=-10 and the estimator is started in:

$$\hat{\theta}_0(-10) = (\hat{a}_1, \hat{b}_1, \hat{b}_2, \hat{d})^T$$
$$= (0,0,0,0)^T$$

$$P_0(-10) = 1000\ I_4$$

where I_4 is the unit matrix.

Example 7: Initial values for plant 2.

The estimation algorithm, used in the implicit pole placement controller (see example 2 and 5) which is used in connection with plant 2, is started in:

$$\tilde{\underline{\theta}}_0(0) = (\hat{f}_0, \hat{f}_1, \hat{g}_1, \hat{d})^T$$

$$P_0(0) = 1000\ I_4$$

3.3 Summary.

=============================

Two different types of self tuning controllers have been described in this chapter.

In the explicit LQG - controller the estimated parameters are transformed, as if they were correct, into a set of control parameters which are used in the control part. If the plant has a more general structure than embodied in the model (2:5) the plant parameters can be estimated with an extention of the general algorithms in Ljung and Söderström (1982). This extention is described in appendix 1. The knowledge of the time delay through the plant is neither required in the identification part nor in the control part. Since the optimal closed loop poles are determined by a spectral factorization, the LQG - algorithm can be used for controlling unstable and nonminimum phase systems.

In the implicit poleplacement controller the estimation is based on a model in the control parameters. Even in the general case with $C(q^{-1}) \neq 1$ the control parameters can be estimated with a leasts square algorithm. For correct parameter estimates the prediction error will be a moving average process of order k-1. The value of the leading coefficient, $\underline{b}_o$, and the time delay, k, have to be determined a'priori. The closed loop poles are determined a'priori and the controller is cancelling the plant zeroes, which thus have to be well damped. The algorithm can accordingly not be used for controlling nonminimum phase plants.

4

Identification
of
Timevarying Plants.

The scope of this chapter is to present methods for estimation of plant parameters which have a regular evolution.

If the plant parameters are constant but unknown it is possible to use an off line identification method and base the design of a conventional controller upon the result of this identification. This procedure is, however, not possible if the plant parameters are varying in time. Then, an adaptive controller which is designed to handle time varying plants might be applied. In a self tuning controller the control and the identification are performed simultaneously. Then, it is possible to use a self tuning controller if just the identification part is able to handle the time variations. In order to handle such time varying plants it is common practice only to modify the identification part and to design the control part as if the parameters were constant. This procedure is not strictly correct for all kinds of controllers. Especially controllers based on a criterium which include several future observations are often founded on an assumed constancy of the plant parameters.

The intention with this chapter is twofold. Firstly, the properties of the methods which are to be described in this chapter can be examined in connection with abruptly changing plant parameters. This investigation might indicate in which directions the self tuning controllers are to be modified in order to handle abruptly changing process dynamics. Secondly, some of the estimators which are to be described in chapter 5 and 6 have several operation modes. One of these modes are designed to handle the normal situation, i.e. when the plant parameters have a regular evolution. Consequently, the estimators which are described in this chapter might be used as the normal operation mode in an algorithm designed to handle abruptly changing parameters.

4.1 Forgetting Methods.

Methods for estimating plant parameters which are slowly
varying will be described in this section. The titel includes
a little bit of abuse of the word, forgetting, because all
estimation methods designed to handle time varying parameters
involve a discounting of old data. Here, the term is reserved
to methods which are based on one assumption, namely that the
variations are so slow that the parameters can be regarded as
constant within a horizon with a reasonable length.

4.1.1 The Resetting Method.

The first and most simple method is to reinitiate the
estimation part of the self tuning controller with suitable
intervals, i.e. the resetting method. The tuning parameters of
this method are the reinitiating point, i.e. $\hat{\theta}(nN_o)$ and
$P(nN_o)$, and the length of the intervals, N_o, which might
be timevarying.

The method is applicable if the variations in the plant
parameters are caused by a batch processing where the dynamic
characteristics may depend on the properties of the actual
batch. Normally, the point of time at which a new batch is
initiated, and consequently N_o, is known.

The method is also useable for estimation of slowly varying
plant parameters. Then the algorithm in practice is
reinitiated in:

$$P(nN_o) = P_o, \quad n = 1, \ldots \qquad\qquad (1)$$

where N_o is the fixed length of the intervals and P_o is a
predefined matrix. One disadvantage of this method is that,
depending on P_o, the estimate may be dithering just after a

reinitiation.

4.1.2 The Rectangular Forgetting Method.

Another method that could be used for handling slow parameter variations could be the rectangular forgetting method in which the estimation is based on a number, N, of observations. This is accomplished by determining the estimate, $\hat{\theta}(t)$, as the argument which minimizes the criterium:

$$ J(t,\theta) = \frac{1}{2} \frac{1-q^{-N}}{1-q^{-1}} [y(t) - \phi^T(t)\theta]^2 \qquad (2) $$

instead of the least squares criterium (3.2:49). The storage and computional requirements for a fixed number, N, is rather heavy and the data are consequently often discounted in batches (see e.g. Jazwinsky (1970) or Goodwin and Payne (1977)).

4.1.3 The Exponential Forgetting Method.

The most popular forgetting method is the exponential method where the criterium to be minimized is a least squares criterium which is given by the recursion:

$$ J(t+1,\theta) = \lambda(t+1)J(t,\theta) + \frac{1}{2} (y(t+1)-\phi^T(t+1)\theta)^2 \qquad (3) $$

Notice that old data for constant $\lambda(t+1)$ is exponentially discounted. The determinination of $\hat{\theta}(t+1)$, i.e. the minimization of $J(t+1,\theta)$, gives (see appendix 2 and 3) exactly the same recursions as the ordinary least squares in (3.2:43) except for the recursion which determines the P(t) matrix. This is in the exponential forgetting method given by:

$$ P^{-1}(t+1) = \lambda(t+1)P^{-1}(t) + \phi(t+1)\phi^T(t+1) \qquad (4) $$

This simple extension explains the popularity of the method. The tuning parameter, $\lambda(t)$, can be determined in several manners. It may be constant, it may be depending on the prediction error or it may be a function of the $P(t)$ matrix, as shown below. The constant forgetting factor is chosen with respect to the equivalent horizon:

$$N_o = \sum_{i=0}^{\infty} \lambda^i = \frac{1}{1-\lambda}$$

and the length of the interval in which the parameters are assumed to be constant, i.e.

$$\lambda = 1 - \frac{1}{N_o} \qquad (5)$$

4.1.4 The Variable Forgetting Method.

In an algorithm proposed by Fortescue et al. (1981), the forgetting factor depends on the magnitude of the prediction error, which statistical properties depend on the difference between the estimated and correct parameters. The basic idea is to determine $\lambda(t)$ in such a manner the criterium, (3), evaluated in the estimate, $\hat{\theta}(t)$, is constant, i.e.:

$$J(t+1,\hat{\theta}(t+1)) = J(t,\hat{\theta}(t)) = J_o \qquad (6)$$

Usually the tuning parameter, J_o, is transformed into:

$$J_o = \frac{1}{2} N_o \sigma_o^2$$

where σ_o^2 is the innovation variance and N_o is an asymptotic horizon which becomes the new tuning parameter. It can easily be shown that (see e.g. Albert and Sittler (1966) or appendix 2 and 3):

$$J(t+1, \hat{\theta}(t+1)) = \lambda(t+1)J(t, \hat{\theta}(t)) \qquad\qquad (7)$$
$$+ \ \frac{1}{2} \ [1 - \phi^T(t+1)P(t+1)\phi(t+1)]\epsilon^2(t+1)$$

Since $P(t+1)$, due to (4), depends on $\lambda(t+1)$, the solution of this equation is equivalent to a solution of a quadratic equation, given in exact form in appendix 3. For small values of $\epsilon(t+1)$ ($\epsilon^2(t+1) \ll N_o\sigma_o^2$) the forgetting factor, $\lambda(t+1)$, has the approximation:

$$\lambda(t+1) = 1 - \frac{1}{N_o} \ \frac{\epsilon^2(t+1)}{[1+\phi^T(t+1)P(t)\phi(t+1)]\sigma_o^2} \qquad (8)$$

Notice the ration between $\epsilon^2(t+1)$ and its expected value in this expression (under the assumption that $P(t)\sigma_o^2$ is the expected value of the conditional variance of $\theta(t)$). Compare (8) with (5). In figure 1 the exact (shown dashed) and the approximative value of $\lambda(t+1)$ are plotted versus $\epsilon(t+1)/\sigma_o$.

The asymptotic horizon, N_o, and λ_{min}, which has to be introduced in order to avoid too small or even negative values of $\lambda(t)$, are tuning parameters in this method. The method requires a knowledge of σ_o^2 which can be predetermined or estimated. In chapter 5.6 a recusive method for estimating the innovation variance, σ_o^2, is described. In a similar method proposed in Wellstaed and Sanoff (1981) the determination of the forgetting factor is also based upon a single value of the prediction error.

Example 8: Simulation with the variable forgetting method.

The following four figures show results obtained in a simulation with the variable forgetting method proposed in Fortescue et al (1981). The control part of the algorithm is
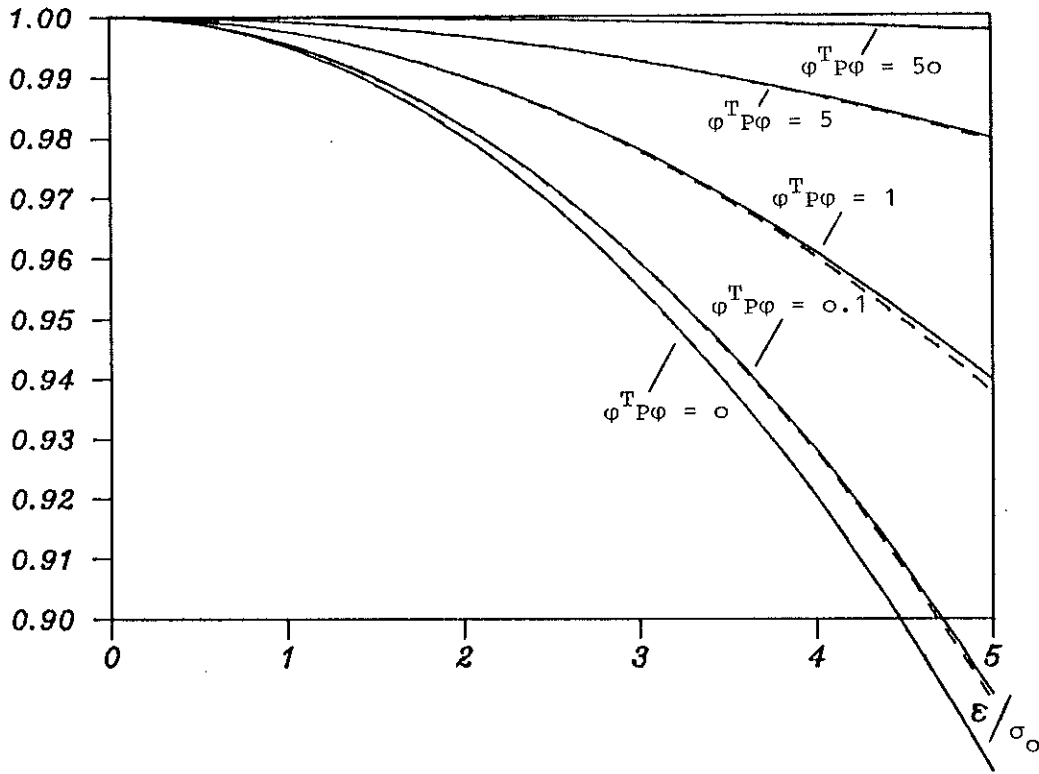
Figure 1: Approximative and exact (shown dashed) value of
$\lambda(t+1)$ plotted versus $\varepsilon(t+1)/\sigma_o$.

the poleplacement controller, described in example 5. The
estimation model is accordingly not expressed in the plant
parameters, but in the control parameters. Plant 2 is the test
object, in which the load disturbance is changing at t=100.
The reason why an abruptly changing plant is used in this
simulation, is because a step change in the parameters exposes
the robustness properties of the algorithm. The tuning
parameter, N_o, i.e. the asymptotic horizon, is chosen to be:

$$N_o = 150$$

which is a reasonable compromise between good tracking
properties, when the parameters might change, and reasonable
stable estimates in the normal situation. The innovation
variance, σ_o^2, is estimated with a recursive variance

estimator (see chapter 5.6 for details and example 11 for simulation results).
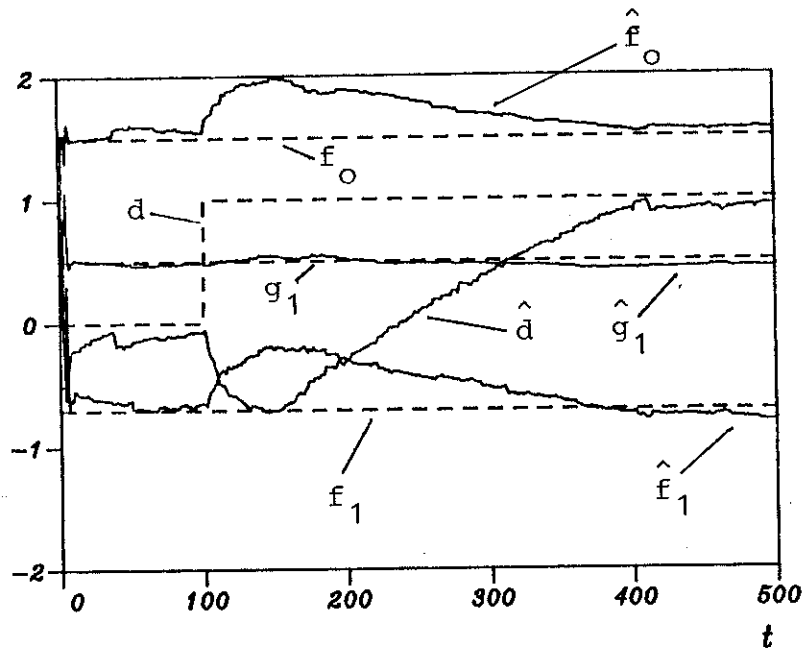
Figure 2: The estimated parameters obtained in a simulation with plant 2 and the poleplacement algorithm.

In figure 2 the estimated parameters are shown as a function of time. Notice, how the estimates of the plant parameters, except g_1, is moving in a wrong direction just after the change. However, the algorithm defeats the influence from the change in the load disturbance after approximately 300 steps.

The diagonal elements of P(t) are plottet versus time in figure 4. The diagonal element corresponding to g_1 has a special status due to the large control actions. The effects of the jump in the load disturbance at t=100 are easily recognized in figure 3 and 4, showing $\lambda(t)$ and P(t).

In figure 5 the loss function, $J_1(t)$, defined through:

$$(1-q^{-1})J_1(t) = [y(t) - w(t)]^2 \qquad\qquad (9)$$
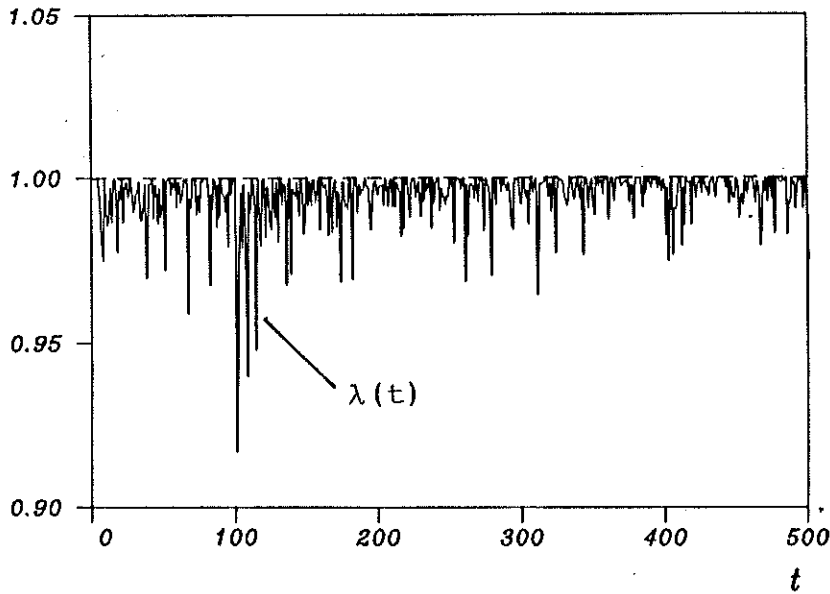
<u>Figure 3:</u> The variable forgetting factor, $\lambda(t)$. The parameters are constant except the load disturbance which has an abrupt change at t=100.

is shown. Notice the soft evolution of $J_1(t)$ after the change. This indicates that the algorithm slowly recovers form the change.

It is typical to methods which have their origin in the criterium (3), that the discounting only takes place in the direction determined by the old data, i.e. only $P^{-1}(t)$ in the recursion, (4), is reduced. This means in particular for constant forgetting factor, λ, that, if the determinant of the matrix:

$$R_N(t) = \sum_{i=1}^{N} \lambda^{i-1} \phi(t+1-i)\phi^T(t+1-i)$$

$$N \gg \dim(\theta)$$

is numerically small for large values of N, some elements in P(t) might become extremely large (for $\lambda < 1$). This situation occurs if the information is only obtained in
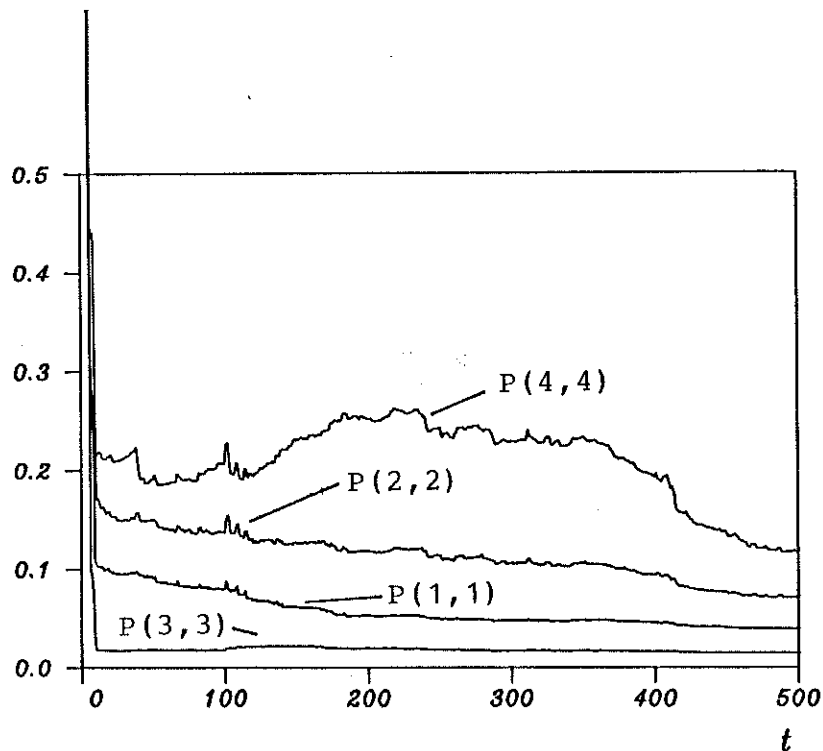
Figure 4: The diagonal elements in the P(t) - matrix
obtained in the simulation with plant 2 and the
poleplacement control algorithm.

certain directions, i.e. $\phi(t)$ is restricted to a subset of
the space. A small exitation in one particular direction then
might change the estimate rather inconveniently, because the
amplification in the feedback from the prediction error has
become extremely large. The phenomenon is often denoted as
covariance wind up (cf. Aström (1980)).

4.1.5 The Trace Algorithms.

There exist certain methods which deal with the problem from
an algorithmic point of view. Basically, the problem in an
ordinary least squares algorithm is that the $P_o(t)$ - matrix
and accordingly the amplification, $K_o(t)$, in (3.2:44) is
vanishing and it is consequently impossible to track the
parameter variations. The objective in these presented
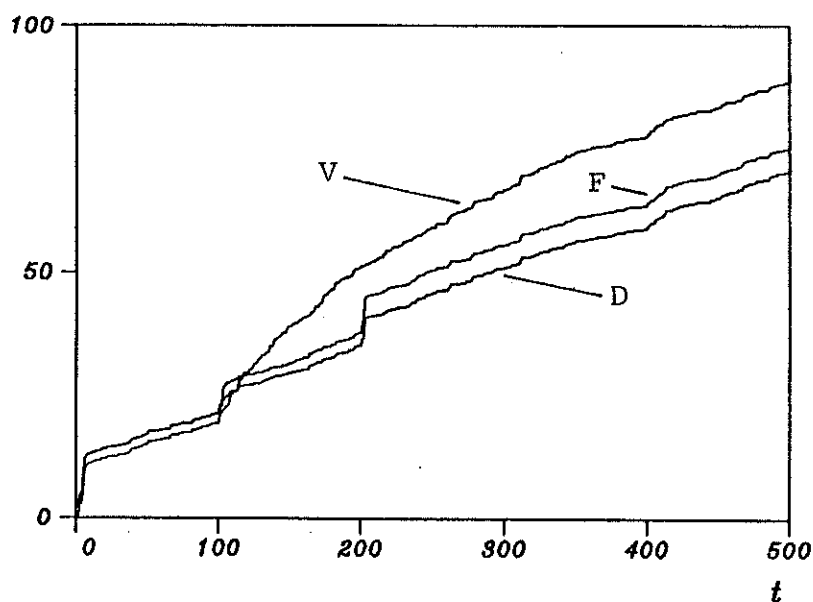
Figure 5: The loss function, $J_1(t)$, obtained in simulations with the variable forgetting method, (V). The curves, (D) and (F), are the loss functions obtained in simulations with the detector algorithm, (D), and with the filter bank algorithm, (F). (see example 13 and 14).

forgetting methods is to keep P(t) away from zero. The recursion for P(t), (4), is in these kinds of methods modified and becomes:

$$P^{-1}(t+1) = \lambda_1(t+1)P^{-1}(t) + \lambda_2(t+1)\phi(t+1)\phi^T(t+1) \qquad (10)$$

where the discounting is done both in the direction determined by old observations (for $\lambda_1 < 1$) and in the direction of incoming information, i.e. in the direction of $\phi(t+1)$ (for $\lambda_2 < 1$).

In a version of the constant trace algorithm which is proposed by Irving and mentioned in Landau (1981), the objective is to keep the trace of P(t) not less than a predetermined value, L_o, i.e to determine λ_1 and λ_2 in such a way that:

$$L(t+1) = tr \ P(t+1) \geqslant L_o \qquad\qquad (11)$$

The inequality is due to the fact that the value of $P(t)$ is rather high in the startup period. Introducing the quantity:

$$\gamma = \frac{\lambda_1(t+1)}{\lambda_2(t+1)}$$

Then the solution to (11) is simply (see appendix 7):

$$\lambda_1(t+1) = Min[\ \frac{1}{L_0} \ \{ \ L(t) - \frac{\phi^T(t+1)P^2(t)\phi(t+1)}{\gamma+\phi^T(t+1)P(t)\phi(t+1)} \ \} ; \ 1 \ ] \qquad (12)$$

where it is costumary to choose $\gamma=1$.

Another method, proposed in Luzano (1982), from this class is the bounded trace algorithm in which the objective is to keep the trace of the Hessian matrix [$P^{-1}(t)$] bounded, i.e. to keep:

$$T(t+1) = tr \ P^{-1}(t+1) < M \qquad (13)$$

This objective is cf. Appendix 7 accomplished by:

$$\lambda_1(t+1) = \frac{\gamma M \lambda_0}{\gamma M + \phi^T(t+1)\phi(t+1)} \qquad (14)$$

Here $0 < \lambda_0 < 1$ and M are tuning parameters. The two methods mentioned above are examples of techniques for gaining a certain property of the $P(t)$ - matrix. The forgetting factors are in contradiction to the variable forgetting method only determined as functions of $P(t)$ and the exitation direction, $\phi(t)$.

4.1.6 The Directional Forgetting Method.

A third method which objective concerns properties of $P(t)$ has been proposed in Hägglund (1983). The goal here is to obtain a $P(t)$ - matrix which is proportional to the unit matrix, i.e.

$$P(t) = a I \qquad (15)$$

where the scalar, a, is a tuning parameter. If this objective is reached, the Hessian matrix has equal eigenvalues, i.e the niveau curves in the loss function are circles. Opposite to the previous methods the discounting of information is only performed in the direction of the exitation:

$$P^{-1}(t+1) = P^{-1}(t) + [1-\alpha(t+1)] \phi(t+1)\phi^T(t+1) \qquad (16)$$

i.e. $\lambda_1 = 1$ and $\lambda_2 = 1 - \alpha(t+1)$. The objective, (15), is obtained by a nonlinear feedback from the error:

$$\delta(t+1) = \frac{\phi^T(t+1)P^3(t)\phi(t+1)}{\phi^T(t+1)P^2(t)\phi(t+1)} - a \qquad (17)$$

which is the difference between the Rayleigh quotient of $P(t)$ (evaluated in $P(t)\phi(t+1)$) and its desired value, a. In order to be more specific, introduce the quantities:

$$S_i(t+1) = \phi^T(t+1)P^i(t)\phi(t+1), \quad i = 1,2 \qquad (18)$$

Then the feedback is given by:

$$\tilde{\alpha}(t+1) = 1 + \frac{\delta(t+1)}{S_1(t+1)\delta(t+1) - S_2(t+1)} \qquad (19)$$

$$\alpha(t+1) = \begin{cases} 0 & \text{for} & \tilde{\alpha}(t+1) \leqslant 0 \\ \tilde{\alpha}(t+1) & \text{for} & 0 < \tilde{\alpha}(t+1) \leqslant S_1^{-1}t+1) \\ S_1^{-1}t+1) & \text{for} & S_1^{-1}t+1) < \tilde{\alpha}(t+1) \leqslant 1+S_1^{-1}t+1) \\ 0 & \text{for } 1+S_1^{-1}t+1) < \tilde{\alpha}(t+1) & (\ 20 \ ) \end{cases}$$

where the limit is chosen in order to fulfil demands on stability and in order to keep the P - matrix positively definite.

In Hägglund (1983) it is shown that the P(t) - matrix will not converge to aI, if the tuning parameter, a, is chosen too large, but then converge to a smaller matrix.

The objective, (15), separates this method from the other forgetting methods. In the exponential forgetting method all elements in the hessian matrix, $P^{-1}(t)$, are for $\lambda(t) < 1$ equally reduced. This means that old information gradually is discounted and the magnitude of the error amplification, $P(t)\phi(t)$, is prevented from vanishing. Now assume that an abrupt change has occured. The estimate, $\hat{\theta}(t)$, will then move in a direction determined by the gradient, $-\phi(t)\varepsilon(t)$, and by the the hessian matrix, $P^{-1}(t)$, which elements mainly are determined by old observations. Since the eigenvectors of $P^{-1}(t)$ reflects properties of the old system the estimates will probably move in a wrong direction just after the change. In the algorithm proposed by Hägglund the P(t) - matrix will under certain conditions approximately be proportional to a unit matrix, i.e. the algorithm asymptotically becomes a gradient algorithm. Since the estimate in such an algorithm is moving in the direction of the gradient, $-\phi(t)\varepsilon(t)$, it is likely that the estimate will move in the right direction just after an abrupt change.

On the other hand, forgetting methods are designed to estimate plant parameters which are almost constant. In the basic least squares case (constant plant parameters), the P(t) matrix among other things, indicates that a movement of the estimate

in certain directions will result in large prediction errors, i.e. in large losses. This accuracy information is discounted in the algorithm proposed by Hägglund and the algorithm becomes a gradient algorithm (see appendix 2) if the objective, (15), is gained.

Example 9: A simulation with plant 2 and directional forgetting.
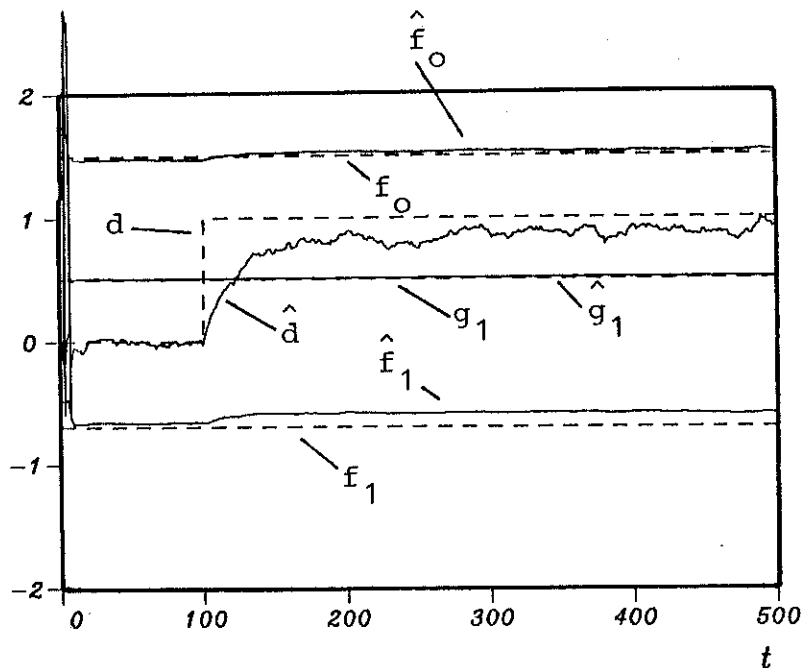
Figure 6: The estimated (shown solid) and the correct control parameters.

The estimates shown in figure 6 are obtained in a simulation with the directional forgetting method and poleplacement control of plant 2 (see example 2 and 5).

Only the load disturbance has an abrupt change. Notice, the estimates which are shown in figure 6 of the load disturbance is moving in the right direction just after the change and the estimate of the dynamic parameters are only slightly effected. If the loss function, $J_1(t)$, which is defined in (9) plottet in figure 7, is compared with the loss function obtained in a simulation with variable forgetting (see
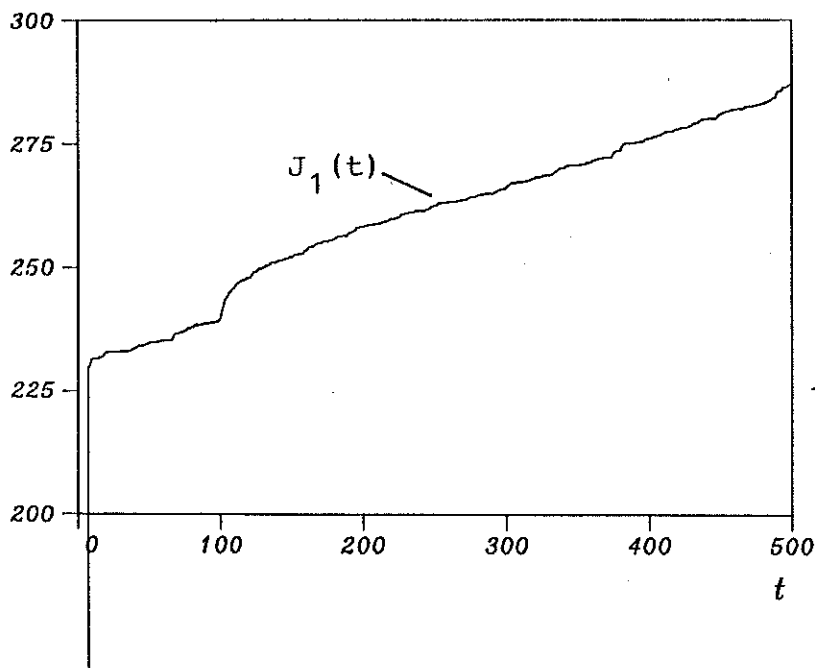
Figure 7: The loss function, $J_1(t)$, given by (9).

example 8) it is seen that the algorithm given here quickly recovers from the change.

The objective in the directional forgetting method is to obtain a $P(t)$ – matrix which is proportional to the unit matrix. The factor, a, which is the tuning parameter in the algorithm, has been chosen to be equal 0.05 in this simulation.

In figure 8 it is seen that only the diagonal element in $P(t)$ corresponding to the load disturbance, i.e. $P(4,4)$, reaches this goal. This is probably due to the large value for the tuning parameter, a. This large choice results especially in good tracking properties with respect to the load disturbance. However, the tracking properties with respect to the dynamic parameters are not as good as the choice of the tuning parameter indicates. The forgetting quantity, $\alpha(t)$, is shown in figure 9.
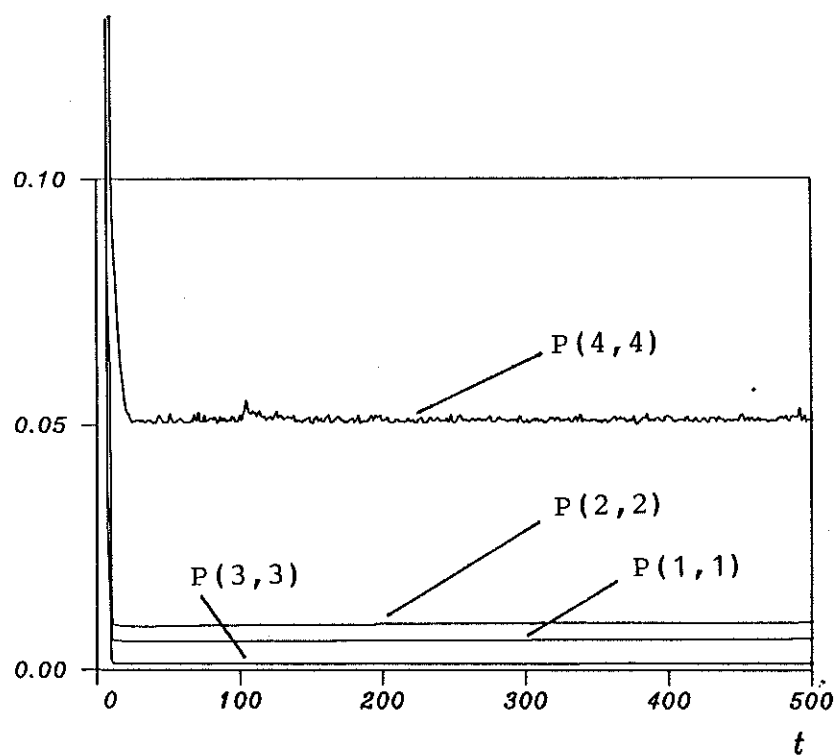
Figure 8: The diagonal elements in the P(t) - matrix.
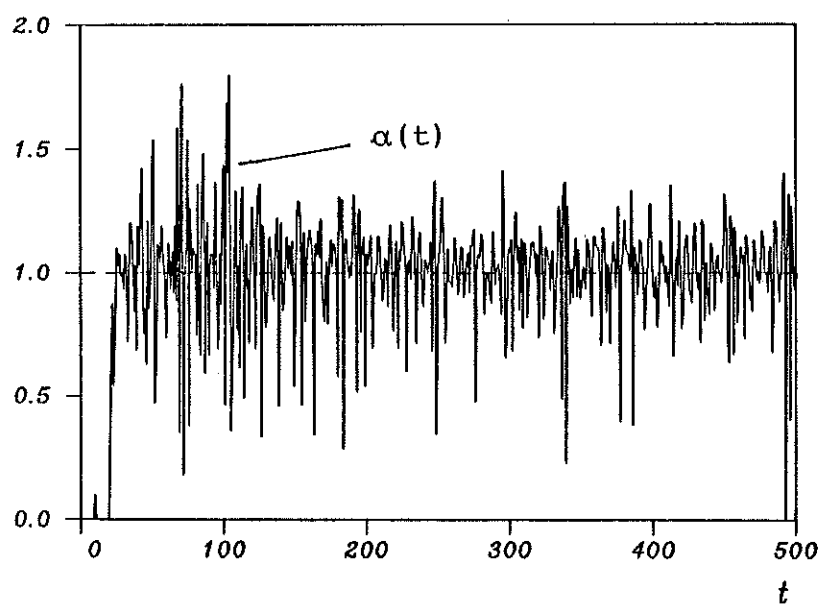
Figure 9: The forgetting quantity, α(t).

4.2 Model Estimators.

The methods which were described in the previous section are all based on the assumption that the plant parameters can be regarded as nearly constant within a suitable interval. If the parameter variations are more quick, this assumption must be abandoned. In this section, two methods designed to handle quick, but regular, parameter variations, will be described. Both methods include a model of the parameter variations.

4.2.1 Piecewise Approximations.

A method which is based on a deterministic model of the parameter variations has been proposed in Evans and Betz (1982). In this method the parameter variation is assumed to be piecewise linear. Hence, it is assumed that the plant is given by:

$$\theta(t+1) = \theta(t) + \omega_n$$

$$y(t) = \phi^T(t+1)\theta(t) + e(t) \qquad (17)$$

$$nN_o \leqslant t < (n+1)N_o$$

where ω_n is constant within the interval with length N_o. If the extended parameter vector, Θ, and the extended regressor vector, Φ, are introduced:

$$\Theta(t) = [\ \theta^T(nN_o),\ \omega_n^T\ ]^T$$

$$\qquad (18)$$

$$\Phi(t) = [\ \phi^T(t),\ (t-nN_o)\phi^T(t)\ ]^T$$

then the plant can be described with the regression equation:

$$y(t) = \Phi^T(t)\Theta(t) + e(t) \qquad (19)$$

within the mentioned interval. In the original paper the elements in the extended parameter vector, θ, were estimated with a modified recursive least square algorithm supplied with resetting in every intervals. Here it is assumed that the parameter vector $\theta(t)$ is constant within an interval. It is, however, possible to apply a forgetting method or another model estimator if $\theta(t)$ is time varying. It can be regarded as if the feature for handling the time variations has been introduced at a higher level.

This method can be further extended and in Xianya and Evans (1984) an algorithm is proposed in which the parameter variations are piecewise approximated by a finite order polynomial.

4.2.2 The Kalmanfilter Method.

Another method which is including a model of the parameter variation is the well known Kalman filter method (see e.g. Jazwinsky (1970) or Åström (1970)). This method is based on the stochastic model (2:9) and (2:10), i.e.

$$\theta(t+1) = \theta(t) + v_o(t+1)$$
$$\tag{20}$$
$$y(t) = \phi^T(t)\theta(t) + e(t)$$

where especially:

$$v_o(t) \; \varepsilon \; N(\; 0, \; R_o\sigma_o^2 \;)$$

Assuming that the matrix R_o is known, then the bayesian approach leads to the estimator given by (3.2:43) except for the recursions for $P(t)$ which in this case become:

$$P(t+1|t) = P(t) + R_o \tag{21}$$

$$P^{-1}(t+1) = P^{-1}(t+1|t) + \phi(t+1)\phi^T(t+1) \qquad (22)$$

Actually the recursion, (21), is the only algorithmic extension of the least squares estimator. Compared to exponential forgetting, which is only designed to handle slow parameter variations, this method not only increases the magnitude of the elements in $P(t)$, but also the principal directions. On the other hand, this more detailed model requires a better knowledge of the plant. Normally R_o is unknown and is predefined as a tuning parameter. If one element in R is chosen erroneously, then the prediction error becomes correlated for correct estimate. This means that the estimate might become biased (see Ljung (1977)). However, it is a fact that if the variations are regular then the method has a relative low sensivity with respect to the choice of R_o.

The Kalman method might, however, be supplied with a superstructure, which estimates the elements in R_o. Such a noise adaptive Kalmanfilter might be based on several approaches which originally are developed for the state estimation problem. Some of these methods are based on the fact that the prediction errors are white for correct value of R_o. Another class of methods are founded on covariance matching. One of these methods consists of a recursively minimizing of the criterium:

$$J(t,R) = \sum_{i=0}^{t} [\; \epsilon^2(i) - E\{(\epsilon^2(i)|R\} \;]^2$$

for a special structure in R_o (diagonal). Since it is impossible to estimate both σ_o^2 and the diagonal of R_o with this method, an a'priori knowledge of σ_o^2 is required. Such an assumption is in general unrealistic in connection with adaptive control.

Example 10: Simulation with the kalmanfilter algorithm.

In this simulation performed with the poleplacement controller
and plant 2, only the load disturbance is changing (at t=100).
In the design of the kalmanfilter it is assumed that only the
load disturbance might change and the corresponding element in
the R_O - matrix is used as a tuning parameter. In the
simualtions the matrix:

$$R_O = \left\{ \begin{array}{cccc} 0 & & & \\ & . & O & \\ & & . & \\ O & 0 & & \\ & & & 0.1 \end{array} \right\}$$

is used. In figure 10 the parameter estimates are shown. The
estimate of the load disturbance is tracking the true value
very well during the change but is rather dithering in the
normal situation.

The large value for the element in R_O which corresponds to
the load disturbance will, for constant and correct estimate,
give correlated prediction errors. Consequently the estimates
of the plant parameters which are constant might be biased.
The small difference between the estimate and the true value
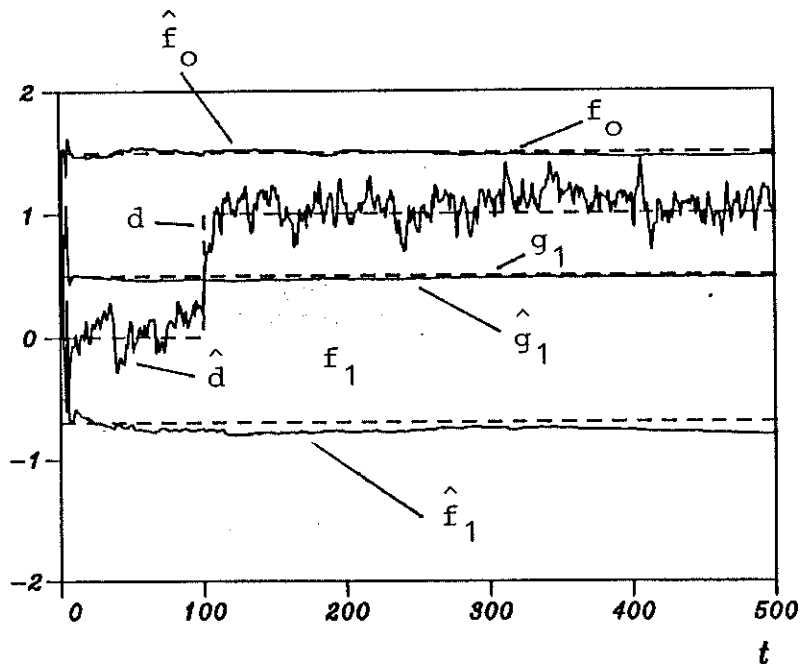for f_1 ($= - a_1$) might be bias.

Figure 10: The estimated control parameter obtained in a simulation with the poleplacement controller and plant 2. The identifcation part consists of a kalmanfilter.

4.3 Summary.

================================

This chapter has been devoted to methods for estimating parameters which have a regular variation.

The methods are divided up into two classes characterized by the type of parameter variation to which the method are designed.

The forgetting methods are based on the assumption that the parameter variation is so slow that the parameters can be regarded as constant within a reasonble horizon. These methods are founded on two basic ideas, either to discount old information or to keep the P(t) - matrix from decending toward zero. The amount of discounting or forgetting can be determined in different ways. It can be determined by the length of the horizon in which the plant parameters can be regarded as constant. The amount of forgetting can also be a function of the statistical properties of the prediction error. Finally, the forgetting is in several methods performed in such a manner that a certain property of the P(t) - matrix is obtained.

The model estimators are designed to estimate plant parameters which have a regular evolution. Since the design of the model estimators is based on a model of the parameter variations, it is possible to apply these methods even if the parameter variations are rather quick compared to the sampling interval. The model have, however, to be a reasonable approximation to the actual parameter variation.

The methods which are described in this chapter are individual algorithms, but can be used as building blocks in the detector algorithm which are to be described in the following chapters.

5

The Detector Algorithm.

The methods which have been described hitherto are designed to estimate the plant parameters when they have a regular evolution. If some of the plant parameters in this evolution have an abrupt change due to e.g. a failure, then the corresponding estimates will be erroneous after the change. Due to the feedback from the prediction error this change might result in errors in all the estimates. These errors are, however, decreasing at a tempo determined by the degree of discounting of old information. Then, one course to follow is to choose the tuning parameter(s) for the actual method in such a manner that the estimates are based on only a small number of observations. This will of course result in a large variance of the estimate. Since an abrupt change in the plant parameter might normally be a rare event, an estimator with a short horizon then will result in unnecessarely large losses. The dilemma occurs because of the difficulty in choosing the tuning parameter(s) in such a manner that the estimation horizon is large when the parameter evolution is regular, and that it is very short when the plant parameters change abruptly.

The algorithm which is to be described in this chapter is intended for estimation of plant parameters which occationally may perform abrupt changes. Such an algorithm can be derived in several ways. The derivation given in this thesis is based on a model of the parameter variations. Firstly, this design model is to be discussed. Secondly, there will be given an optimal solution to the identification problem when the variations of the plant parameters are given by the design model. The solution require, however, an exponentially growing bank of kalmanfilters. Since the implementation of such a bank is unrealistic there will be searched for an approximation which will result in the detector algorithm. The algorithm is presented as the identification part of an explicit self tuning controller, but is also applicable in connection to the implicit poleplacement controller and to other adaptive

problems.

5.1 The Designmodel.

This presentation takes its point of origin in a simple model of a plant in which the parameters might perform abrupt changes. The resulting algorithm is, however, not restricted to the described model. It might be used when the plant parameters normally perform a smooth evolution, but at certain points of time have abrupt changes or jumps.

It is, as in chapter 2, assumed that the time evolution in the plant parameters, $\theta(t)$, is given by the Markov process, (2:12), which satisfies the following simple process equation:

$$\theta(t+1) = \theta(t) + v(t+1) \qquad\qquad (1)$$

The output equation is still assumed to be the linear regression model, (2:9), i.e.:

$$y(t) = \phi^{T}(t)\theta(t) + e(t) \qquad\qquad (2)$$

Here the two white noise sequences, $v(t)$ and $e(t)$, are mutually independent. The output equation, (2), forms together with the process equation, (1), a state space model of the plant.

In chapter 2 the different kinds of parameter variations were discussed. In the model, (2:12), which describes abruptly changing parameters the different directions of parameter changes were divided into different classes or situations. In the normal situation, $S_{o}(t)$, the plant parameters are performing a regular evolution. The exceptional situation, $S_{3}(t)$, in which all parameters might change abruptly, is subdivided into two situations, $S_{1}(t)$ and $S_{2}(t)$. In the

situation, $S_1(t)$, only the load disturbance is abruptly changing and in $S_2(t)$ only some of the dynamic parameters might change abruptly.

There exist, as indicated in the proceeding sections, several methods for handling regular variations in the plant parameters. Since it is possible to include these methods in the final algorithm, it will, in the sequel, be assumed that the parameters are constant in the normal situation.

If, as temporarily assumed for the normal situation, $S_o(t)$, the process noise, $v(t)$, is identically to zero, the plant has constant parameters. If $v(t)$ on the other hand is a white noise vector sequence, the plant parameter will perform a random walk process according to the structure of the variance matrices, $R_i(t)$, $i=1,2,3$. The algorithm which is to be described in the sequel deals with adaptive control when the variance of the process noise, $v(t)$, is mainly equal to zero, but at certain instants of time takes large values, i.e. when the plant parameters perform an abrupt change. This can be treated as if the characteristics of the process noise are changing between the two extreme situations mentioned above.

Let $\phi_N(x;\mu,\sigma^2)$ be the gaussian density function with mean equal to μ and variance equal to σ^2. Assume that the process noise, $v(t)$, can be modelled as a stochastic variable which distribution is a compound of normal distributions, i.e. that the density function, $f(x)$, for $v(t)$ is given by:

$$f(x) = \sum_{i=0}^{3} p_i(t)\phi_N[x;0,R_i(t)\sigma_o^2] \qquad (3)$$

where $R_0(t)$ for constant parameters in the normal situation is a zero matrix and $R_i(t)$, $i=1,2,3$, are nonzero matrices. The variance matrices have been factorized in the innovation variance, σ_o^2, since this temperarily is assumed to be constant. Futhermore, the weights, $p_1(t)$, $p_2(t)$ and

$p_3(t)$ can be interpreted as probabililities for jumps in the plant parameters, corresponding to the situations, $S_1(t)$, $S_2(t)$ and $S_3(t)$, respectively. Since $f(x)$ is a density function:

$$\sum_{i=0}^{3} p_i = 1. \qquad\qquad (4)$$

Compared to the model given in chapter 2 the problem which is to be solved here is more clearcut due to the assumed constancy of the plant parameter in the normal situation. However, The resulting algorithm is easily extended in order to handle regular parameter variations in the normal situation.

5.2 The Optimal Solution.

The problem which is to be solved in this chapter is to determine the estimate, $\hat{\theta}(t)$, of the plant parameters, $\theta(t)$ in such a manner that the expected loss:

$$J(t) = E\{ [(\theta(t)-\hat{\theta}(t)]^T [(\theta(t)-\hat{\theta}(t)] \} \qquad (1)$$

is minimized. Here a quadratic criterium has been used but others might (cf. Jazwinski (1970)) have been used as well. The resulting method is based on the designmodel which was stated in the previous chapter.

Assume for a while that the time evolution of the statistics of the process noise, $v(t)$, is known. This implies in particular that the time variation of the matrices, $R_i(t)$, $i=1,2,3$, is known. Assume further that:

$$\theta(0) \ \epsilon \ N(\hat{\theta}(0), P(0)\sigma_o^2) \qquad (2)$$

The available information at the current time t is also here symbolized by Y_t. Since the conditional distribution, $p(\theta(t)|Y_t)$, of $\theta(t)$ embodies all the statistical information about $\theta(t)$ which is avaible in the observed data and in its initial conditions (here $\hat{\theta}(0)$, $P(0)$), the solution to the estimation problem simply is contained in the conditional distribution $p(\theta(t)|Y_t)$. The problem left is to determine the evolution of the conditional distribution and to determine in which way the estimate is obtained from the distribution. The details in the deviation given in the sequel of this chapter can be found in Andersson (1983). Under the assumptions given here and in the previous chapter, it can be shown that:

$$p(\theta(t)|Y_t) = g[\theta(t)]$$
$$= \sum_{i=0}^{m_t-1} \alpha_i(t)\ \phi_N[\theta(t);\hat{\theta}_i(t),P_i(t)\sigma_o^2] \qquad (3)$$

i.e. that the conditional distribution of $\theta(t)$ is a compound of gaussian distributions. The weights, $\alpha_i(t)$, can be interpreted as the probability for $\theta(t)|Y_t$ to be $N(\hat{\theta}_i(t),P_i(t)\sigma_o^2)$ distributed. The time evolution of the distribution is determined by the process equation (5.1:1), which contains the designmodel of the parameter variations. The predictive distribution is equal to:

$$p(\theta(t+1)|Y_t) = \int g[v]\ f[\theta(t+1)-v]\ dv$$
$$= \sum_{i=0}^{m_{t+1}-1} \alpha_i(t+1|t)\ \phi_N[\theta(t+1);\hat{\theta}_i(t+1|t),P_i(t+1|t)\sigma_o^2] \qquad (4)$$

where $f[x]$ is defined in (5.1:3) and where:

$$\hat{\theta}_k(t+1|t) = \hat{\theta}_i(t|t) = \hat{\theta}_i(t) \quad ; \quad k=0,\ \ldots\ ,m_{t+1}$$

$$P_k(t+1|t) = P_i(t|t) + R_j \quad ; \quad i=0,\ \ldots\ ,m_t \qquad (5)$$

$$\alpha_k(t+1|t) = \alpha_i(t)p_j \quad ; \quad j=0,\ \ldots\ ,3$$

Here i and j are independent indices and k=4i+j. This means that the number, m_t, of elements in the conditional distribution satisfies:

$$m_{t+1} = 4\ m_t \qquad (6)$$

The recursions in (5) state that each member of the a'priori distribution results in 4 new members in the predictive distribution. It is important to notice that the predictive distribution consists of a compound of gaussian distributions which number is exponentially growing.

The inference or the data updating is, as in the ordinary
least squares case, governed by the output equation, (5.1:2).
Using Bayes theorem it can be shown that:

$$p(\theta(t+1)|Y_{t+1}) = \frac{p(y(t+1)|\theta(t+1),Y_t)p(\theta(t+1)|Y_t)}{p(y(t+1)|Y_t)}$$

$$= \underline{C}\ p(y(t+1)|\theta(t+1),Y_t)p(\theta(t+1)|Y_t)$$

$$= \sum_{i=0}^{m_{t+1}-1} \alpha_i(t+1)\ \phi_N[\theta(t+1);\hat{\theta}_i(t+1),P_i(t+1)\sigma_o^2]$$

$(\ 7\ )$

where

$$\hat{\theta}_i(t+1) = \hat{\theta}_i(t+1|t) + P_i(t+1)\phi(t+1)\varepsilon_i(t+1)$$

$$\varepsilon_i(t+1) = y(t+1) - \phi^T(t+1)\hat{\theta}_i(t+1|t)$$

$$P_i^{-1}(t+1) = P_i^{-1}(t+1|t+1)$$

$$= P_i^{-1}(t+1|t) + \phi(t+1)\phi^T(t+1)$$

$(\ 8\ )$

$$\alpha_i(t+1) = \underline{C}\ \phi_N[y(t+1);\phi^T(t+1)\hat{\theta}_i(t+1),\beta_i(t+1)]\ \alpha_i(t+1|t)$$

$$\beta_i(t+1) = \{\ 1 + \phi^T(t+1)P_i(t+1|t)\phi(t+1)\ \}\ \sigma_o^2$$

Here $\underline{C}^{-1}$ is the likelihood of the predictive distribution.
$\underline{C}$ is independent of $\theta(t+1)$ and insures that:

$$\sum_{i=0}^{m_t-1}\alpha_i(t) = 1.$$

$(\ 9\ )$

Since also the a'posteriori distribution is a compound of
gaussian distributions, the assumption, (3), is verified by
induction.

The recursions in (5) and (8) can be interpreted as an

exponentially growing bank of kalmanfilters corresponding to the members of the conditional distribution, $p(\theta(t)|Y_t)$.
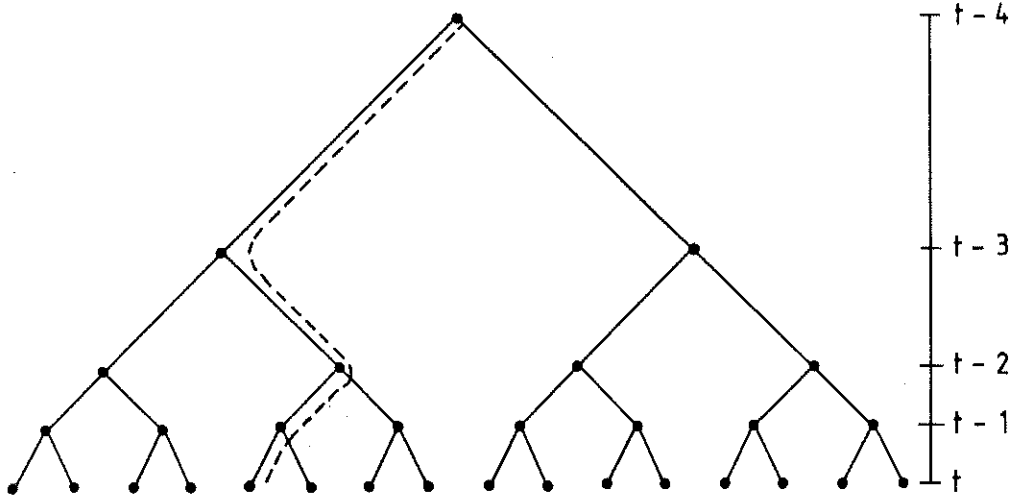
Figure 1: Tree showing the evolution of the conditional distribution. It is here assumed that $\theta(t-4)$ is gaussianly distributed. For the sake of simplicity it is also assumed that only one exceptional situation exists. The evolution of $\theta(s)|\Delta(s)$, $s \leqslant t$ is indicated dashed.

The weight, $\alpha_i(t)$, is the probability for the situation that the statistics of $\theta(t)$ are given by the i'th kalmanfilter in the bank. Notice in which way the probabilities are transformed when an observation is done. If the i'th kalmanfilter has a large prediction error, then the corresponding weight, $\alpha_i(t)$, is reduced because the predictive likelihood, $\alpha_i(t|t-1)$, is multiplied by the ratio between the likelihood of the i'th kalmanfilter and the likelihood of the whole bank, i.e. the ratio between:

$$\phi_N[y(t);\phi^T(t)\hat{\theta}_i(t),\beta_i(t)]$$

and $p(y(t)|Y_{t-1}) = \underline{c}^{-1}$.

Hence, based on the assumption that the time variation of $R_i(t)$ is known, it is possible to calculate the evolution of

the conditonal distribution for $\theta(t)$.

The optimal solution to the estimation problem can now be stated. It is known (cf. eg. Jazwinsky (1970)) that the conditonal expectation of $\theta(t)$ is the optimal solution to a large class of criteria including (1). This means that the optimal estimate is given by:

$$\hat{\theta}(t|t) = E\{ \theta(t)|Y_t \} = \sum_{i=0}^{m_t-1} \alpha_i(t)\hat{\theta}_i(t|t) \qquad (10)$$

i.e. the result $\hat{\theta}(t)$ is obtained by weighting together the results, $\hat{\theta}_i(t)$, from the different kalmanfilters in the bank. Consequently, the weights, $\alpha_i(t)$, can be interpreted as a selection rule.

Notice that the optimal solution consists of a bank of kalmanfilters which number unfortunately is exponentially increasing. Also notice that one specific kalmanfilter is designed for one specific realisation of the stochastic process, $\Delta(t)$, from (2:12). In figure 1 the evolution tree, corresponding to the the bank of kalmanfilters, is shown for the case in which the number of exceptional situations for simplicity is chosen equal to one.

Also notice that this optimal solution is based on the assumption that the variance matrices $R_i(t)$, $i=1,2,3$, are known. Consequently, the application of the optimal kalmanfilter bank algorithm is meets two different problems, namely the exponentially growing number of kalmanfilters and the unknown variance matrices. In the sequel the discussion will be focussed on approximative methods for treating these two problems.

5.3 The Detector Algorithm.

In this chapter the basic structure of the detector algorithm will be described. The algorithm can be regarded as an approximation to the optimal kalmanfilter bank algorithm which is based on the model, (5.1:1 and 2), for the parameter variations. This approximation requires that the normal situation, i.e. when the plant parameters are constant (or in general perform a regular evolution), is the most likely situation. If this assumption is not satisfied, then a model of the type (2:10) is more appropriate and a single kalmanfilter is then applicable. The detector algorithm might also be obtained via approaches which are not based on a model for the parameter variations. The only assumption is then that the plant parameters normally perform a regular evolution. Here the detector algorithm will be presented as an approximation to the optimal filter bank algorithm.

First focus on the problems caused by the exponentially increasing number of filters in the bank. It has been mentioned that $\alpha_i(t)$ can be interpreted as a selection rule, i.e. a rule which determines the weight with which a result, $\hat{\theta}_i(t)$, from one kalmanfilter enters into the final result, $\hat{\theta}(t)$. Of course the weight, $\alpha_i(t)$, reflects how close the estimate, $\hat{\theta}_i(t)$, is to the correct value, i.e. $\alpha_k(t)$ is approximately equal to one if the statement:

$$\theta(t) \ \varepsilon \ N(\hat{\theta}_k(t), P_k(t)\sigma_o^2)$$

is approximately correct.

Due to the basic assumptions the normal situation with constant parameters is the most likely situation. Then one specific kalmanfilter will entirely determine the resulting

estimate, $\hat{\theta}(t)$, except the case when the parameters are
changing. This means that all $\alpha_i(t)$ according to (5.2:8)
are nearly equal to zero except for the one which corresponds
to the kalmanfilter designed for the normal situation. More
precisely, only one kalmanfilter is actually needed when the
parameters are constant. Due to the exponentially growing bank
of kalmanfilters an approximation is needed. If the selection
rule is modified and is based on data which has its origin in
the kalmanfilter designed for the normal situation, then the
whole exponentially growing bank of filters does not have to
be implemented. Since the modified selection rule becomes a
test of the assumption on which the actual kalmanfilter is
designed it will, in the sequel, be denoted as the detector.
If the selection rule at time t detects a change in the
parameters and requires the results from an unimplemented
kalmanfilter, then the specific branch of the evolution tree
has to be realized. This is done by starting a kalmanfilter in
that point of time, $T_c < t$, at which the change has occured.
Consequently data has to be stored because the kalmanfilter
which is designed to handle the actual exceptional situation,
performs the branch estimation on late observations.

Hence, the detector has two tasks, firstly, to test if the
parameters are constant (or has an evolution corresponding to
the normal situation) and secondly to estimate the time at
which the change might has occured.

The second problem connected to the optimal solution is that
the matrices, $R_i(t)$, i=1,2,3, are unknown. The elements in
these variance matrices could be determined a'priori and be
treated as tuning parameters. However, in the algorithm
presented here the elements in the variance matrices are
estimated in a noise adaptive kalmanfilter which will be
described in chapter 5.5. If the detector only is able to
determine if a change has occured then the noise adaptive
kalmanfilter both has to estimate which parameters have
changed and to estimate the variance of the actual parameters.

This is a possible course to follow, but a better performance can be achieved if the detector is able to perform a discrimination between the different directions of abrupt changes. Since the load disturbance has a special status, i.e. it is likely that this parameter might change more frequently than the others, it is convenient if the detector can discriminate between the situations, $S_1(t)$, $S_2(t)$ and $S_3(t)$. Hence, the detector has a third task, namely to discriminate between the three exceptional situations.

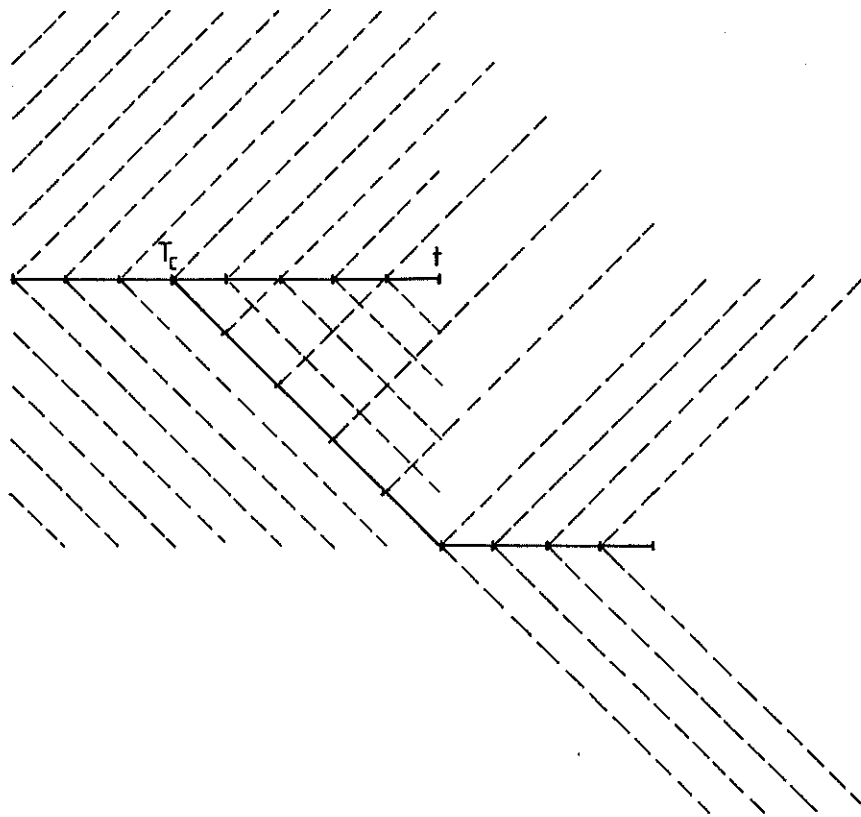Figure 1: The evolution tree showing the mechanism in the detector algorithm when an abrupt change at T_c has been detected (at t). The figure is shown for two exceptional situations (sloping lines). The normal situation is indicated by a horizontal line and the unimplemented branches are shown dashed.

The mechanism in the algorithm is as follows. In the normal mode, i.e. when the parameters are assumed to be constant (or

have a regular evolution), only the kalmanfilter or the estimator designed to handle this situation is active. Simultaneously a detector tests the assumptions on which the kalmanfilter is designed, i.e. if the parameters are constant (or have a regular evolution). If the detector alarms, the mode of the estimator is changed in accordance to the detected exceptional situation. Then the estimator is started in that point of time which is assigned to the parameter change by the detector. The exceptional modes consist in this thesis of noise adaptive kalmanfilters, each designed to handle one specific situation. The actual noise adaptive kalmanfilter has to perform the branch estimation on stored data and has to track the change in the parameters. This is illustrated in figure 2 where the number of exceptional situations for the sake of simplicity is equal to two.

In general, it can be regarded as if the detector algorithm consists of an estimator which has several modes controlled by a detector. In the version presented here, the estimator is in each operation mode a kalmanfilter designed for the corresponding situation. It is, however, possible to use other methods for estimating the parameters in the different situations. Also the design of the detector can be founded on several approaches different from those described in the following chapters. A survey of design methods for failure detection can be found in Willsky (1976). This survey is, however, restricted to the state estimation problem. Other approaches can be found in Basseville and Benveniste (1982). Futhermore, the main part of the estimation of the parameter change, i.e. the branch estimation, is in this version performed recursively on stored data. Alternatively, this tracking might, as in a method which is described later, be performed on line. When a change has been detected then nothing exceptional is carried out except for a change in the operation mode.

5.4 The Detector.

In this chapter we will deal with the detector problem and the
goal is according to the previous chapter a detector which
fulfil the following 3 demands:

1) To alarm one of the exceptional modes, if the plant
parameters have changed. This has to be done with
respect to both a quick and to a secure detection.

2) To discriminate between the different situations,
$S_1 - S_3$, if a change has occured.

3) To estimate the point of time, T_c, at which the
change has occured.

A demand of practical importance is that the detector
procedure should be simple and not require too heavy
computations.

Assume temporarily that the innovation, e(t), in the output
equation:

$$y(t) = \phi^T(t)\theta(t) + e(t) \qquad\qquad (1)$$

is a white $N(0,\sigma_o^2)$ noise sequence. Later we will discuss some
methods designed to handle non white noise sequences and some
methods designed to handle some kinds of timevariations in the
innovation variance. In Poulsen (1982) and in Poulsen and
Holst (1982) robust detectors are used. These detectors are
designed to handle deviations between the datagenerating
distribution and the assumed gaussian distribution of e(t).
Such detectors are especially robust with respect to outliers
or gross errors.

As mentioned in the previous chapters, the detection is based on data from the kalmanfilter which is designed to handle the normal situation, S_o. In order to simplify the representation, assume that the parameter estimate, $\hat{\theta}_o(t)$, is nearly correct and that the estimate of the variance matrix $P_o(t)$ is correspondingly small. This implies that a change in the plant parameters only effects the estimates slightly, if the algorithm is fixed in its normal mode. Later the implications of this assumption will be discussed.

The residual, $\varepsilon_r(t)$, which is related to the prediction error, $\varepsilon_o(t)$, in the following manner

$$\varepsilon_r(t+1) = \frac{\varepsilon_o(t+1)}{1 + \phi^T(t+1)P_o(t+1|t)\phi(t+1)} \qquad (2)$$

is by definition equal to the variation in the output signal, $y(t)$, which is not included in the model, i.e.:

$$y(t) = \phi^T(t)\hat{\theta}_o(t) + \varepsilon_r(t) \qquad (3)$$

Since $\varepsilon_r(t)$ contains that information which is not included in the model, it is natural to build the detection upon the statistical properties of this signal.

In S_o, i.e. when the hypothesis H_o is true, the prediction error $\varepsilon_o(t)$ is a white noise signal and

$$\varepsilon_o(t+1)|Y_t \quad \varepsilon \quad N(0,\sigma_o^2[1+\phi^T(t+1)P_o(t+1|t)\phi(t+1)]) \qquad : H_o$$

where the condition is on the information, Y_t, which is available at time t. It might be possible to investigate the statistical properties of $\varepsilon_r(t)|Y_{t-1}$ in the three exceptional situations and base the design of the detector upon the results of this investigation.

However, here another course will be taken and the detection will not involve Y_t directly. Basicly, the aim of the detector is to investigate the estimation error, $\tilde{\theta}(t)$:

$$\tilde{\theta}(t) = \theta(t) - \hat{\theta}(t) \qquad\qquad (4)$$

through the statistical properties of residual signal, $\varepsilon_r(t)$.

The residual, $\varepsilon_r(t)$, and the estimation error, $\tilde{\theta}(t)$, is related through:

$$\varepsilon_r(t) = \phi^T(t)\tilde{\theta}(t) + e(t) \qquad\qquad (5)$$

where the statistical properties of $\phi(t)$ are dependent on the evolution of the plant dynamic and on the applied control sequence. Let $\Theta(t)$ denote the unknown sequence of plant parameters:

$$\theta(s), \ s < t$$

and the known sequence of estimate:

$$\hat{\theta}(s), \ s < t$$

Additionally, let $\underline{\tilde{\theta}}(t)$ denote the information contained in the estimation error, $\tilde{\theta}(t)$, and in $\Theta(t)$. Consequently, the detection will be based on the properties of $\varepsilon_r(t)|\underline{\tilde{\theta}}(t)$.

Conditioned on $\Theta(t)$, the regressors in $\phi(t)$ are gaussianly distributed, i.e.:

$$\phi(t)|\Theta(t) \ \varepsilon \ N[\ \underline{m}(t), \ G(t) \ ] \qquad\qquad (6)$$

where:

$$\underline{m}(t) = E\{ \ \phi(t) \ | \ \Theta(t) \ \} \qquad\qquad (7)$$

$$G(t) = E\{ [\phi(t) - \underline{m}(t)][\phi(t) - \underline{m}(t)]^{T} \mid \theta(t) \} \qquad (8)$$

Then, equation (5) and (7) implies that:

$$\varepsilon_{r}(t) \mid \tilde{\underline{\theta}}(t) \in N[\mu(t), \sigma^{2}(t)] \qquad\qquad (9)$$

where:

$$\mu(t) = \underline{m}^{T}(t)\tilde{\theta}(t)$$

$$\qquad\qquad (10)$$

$$\sigma^{2}(t) = \tilde{\theta}^{T}(t)G(t)\tilde{\theta}(t) + \sigma_{o}^{2}$$

Due to the dependence of $\theta(s)$ and $\hat{\theta}(s)$ for s<t in $\phi(t)$ it is in general not an easy task to calculate all elements in G(t) and $\underline{m}(t)$. However, since the regressor which corresponds to the load disturbance is equal to 1, it is easy to determine some of the elements in G(t) and $\underline{m}(t)$. With the definition of $\phi(t)$ given in (2:7) the results:

$$G(t) = \left\{ \begin{array}{ccccc} \times & \ldots & \times & \mid & 0 \\ . & & . & \mid & . \\ . & & . & \mid & . \\ \times & \ldots & \times & \mid & 0 \\ \text{------} & & & & \\ 0 & \ldots & 0 & \mid & 0 \end{array} \right\} \underline{m}^{T}(t) = (\times \ldots \times \mid 1) \quad (11)$$

are easily obtained. Here × denotes an element which is not necessarily equal to zero. Based on this knowledge of the structure of G(t) and $\underline{m}(t)$ it is due to (8) and (9) possible to connect the different situations concerning the behaviour of the plant parameters with statistical properties of $\varepsilon_{r}(t)$ and base the discrimination upon this.

Under the hypothesis, H_{o}, it is as mentioned before assumed that $\tilde{\theta}(t) = 0$ and consequently that:

$$\varepsilon_r(t)|\tilde{\underline{\theta}}(t) \quad \varepsilon \quad N(0,\sigma_o^2) \qquad\qquad : \qquad\qquad H_o$$

Now assume that only the load disturbance is changing, i.e. situation $S_1(T_c)$, then the residuals have a non zero mean value but unchanged variance, i.e.

$$\varepsilon_r(t)|\tilde{\underline{\theta}}(t) \quad \varepsilon \quad N(\mu(t),\sigma_o^2) \qquad\qquad : \qquad\qquad H_1(T_c)$$

where $\mu(t) \neq 0$. A change in the load disturbance will also effect the estimates of the dynamic parameters, unless the $P_o(t)$ matrix is equal to zero. We neglect this problem since $P_o(t)$ quickly decreases to a small value if, as assumed, the system is properly excited.

Assume next that only the dynamic parameters are changing, i.e. situation $S_2(T_c)$. Then the variance of $\varepsilon_r(t)$ will increase because only the correct estimates give minimum residual variance, i.e.

$$\varepsilon_r(t)|\tilde{\underline{\theta}}(t) \quad \varepsilon \quad N(0,\sigma^2(t)) \qquad\qquad : \qquad\qquad H_2(T_c)$$

where $\sigma(t) > \sigma_o$. Depending on the nature of the reference signal and the direction of the change, also the mean value of the residual, $\varepsilon_r(t)$, might change. If, however, the mean value is equal to zero we conclude that only the dynamic parameters have changed.

For a total change, $S_3(T_c)$, we have that:

$$\varepsilon_r(t)|\tilde{\underline{\theta}}(t) \quad \varepsilon \quad N(\mu(t),\sigma^2(t)) \qquad\qquad : \qquad\qquad H_3(T_c)$$

i.e. that both the mean and the variance have changed. Notice that we might classify a large part of $S_2(T_c)$ as total change, $S_3(T_c)$. This is one of the prices which have to be paid, for the use of a detector algorithm instead of the optimal but unrealistic kalmanfilter bank algorithm.

In all the detectors which will be proposed in this chapter a knowledge of the innovation variance, σ_o^2, is needed. Such an information is rarely at hand and we are consequently forced to estimate it. In chapter 5.6 a recursive variance estimator will be described. Such an estimator must be designed in such a manner, that a change in the plant parameters do not significantly effects the variance estimate, since a false increase in this estimate would make a detector fault more likely. In chapter 5.6 some methods for handling this problem will also be discussed.

In the sequel 3 kinds of detectors are to be designed. Firstly, some simple detectors will be described. The design of these does not include the second demand, i.e. the detectors do not perform a discrimination between the three exceptional situations concerning the behaviour of the plant parameters. Secondly, a maximum likelihood (ML) detector, which on a fixed number of observations is able to discriminate between the four situations, is to be derived. Finally, a detector based on the cusum techniques is derived.

In the detectors which are to be described and presented in the sequel a test based on the N latest independent residuals is included. Due to the instationarity in $\varepsilon_r(t)$ the design of the detector is carried out on a related but more generic problem. Accordingly, consider the problem of testing if the N independent and identically distributed stochastic variables, ξ_i, are generated under one of the 4 situations. The stochastic variables are:

$$\xi_i \ \varepsilon \ N(\mu,\sigma^2)$$

$$(\mu,\sigma) \ \varepsilon \ \Omega = \{ \ \mu,\sigma \ | \ \sigma \geqslant \sigma_o \ \}$$

(12)

Let furthermore Ω_i , $i \ \varepsilon \ \{0,1,2,3\}$, be the subsets of Ω which correspond to the 4 situations, i.e.

$$\Omega_0 = \{ \mu,\sigma \mid \mu=0, \ \sigma=\sigma_0 \}$$
$$\Omega_1 = \{ \mu,\sigma \mid \mu\neq0, \ \sigma=\sigma_0 \}$$
$$\Omega_2 = \{ \mu,\sigma \mid \mu=0, \ \sigma>\sigma_0 \} \qquad (13)$$
$$\Omega_3 = \{ \mu,\sigma \mid \mu\neq0, \ \sigma>\sigma_0 \}$$

Let additionally $\underline{\xi}$ denote the vector:

$$\underline{\xi} = (\ \xi_1, \ \ldots \ , \xi_n \ )^T \qquad (14)$$

Due to the independence assumption the joint density function, $f(\underline{\xi};\mu,\sigma^2)$, satisfies:

$$\log(\ f(\underline{\xi};\mu,\sigma^2) = \sum_{i=1}^{N} \log \ \phi_N[\xi_i;\mu,\sigma^2] \qquad (15)$$

The difference between the statistical properties of ξ_i and $\varepsilon_r(t)|\tilde{\underline{\theta}}(t)$ is that these properties are constant for ξ_i, which not necessarrily are the case for $\varepsilon_r(t)|\tilde{\underline{\theta}}(t)$. The detectors are designed for ξ_i but the detection itself is based on $\varepsilon_r(t)$.

5.4.1 A Mean Value Detector - 1.

If it is known that only the load disturbance might change in an abrupt manner and if the slow variations in the dynamic plant parameters are handled by other methods, it is possible to apply a simple maximum likelihood ratio test in the mean value of the latest N residuals. This test is based on a constant innovation variance. Assuming that the time of change, T_c, is equal to t-N, then the goal is a test in the hypothesis:

$$H_0(t,N) : \varepsilon_r(T)|\tilde{\underline{\theta}}(T) \ \varepsilon \ N(0,\sigma_0^2) \qquad t-N < T \leqslant t$$

against the alternative:

$$H_1(t,N) : \varepsilon_r(T)|\tilde{\underline{\theta}}(T) \ \varepsilon \ N(\mu(t),\sigma_0^2) \qquad t-N < T \leqslant t$$

where $\mu(t) \neq 0$. Using the maximum likelihood ratio (MLR) test principle on the generic problem, the hypothesis, H_0, is accepted if the inequality:

$$L_0(\underline{\xi}) - \tilde{L}_1(\underline{\xi}) > \underline{K}_1 \qquad\qquad (16)$$

holds. Here the definitions:

$$L_0(\underline{\xi}) = \log[\sup_{(\mu,\sigma) \epsilon \Omega_0} f(\underline{\xi},\mu,\sigma^2)] \qquad\qquad (17)$$

and:

$$\tilde{L}_1(\underline{\xi}) = \log[\sup_{(\mu,\sigma) \epsilon \tilde{\Omega}_1} f(\underline{\xi},\mu,\sigma^2)] \qquad\qquad (18)$$

$$\tilde{\Omega}_1 = \Omega_0 \ \mathrm{U} \ \Omega_1$$

are used and $\underline{K}_1$ is a design variable. Notice, that the density function:

$$f_i(\underline{\xi}) = \sup_{(\mu,\sigma) \epsilon \tilde{\Omega}_i} f(\underline{\xi};\mu,\sigma) \qquad\qquad (19)$$

which enters in L_i can be interpreted as a principal element from the set of density functions which corresponds to Ω_i. This principal element is, in this test principle, chosen as the density function which describes the observed data at the largest likelihood. Direct calculations show (see eg. appendix 4 for details) that the inequality (16) is equivalent to:

$$\underline{M}^2 < \frac{\nu^2}{N} \qquad\qquad (20)$$

where:

$$\underline{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\sigma_o} \qquad\qquad (21)$$

Under the hypothesis, H_o, the normalized sum, $\underline{M}$, is gaussian. Hence, the tuning parameter, ν ($\nu^2 = 2K_1$), can be chosen equal to:

$$\nu = u_{1-\alpha/2} \qquad\qquad (22)$$

where α is the probability for erroneous rejection of H_o and u_x is a gaussian quantile. The resulting detector, MLRm, consists accordingly only of one signal, $M(t)$, given by the recursion:

$$(1-q^{-1})M(t) = (1-q^{-N})\varepsilon_r(t) \qquad\qquad (23)$$

and the inequality:

$$M^2(t) < N \sigma_o^2 \nu \qquad\qquad (24)$$

If this inequality holds then the load disturbance can be assumed to be unchanged. If the inequality fails to hold, then the estimate of the time, T_c, at which the change has occured is simply t-N. Normally this estimate will be less than T_c, which means that a change is detected quicker than indicated by the length of the horizon, N.

5.4.2 A Mean Value Detector - 2.

The detector which was described in the previous section bases its detection only on a fixed number of observations. If stated a little different, the detector does not involve a search for T_c. In this chapter a Cusum detector will be described. This Cusum detector is a special mechanization of a sequential test and consequently involve a search for T_c.

One possible method is at each instant of time, t, to evaluate the inequality (16) for a successively increasing number,N, of observations. Assume that the inequality, (16), holds for the N-1 latest residuals. For a given N there exist two possibilities. If the inequality fails to hold, then the change has occured at t-N. If, on the other hand, the inequality holds then the number of (late) residuals is increased and the inequality is evaluated again. This method gives a better performance with respect to the first demand, i.e. to alarm if a change has occured. If this procedure is based on the MLR test principle then it will be too labourious. A Cusum detector which is based on the probability ratio (PR) test principle utilizes the sequential structure and is very easy to implement.

Let us again turn to the generic problem. Assume that $H_0(t,N-1)$ has been accepted and $H_0(t,N)$ is going to be tested against $H_1(t,N)$. The design of the detector is based on the two following tests between the simple hypothesis:

$$H_0: \quad \xi_i \ \varepsilon \ N(0,\sigma_0^2), \qquad\qquad i=1, \ \dots \ , \ N$$

against the simple alternatives:

$$H_1^+: \quad \xi_i \ \varepsilon \ N(\gamma\sigma_0,\sigma_0^2)$$
and:
$$H_1^-: \quad \xi_i \ \varepsilon \ N(-\gamma\sigma_0,\sigma_0^2)$$

The principal elements from the sets of density functions, f_0, f_1^+ and f_1^-, are chosen a'priori and γ becomes a tuning parameter. Defining the quatities:

$$L_0(\underline{\xi}) = \log[\ f(\underline{\xi};0,\sigma_0^2) \ ]$$
and:
$$L_1^+(\underline{\xi}) = \log[\ f(\underline{\xi};\gamma\sigma_0,\sigma_0^2) \ ]$$

(25)

then the hypothesis, H_o, is accepted and H_1^+ is rejected
for the test based on N observations (N latest residuals),
if the inequality:

$$L_o(\underline{\xi}) - L_1^+(\underline{\xi}) > \log(B) \qquad (26)$$

holds. The quantity, B, which also is a tuning parameter in
the test, has an interpretation as the expected length between
erroneous rejections of H_o. This inequality is (c.f.
Appendix 4) eqvivalent with:

$$\underline{\underline{M}} < \frac{\log(B)}{\gamma} + \frac{\gamma}{2} N = \underline{A} + \underline{B} N \qquad (27)$$

where:

$$\underline{\underline{M}} = N \underline{M} = \sum_{i=1}^{N} \frac{\xi_i}{\sigma_o} \qquad (28)$$

An analog set of inequalities exist for detection of change in
the negative direction.

A study in the mechanisation of a sequential evaluation of the
inequalities (see e.g. Kemp (1962) or Hinkley (1971)) shows
that the detector, PRm, simply is given by the two set of
recursions:

$$M_+(t+1) = \text{Max}[\ M_+(t) + \frac{\varepsilon_r(t+1)}{\sigma_o} - \frac{\gamma}{2} \ , \ 0 \ ]$$

$$N_+(t+1) = [\ N_+(t)+1 \ ] \ \text{sign}(M_+(t+1)) \qquad (29)$$
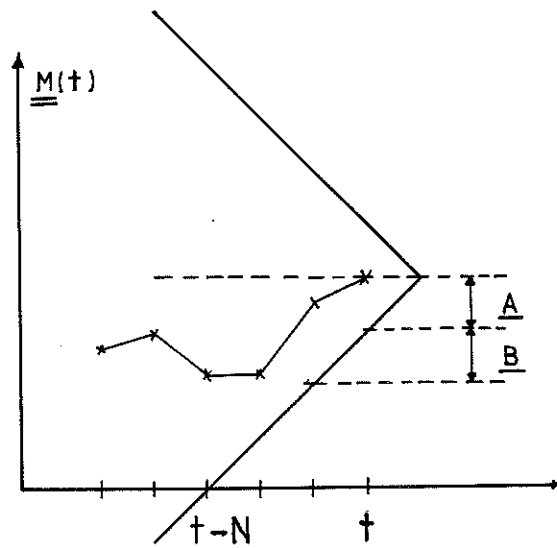
$$M_+(t) < \frac{\log(B)}{\gamma}$$

and:

Figure 1: The accumulated sum, $\underline{\underline{M}}(t)$, should be inside
the mask given by (18) and the anolog inequality.

$$M_{-}(t+1) = Max[\ M_{-}(t) - \frac{\varepsilon_r(t+1)}{\sigma_0} - \frac{\gamma}{2}\ ,\ 0\ ]$$

$$N_{-}(t+1) = [\ N_{-}(t)+1\ ]\ sign(M_{-}(t+1)) \qquad\qquad (\ 30\ )$$

$$M_{-}(t) < \frac{log(B)}{\gamma}$$

Here sign(0) = 0. The states in the recursions, (29) and
(30), i.e. the counters, $N_{+}(t)$ and $N_{-}(t)$, and the reduced
accumulated sums, $M_{+}(t)$ and $M_{-}(t)$, are initilized in zero:

$$M_{+}(0) = M_{-}(0) = 0$$

$$N_{+}(0) = N_{-}(0) = 0$$

If both inequalities hold then the load disturbance is assumed
to be constant. If a change has been detected, i.e. one of the
inequalities in (29) and (30) are violated, then the
estimate of T_c is equal to $t-N_{-}(t)$ or $t-N_{+}(t)$. If a change

has been detected in the positive direction at $\underline{t}$ the
corresponding states are zeroed, i.e.

$$N_+(\underline{t}) = M_+(\underline{t}) = 0$$

A suitable choice of the tuning parameters is:

$$B = 1000 \text{ and } \gamma = 1.$$

In the detector algorithm in which the estimation of the
change is performed on stored data there is a practical limit
for the point in which the exceptional mode is started. Due to
the finite storage there must be introduced a N_{max} which
limits the length of the interval for exceptional estimation.
In the shown simulations the value:

$$N_{max} = 20$$

has been used. If a change has been detected and N_+ (or
N_-) is larger than N_{max} then the exceptional mode is
started in $t-N_{max}$. In such a situation it is likely that
the change only affects the statistical properties of the
prediction error insignificantly.

5.4.3 A Simple Mean Value and Variance Detector.

In this chapter a detector which is based on a test for
increasing variance will be described. The detector is
designed directly for detection of changes in the dynamical
parameters. It is, however, also able to detect changes in the
mean value. Since the situations, $S_1(t)$ and $S_2(t)$, can be
regarded as special cases of $S_3(t)$ the detector can be used
in the general case for detecting a deviation from the normal
situation, i.e. for detection of an abrupt parameter change.
The discrimination between the exceptional situations then has
to be performed entirely as an estimation problem in the
exceptional mode (see chapter 5.5).

When the tracking of the parameter change is completely handled by the exceptional mode, the aim of the detection is to find out if the hypothesis:

$$H_o(t,N) : \varepsilon_r(T)|\tilde{\underline{\theta}}(T) \; \varepsilon \; N(0,\sigma_o^2) \quad t-N < T \leqslant t$$

fails to hold. The design, which is carried out on the generic problem is based on a MLR test in the hypothesis:

$$H_o: \quad \xi_i \; \varepsilon \; N(0,\sigma_o^2)$$

against the alternative:

$$H_2: \quad \xi_i \; \varepsilon \; N(0,\sigma^2)$$

where $\sigma > \sigma_o$. Using the maximum likelihood ration test principle (see appendix 4) the hypothesis, H_o, is accepted if the normalized sum:

$$\underline{Z} = \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i^2}{\sigma_o^2} \quad\quad\quad (\; 31 \;)$$

satisfies the inequality:

$$\underline{Z} < \frac{\beta}{N} \quad\quad\quad (\; 32 \;)$$

Under H_o, the normalized sum of squares is $\chi^2(N)$ distributed and the tuning parameter, β, may then be chosen as a quantile in this distribution, i.e.:

$$\beta = \chi^2(N)_{1-\alpha} \quad\quad\quad (\; 33 \;)$$

A suitable choise for $N = 6$ is $\beta = 22.5$ which corresponds to $\alpha = 0.001$.

Assuming that T_c = t-N then the simple detector, MLRv, consists of the signal:

$$(1-q^{-1})Z(t) = (1-q^{-N})\varepsilon_r^2(t) \qquad (34)$$

and the inequality:

$$Z(t) < \sigma_o^2 \beta \qquad (35)$$

i.e. we accept H_o, the normal situation, if the inequality, (35), holds.

The Cusum detector, PRv, which correspopnds to MLRv, consist of the signal given by the recursions.

$$Z(t+1) = Max[\ Z(t) + \frac{\varepsilon_r^2(t+1)}{\sigma_o^2} - \frac{\kappa}{\kappa-1}\ \log(\kappa)\ ,\ 0\ ] \qquad (36)$$

$$N_z(t+1) = [\ N_z(t)+1\ ]\ sign(Z(t+1))$$

and the inequality:

$$Z(t) < 2\ \frac{\kappa}{\kappa-1}\ \log(B) \qquad (37)$$

where the quantity, B, again can be interpreted as the expected length between erroneous rejections of S_o. Notice, that this detector may be used for sequential testing of the hypothesis:

$$H_o: \qquad \varepsilon_r(t)|\tilde{\underline{\theta}}(t)\ \varepsilon\ N(0,\sigma_o^2)$$

against all alternatives, i.e. against change in the mean of $\varepsilon_r(t)$, against change in the variance in $\varepsilon_r(t)$ and against change in both. However, the design of this detector (see appendix 4) is based on a sequential probability ratio test

between the simple hypothesis:

$$\xi_i \; \varepsilon \; N(0, \sigma_0^2)$$

against the simple alternative:

$$\xi \; \varepsilon \; N(0, \kappa \sigma_0^2)$$

where κ is a design parameter.

5.4.4 Discriminators.

In the previous chapter some useful detectors were discussed. Their main disadvange is that they do not include a discrimination between the 4 situations which exist in this application. An algorithm based on these detectors must then heavily rely on the abilities of the exceptional mode, or require that the changes might only take place in one particular direction. However, in the next three sections two new detectors will be investigated. These two detectors are also able to fulfill the second demand, i.e. to discriminate between the different situations and in this section their basic design will be given. Both the ML and the Cusum detector are based upon this theory. The approach which is applied here takes its origin in the discrimination analysis and this section ends with a discussion of the result and the way in which it is is related to other approaches.

Both the design of a fixed horizon detector, the ML detector, and a sequential detector, the Cusum detector, involve a test or a discimination which bases the decision upon the N latest residuals. Consequently focus on the generic problem of discriminating between the 4 situations when N independent identically distributed stochastic variable, ξ_i are observed.

The first step toward the discriminator is to choose a

suitable test principle.

Introduce the nonnegative lossfunction, $A(i,j)$, i.e. the loss involved by choosing S_j when S_i is correct, and assume that the a'priori probability for S_k is p_k, $k=0, \ldots, 3$.

Assume temporariry that all subsets, Ω_i, in (13) only contain one member, say the principal and let f_i denote the corresponding density function. Then the expected loss, L_i, which is involved by choosing S_i is given by:

$$L_i = h(\underline{\xi}) \sum_{k=0}^{3} A(k,i) \ p_k \ f_k(\underline{\xi})$$

$$(38)$$

$$h^{-1}(\underline{\xi}) = \sum_{k=0}^{3} p_k \ f_k(\underline{\xi})$$

where $f_k(\underline{\xi})$ is the joint density function corresponding to the principal element in Ω_k.

The hypothesis H_i, that S_i is the actual situation, is accepted if

$$L_i \leqslant L_j \qquad\qquad (39)$$

for all value of j, i.e. the hypothesis which involves the smallest expected loss is chosen.

If, in particular an erroneous rejection of H_j implies a loss, which is independent of the accepted hypothesis, i.e:

$$A(i,j) = \begin{array}{ll} A_i & \text{for } i \neq j \\ 0 & \text{for } i = j \end{array} \qquad (40)$$

the previous inequality, (39), is equivalent to:

$$A_i p_i f_i(\underline{\xi})) \geqslant A_j p_j f_j(\underline{\xi})) \qquad (41)$$

Introduce the quantity

$$L_i(\underline{\xi})) = \log(f_i(\underline{\xi}))) + K_i \qquad (\ 42\ )$$

where

$$K_i = \log(\ p_i A_i\ ) \qquad (\ 43\ )$$

then the test principle can be stated in a more general manner. The hypothesis, H_i, is accepted if there exists an i for which:

$$L_i(\underline{\xi}) - L_j(\underline{\xi}) \geqslant C \geqslant 0 \qquad (\ 44\ )$$

for all j. Here C is a nonnegative quantity.

Now, the subsets Ω_i do not contain only one member. The test principle can then be applied as an ad hoc principle and the manner in which the principal elements are chosen becomes a characteristic property of the test procedure. One method can be a'priori to choose 4 specific members from the subsets. Then the principle density functions are given by:

$$f_i(\underline{\xi}) = f(\underline{\xi};\mu_i,\sigma_i) \qquad (\ 45\ )$$

where:

$$(\mu_i,\sigma_i)\ \varepsilon\ \Omega_i$$

and the parameters μ_i and σ_i, i = 0, ... ,3, are tuning parameters in the test. Another method consists of choosing, from each subset, that density function which describe the observations with largest likelihood, i.e.

$$f_i(\underline{\xi}) = \sup_{(\mu,\sigma)\varepsilon\Omega_i} f(\underline{\xi};\mu,\sigma) \qquad (\ 46\ )$$

Evaluated in a realization of ξ this function is the maximum likelihood function.

The quantity, $L(\xi)$, in (42) could also be interpreted as a weighted log likelihood function. A different approach might then take its origin in this interpretation and the (log) weights, K_i, might be chosen accordingly. For C=0 the test principle states that H_i is accepted if it has the largest likelihood. Notice the relation (and the difference) between the MLR test principle in (16) and the ML test principle in (44) and (46).

When C>0 then H_i is only accepted if the corresponding likelihood is significantly larger than the others. In this case it might happen that none of the hypothesis are accepted. This feature can be used to indicate when the decision can not be based on the present data and additional observations have to be taken.

5.4.5 The ML Detector.

The detector which is to be presented in this section is derived from a discriminator based on the choise:

$$f_i(\xi) = \sup_{(\mu,\sigma)\varepsilon\Omega_i} f(\xi;\mu,\sigma) \qquad\qquad (47)$$

for the principal likelihood function. The detector is denoted as a ML detector due to choice of principal element and due to the relation to the detector based on the MLR test principle. The resulting detector is presented in the end of the section and has a fixed horizon.

The second step toward the discriminator is to investigate in which way the testprinciple divides the sample space into subsets. Using the signals:
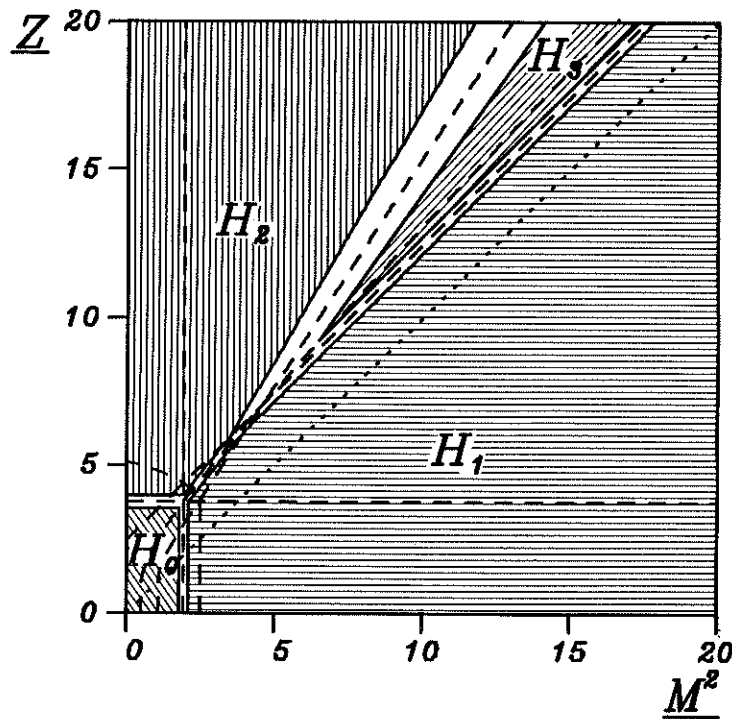
Figure 2: The sample space for $N=6$, $C=0.5$, $K_1=-5.75$. $K_2=-4.27$ $K_3=-7.42$ and $K_0=0$.

$$\underline{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i}{\sigma_o} \qquad \underline{\underline{M}} = N \underline{M}$$

$$\underline{Z} = \frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i^2}{\sigma_o^2} \qquad \underline{\underline{Z}} = N \underline{Z}$$

$$(48)$$

which, in the gaussian case are sufficient statistics for mean and variance, the test principle divides the sample space into subsets as illustrated in figure 2 - 4 for different values of C and K_i.

This discriminator is charactarized by the tuning parameters C, N and K_i. The choise of these parameters can according to the previous part be based on different approaches. In the original approach the weights, K_i, ($C = 0$), are chosen according to the knowledge of the loss function, $A(i,j)$, and the a'priori probabilities, p_i. In the likelihood approach
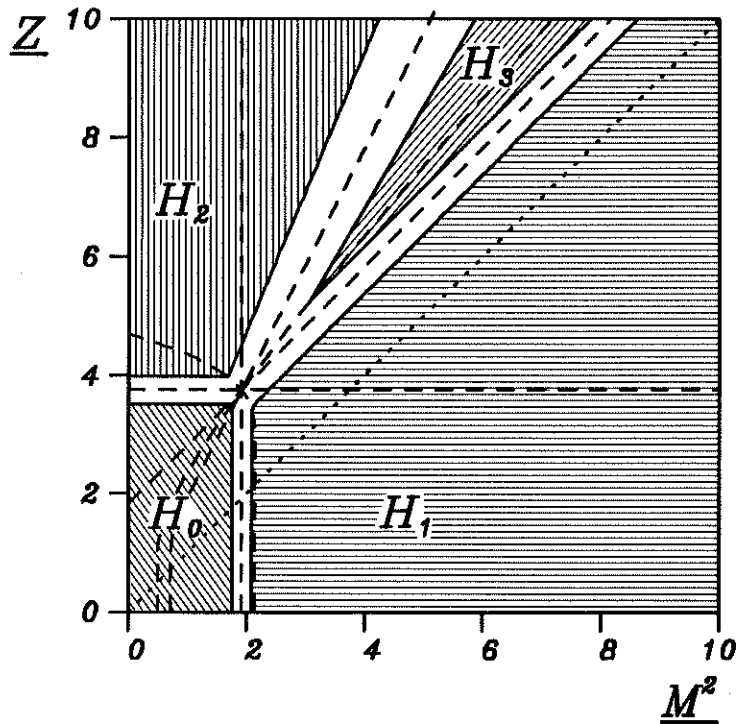
<u>Figure 3:</u> The sample space for $N=6$, $C=0.5$, $K_1=-5.75$.
$K_2=-4.27$ $K_3=-6.42$ and $K_o=0$.

the K_i are chosen as weights or in order to obtain certain properties of the test, as e.g. a certain probability for erroneous rejection of H_o. Here the tuning parameters will be chosen first in order to obtain a simple detector. The remaining tuning parameters are fixed by using the relation to the simple detectors. Let $C = 0$, then the number of limits is significantly reduced. The six limits can, cf. appendix 4, easily be determined and are given by:

$$\underline{M}^2 = \frac{2}{N} (K_o - K_1) \qquad\qquad (49)$$

$$\underline{Z} = g(\frac{2}{N} (K_o - K_2)) \qquad\qquad (50)$$

$$\underline{Z} = g(\frac{2}{N} (K_1 - K_3)) + \underline{M}^2 \qquad\qquad (51)$$

$$\underline{M}^2 = \underline{Z} (1 - \exp(\frac{2}{N} (K_3 - K_2))) \qquad\qquad (52)$$

and
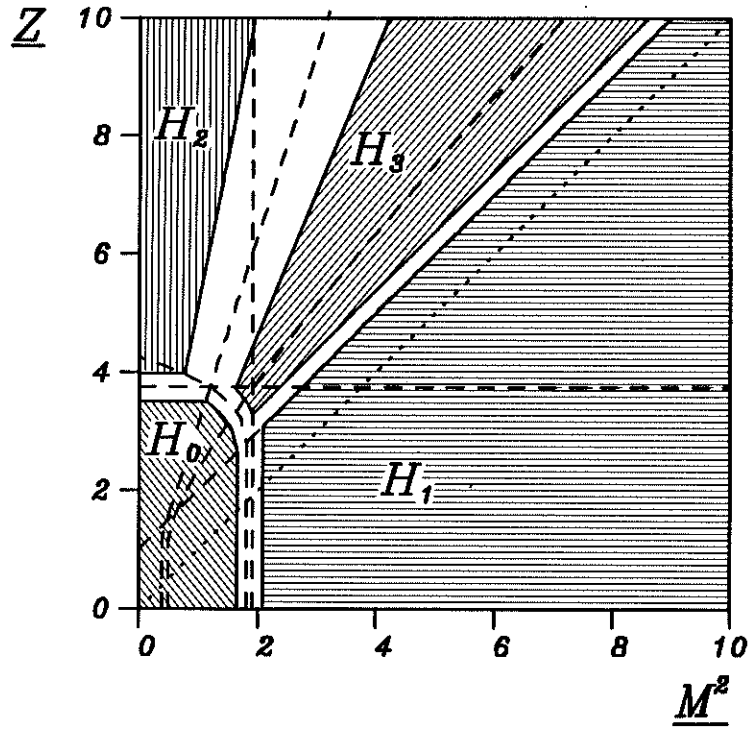
Figure 4: The sample space for $N=6$, $C=0.5$, $K_1=-5.75$, $K_2=-4.27$ $K_3=-5.42$ and $K_0=0$.

$$Z = g(\frac{2}{N} (K_0-K_3) - M^2) + M^2 \qquad (53)$$

$$M^2 = \frac{2}{N} (K_2-K_1) + f(Z) \qquad (54)$$

Here the functions $g(x)$ and $f(x)$ ared defined through:

$$f(x) = \log(x) - x - 1 \qquad (55)$$

and

$$x = \log(g)-g-1 \quad ; \quad g \geqslant 1 \qquad (56)$$

The limits meet in one specific point,

$$(M^2,Z) = (\frac{\nu^2}{N}, \frac{\beta}{N}) \qquad (57)$$

as shown in figure 3, if $K_i = \tilde{K}_i$ for i=0, ... ,3, where:

$$\tilde{K}_1 = \tilde{K}_o - \frac{\nu^2}{2} \qquad\qquad (58)$$

$$\tilde{K}_2 = \tilde{K}_o - \frac{2}{N} \pounds(\frac{\beta}{N}) \qquad\qquad (59)$$

$$\tilde{K}_3 = \tilde{K}_2 + \frac{2}{N}\log(1 - \frac{\nu^2}{\beta}) \qquad\qquad (60)$$

and $\tilde{K}_o$ is abitrary. Notice the tuning parameters, ν and β have to fulfil the inequality:

$$\beta > \nu^2 \qquad\qquad (61)$$

With the choise, (58) - (59) a very simple discriminator results as indicated in figure 3 where $K_i=\tilde{K}_i$, i=0,...,3. In figure 2 and 4 the quantity, K_3, is equal to $\tilde{K}_3-1$ and $\tilde{K}_3+1$, respectively. If ν is chosen equal to a gaussian quantile and β is chosen equal to a $\chi^2(N)$ quantile, the connection to the simpel detector is established. (compare (49), (50) and figure 3 with (20) and (32)).

A suitable choise of tuning parameter is:

$$N=6$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (62)$$

$$\nu = u_{0.9995} = 3.26, \quad \beta=\chi(6)_{0.999} = 22.5$$

which for $\tilde{K}_o=0$ give the values:

$$\tilde{K}_1 = -5.75, \quad \tilde{K}_2 = -4.27, \quad \tilde{K}_3 = -6.42 \qquad\qquad (63)$$

used in figure 2.

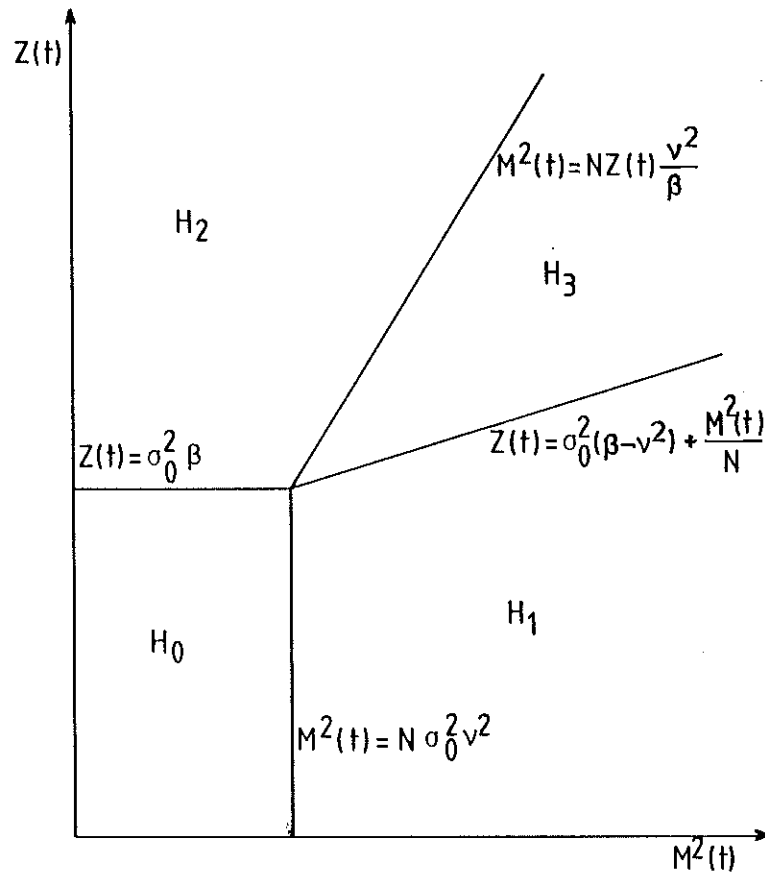The resulting detector, MLD, only consists of the two signals,

Figure 5: The four acceptance regions in the sample space for $C=0$ and $K_i = \tilde{K}_i$, $i=0, \dots ,3$.

$Z(t)$ and $M(t)$, given by the recursions:

$$(1-q^{-1})Z(t) = (1-q^{-N})\varepsilon_r^2(t)$$

(64)

$$(1-q^{-1})M(t) = (1-q^{-N})\varepsilon_r(t)$$

The detection is simply reduced to determine to which acceptance area the point, $(M(t),Z(t))$, belongs. This is, especially with the the choise $K_i = \tilde{K}_i$ and $C=0$, a very easy task.(see figure 5).

Figure 6 shows (for $K_i = \tilde{K}_i$, $C=0$ and the choice, (61)) the probability of rejection of H_0 as a function of μ/σ_0 and σ/σ_0. For the same tuningparameters the probability of accepting H_1 - H_3 are shown in figure 7 - 9.
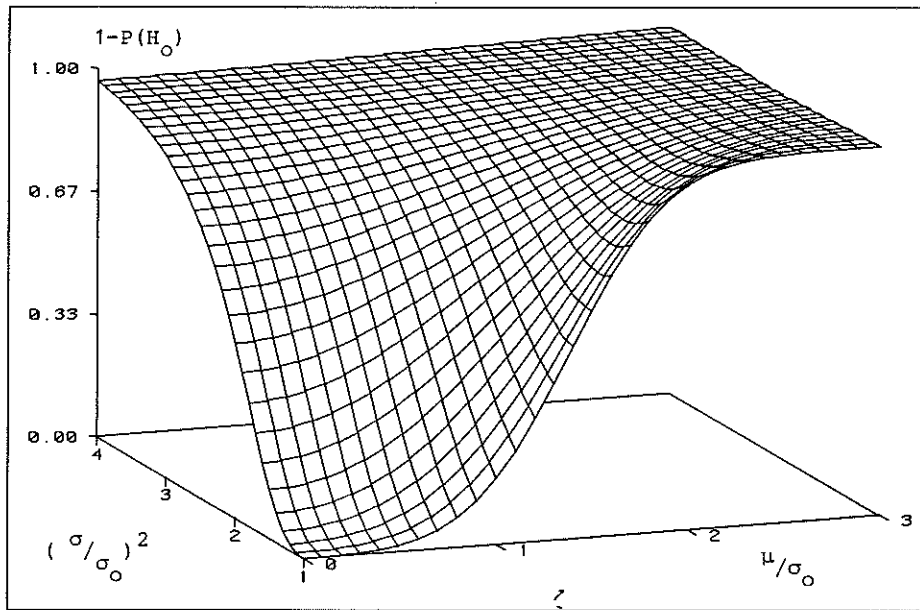
Figure 6: The probability of rejecting H_O as function of μ/σ_O and σ/σ_O.
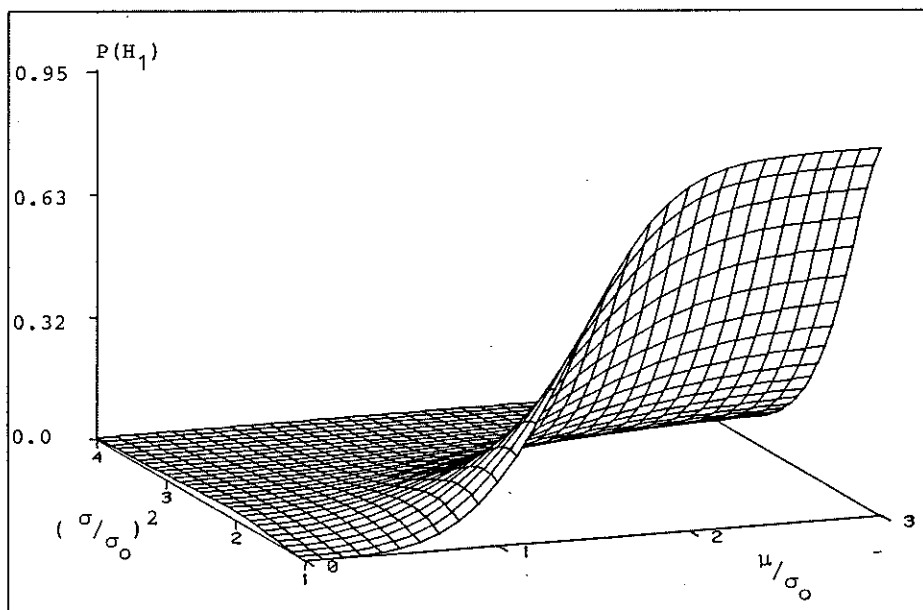
Figure 7: The probability, $P(H_1)$, of accepting H_1 as function of μ/σ_O and σ/σ_O.
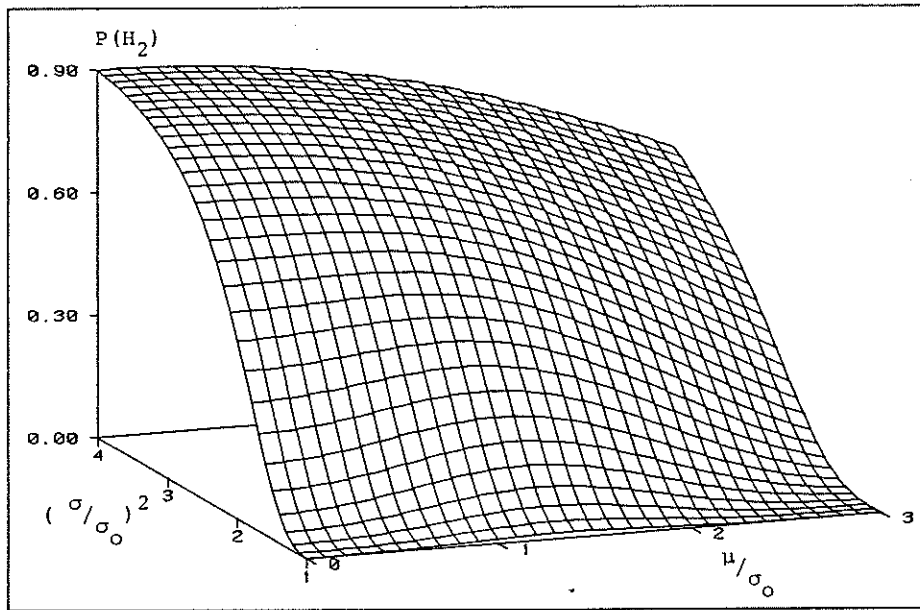
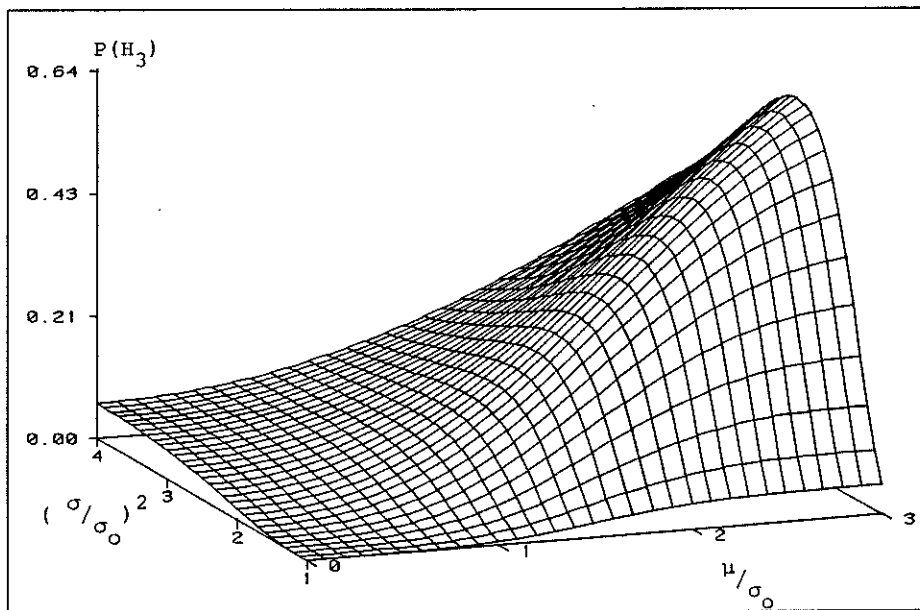Figure 8: The probability, P(H$_2$), of accepting H$_2$ as

Figure 9: The probability, P(H$_3$), of accepting H$_3$ as

5.4.6 The Cusum Detector.

The Cusum detector, CD, which is to be described in this section, is based on a sequential discriminator. Assume that

H_0 has been accepted for the test based on the N-1 latest residuals and focus on the generic discrimination problem which corresponds to basing a determination of the actual situation on the N latest residuals. The principal elements are in this detector chosen a'priori, i.e.

$$f_i(\underline{\xi}) = f(\underline{\xi}; \mu_i, \sigma_i) \qquad\qquad (65)$$

where μ_i and σ_i are chosen from Ω_i. Since the set Ω_1 is not connected it is divided into two connected subsets, Ω_1^- and Ω_1^+, corresponding to positive and negative values of μ_1. The set Ω_3 is correspondingly divided into Ω_3^- and Ω_3^+. The discriminator is for the special choice of μ_i and σ_i then designed as a test between the simple hypothesis:

$$\xi_i \quad \varepsilon \quad N(0, \sigma_o^2) \qquad\qquad : H_0$$

and the simple alternatives:

$$\xi_i \quad \varepsilon \quad N(\gamma\sigma_o, \sigma_o^2) \qquad\qquad : H_1^+$$

$$\xi_i \quad \varepsilon \quad N(-\gamma\sigma_o, \sigma_o^2) \qquad\qquad : H_1^-$$

$$\xi_i \quad \varepsilon \quad N(0, \kappa\sigma_o^2) \qquad\qquad : H_2$$

$$\xi_i \quad \varepsilon \quad N(\gamma\sigma_o, \kappa\sigma_o^2) \qquad\qquad : H_3^+$$

$$\xi_i \quad \varepsilon \quad N(-\gamma\sigma_o, \kappa\sigma_o^2) \qquad\qquad : H_3^-$$

where the tuning parameters fulfil:

$$\gamma > 0 \quad \text{and} \quad \kappa \geqslant 1 \qquad\qquad (66)$$

Again, the second step toward the Cusum detector is for a particular number, N, of observations to investigate in which way the test principle divides the sample space into subsets. Using the normalized sums, $\underline{Z}$ and $\underline{M}$, defined in (48), the test

principle with the actual choise of f_i in (65) is for different values of K_i in (42) dividing the sample space into subsets as indicated in figure 10 - 12.

Direct calculations of (44) show, cf. appendix 4, that the limits which define the regions for accepting the hypotheses are given by:

$$\underline{\underline{M}} = \frac{K_0 - K_1 \mp C}{\gamma} + \frac{\gamma}{2} N \qquad (67)$$

$$\underline{\underline{M}} = \kappa \frac{K_2 - K_3 \mp C}{\gamma} + \frac{\gamma}{2} N \qquad (68)$$

$$\underline{\underline{Z}} = \frac{\kappa}{\kappa - 1} \{ 2[K_0 - K_2 \mp C] + N \log(\kappa) \} \qquad (69)$$

$$\underline{\underline{Z}} = \frac{\kappa}{\kappa - 1} \{ 2[K_1 - K_3 \mp C] + N \log(\kappa) \} + 2\gamma |\underline{\underline{M}}| - N\gamma^2 \qquad (70)$$

$$\underline{\underline{Z}} = \frac{\kappa}{\kappa - 1} \{ 2(K_1 - K_2 \mp C] + N \log(\kappa) + 2\gamma |\underline{\underline{M}}| - N\gamma^2 \} \qquad (71)$$

$$\underline{\underline{Z}} = \frac{\kappa}{\kappa - 1} \{ 2[K_0 - K_3 \mp C] + N \log(\kappa) - 2\frac{\gamma}{\kappa} |\underline{\underline{M}}| + N\frac{\gamma^2}{\kappa} \} \qquad (72)$$

Notice that not all the limits are active at the same time. For C=0 only five limits defines the four domains. As for the ML detector the weights, K_i, can be chosen in accordance to the a'priori distribution, p_i, and the loss function, $\lambda(i,j)$. Here some of this freedom will be exploited for obtaining simple definitions of the acceptance regions. For the sake of simplicity the choice:

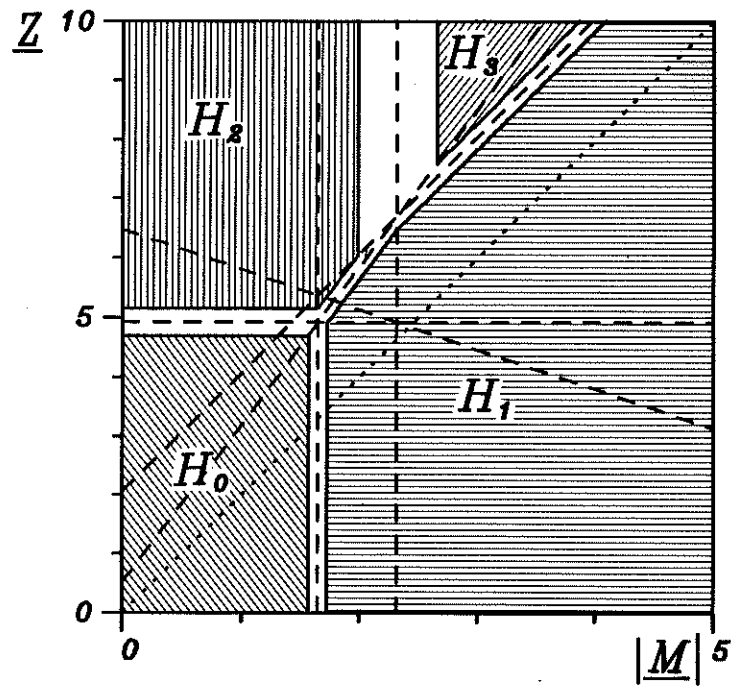$$C = 0 \quad \text{and} \quad K_i = \tilde{K}_i, \; i=0, \ldots, 3 \qquad (73)$$

Figure 10: Sample space for N=6, C=0.5, $K_1 = K_2 = 0$, $K_3 = -2.73$ and $K_0 = 6.91$.
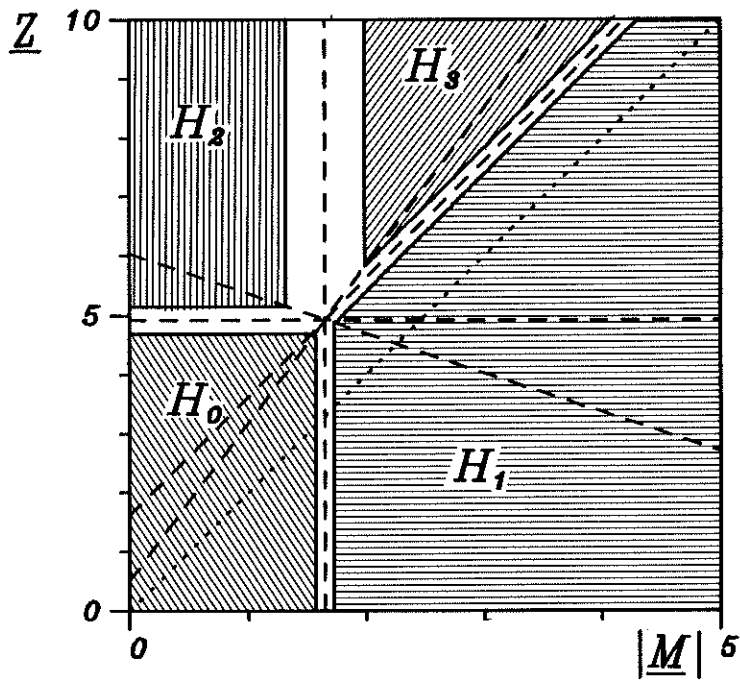

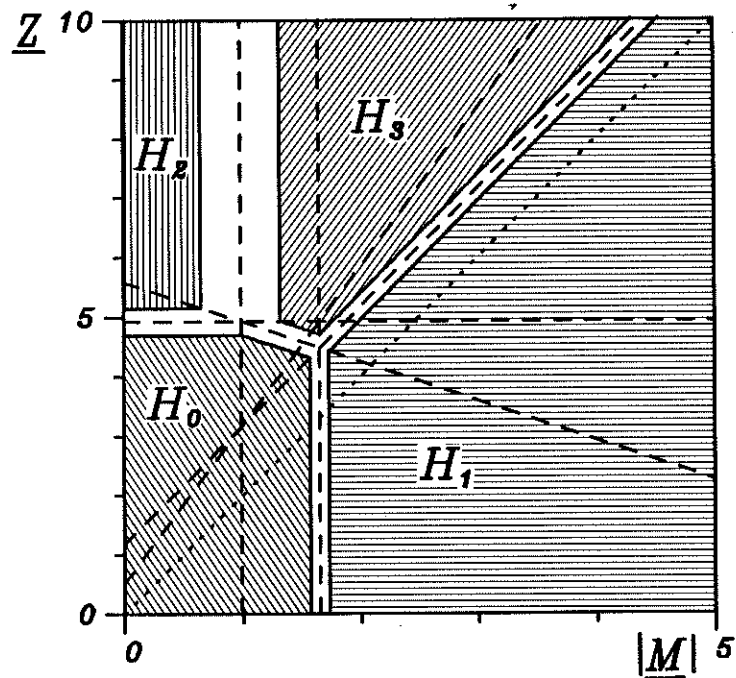Figure 11: Sample space for N=6, C=0.5, $K_1 = K_2 = 0$, $K_3 = -1.73$ and $K_0 = 6.91$.

where:

Figure 12: Sample space for N=6, C=0.5, $K_1=K_2=0$, $K_3=-0.73$ and $K_0=6.91$.

$$\tilde{K}_1 = \tilde{K}_2 = 0$$

$$\tilde{K}_3 = - \frac{\tilde{K}_0}{\kappa}$$

(74)

will be used in the sequel. The acceptance areas in figure 11 are obtained with $K_i=\tilde{K}_i$ and figure 12 and 13 show the acceptance areas for $K_3=\tilde{K}_3-1$ and $K_3=\tilde{K}_3+1$, respectively. This particular choise of K_i makes it easier to apply the Cusum techniques (see figure 12), because the number of detector boundaries is as small as possible. For $K_i=\tilde{K}_i$ and C=0 and a particular number, N, of observations the hypothesis, H_0, is accepted if:

$$|\underline{M}| < \frac{K_0}{\gamma} + \frac{\gamma}{2} N$$

(75)

$$\underline{\underline{Z}} \ < \ \frac{\kappa}{\kappa-1} \ \{ \ 2 \ K_O \ + \ N \ \log(\kappa) \ \} \qquad\qquad (\ 76 \ )$$

The hypothesis H_1 is accepted if only (76) holds. If only the ineqaulity in (76) is violated then H_2 is accepted. Only if both ineqaulities are violated a discrimination between H_1 and H_3 has to be performed. The limit which characterizes this discrimination is given by:

$$\underline{\underline{Z}} \ - \ 2\gamma|\underline{\underline{M}}| \ + \ N \ \gamma^2 \ = \ \frac{\kappa}{\kappa-1} \ \{ \ 2 \ \frac{K_O}{\kappa} \ + \ N \ \log(\kappa) \ \} \qquad (\ 77 \ )$$

In the sequel a special mechanisation (see e.g. Kemp (1962) and Hinkley (1971)) of the sequential test will be used. Since it is difficult to apply this techniques to the quantity, $|\underline{\underline{M}}|$, the test in mean is divided into a test for positive and a test for negative mean. Consequently, the test principle for the choise of parameters given in (73) divides the sample space into six regions. These six regions are illustrated in figure 13. Notice that this discrimination is based on a fixed number, N, of observations. As for the simple detector the weight, K_O, is transformed into:

$$K_O \ = \ \log(B) \qquad\qquad (\ 78 \ )$$

where B is the new tuning parameter which approximately can be interpreted as the expected length between erroneous rejections of H_O.

Now, the discrimination procedure involves a sequentiel test in H_O for an increasing number of observations. In the mechaninsation of the discriminator the cusum technique is applied. Notice that H_O is accepted for all values of N, if the realization of ($\underline{\underline{M}}$,$\underline{\underline{Z}}$) belongs to the pyramid delimited by (75) and (76) as shown in figure 14.

If the hypothesis, H_O, is rejected for one specific number,

$$\frac{2\varkappa}{\varkappa-1}\log(B\varkappa^{N/2})$$

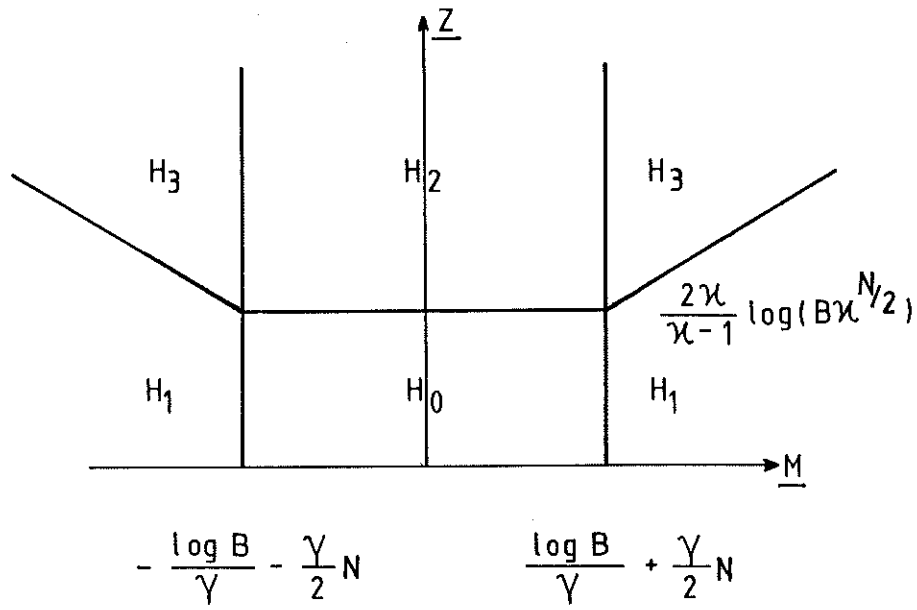$$-\frac{\log B}{\gamma}-\frac{\gamma}{2}N \qquad\qquad \frac{\log B}{\gamma}+\frac{\gamma}{2}N$$

Figure 13: The six domains in the sample space for $C=0$ and $K_i=\tilde{K}_i$.
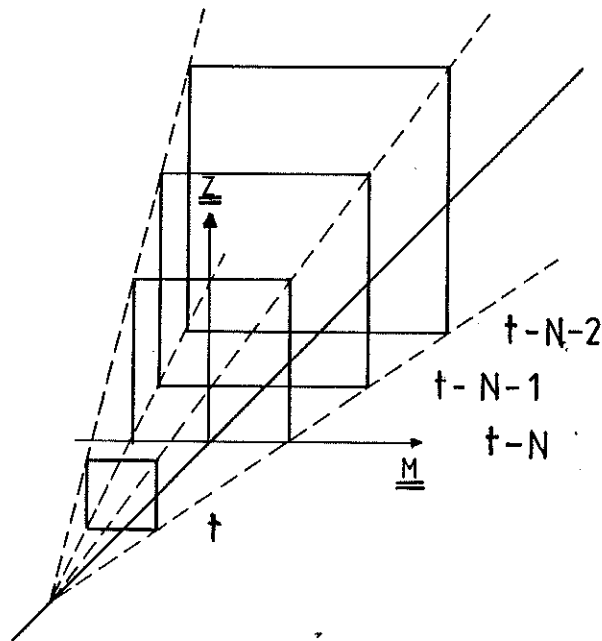
Figure 14: The realization of $(\underline{M},\underline{Z})$ should for all values of N be inside the pyramid.

N_O, of observations then the following test is stopped and a discrimination between the alternatives is performed for this N_O, which for the actual choice of K_i and C is rather

easy.

The resulting detector, CD, can be described by 4 signals and 3 counters given by the recursions:

$$Z(t+1) = \text{Max}[\ Z(t) + \frac{\varepsilon_r^2(t+1)}{\sigma_o^2} - \frac{\kappa}{\kappa-1} \log(\kappa)\ ,\ 0\ ]$$

$$M_z(t+1) = [\ M_z(t) + \frac{\varepsilon_r(t+1)}{\sigma_o}\ ]\ \text{sign}(Z(t+1))$$

$$(79)$$

$$N_z(t+1) = [\ N_z(t)+1\ ]\ \text{sign}(Z(t+1))$$

$$M_+(t+1) = \text{Max}[\ M_+(t) + \frac{\varepsilon_r(t+1)}{\sigma_o} - \frac{\gamma}{2}\ ,\ 0\ ]$$

$$(80)$$

$$N_+(t+1) = [\ N_+(t)+1\ ]\ \text{sign}(M_+(t+1))$$

$$M_-(t+1) = \text{Max}[\ M_-(t) - \frac{\varepsilon_r(t+1)}{\sigma_0} - \frac{\gamma}{2}\ ,\ 0\ ]$$

$$(81)$$

$$N_-(t+1) = [\ N_-(t)+1\ ]\ \text{sign}(M_-(t+1))$$

If the all inequalities:

$$M_+(t) < \frac{\log(B)}{\gamma} \qquad\qquad (82)$$

$$M_-(t) < \frac{\log(B)}{\gamma} \qquad\qquad (83)$$

$$Z(t) < 2\ \frac{\kappa}{\kappa-1}\ \log(B) \qquad\qquad (84)$$

are satisfied, then H_o is accepted for all values of N. The states of the Cusum detector is initiated in zero. If one of the inequalities in (82) - (84) is violated then the corresponding signal and states in the detector are zeroed.

For example, if the inequality in (84) is violated then Z(t), $M_z(t)$ and $N_z(t)$ are put to zero.

Notice, that the detection in the normal situation only consists of an updating of the recursions in (79) - (81) and an evaluation of the inequlities in (82) - (84). Only if one of these inequalities fails to hold, a discrimination is performed. If only one inequality is violated, the discrimination is very simple. If only (82) [or (83)] is violated then H_1 is accepted and $\hat{T}_c = t - N_{\pm}(t)$. The hypothesis, H_2, is accepted if only (82) is violated and then $\hat{T}_c = t - N_z(t)$. If both (84) and (82) [or (83)] fail to hold, a discrimination between H_1 and H_3 has to be performed. If:

$$Z(t) - 2\gamma |M_z(t)| + N_z(t)\gamma^2 < 2 \frac{\kappa}{\kappa - 1} \log(B) \qquad (85)$$

is satisfied the hypothesis H_1 is accepted. The values,

$$\gamma = 1, \qquad \kappa = 4, \qquad B = 1000 \qquad (86)$$

have in simulations shown to give good results.

5.4.7 The Correlation Problem.

The detectors described in the previous part of this chapter are constructed under the assumption that the innovation, e(t), in the output equation:

$$y(t) = \phi^T(t)\theta(t) + e(t) \qquad (87)$$

is a white, $N(0, \sigma_o^2)$ distributed noise sequence. This is fulfilled for the model used in connection with the explicit self tuning controllers. In the implicit poleplacement controller, which was described in chapter 3.1.2, the control parameters $\underline{\theta}(t)$ enter the model:

$$\underline{y}_1(t) = \underline{\phi}^T(t)\underline{\theta}(t) + \underline{e}(t) \qquad (88)$$

which is used in the identification part of the algorithm. Here the signal $\underline{e}(t)$ is a moving average process of order $k-1$, i.e.

$$\underline{e}(t) = E(q^{-1})e(t)$$
$$= [\ 1 + e_1 q^{-1} + \ldots + e_{k-1} q^{1-k}\ ]\ e(t) \qquad (89)$$

The time delay, k, through the plant has to be determined a'priori and enters in the control law explicitly. Since the regressor vector, $\underline{\phi}(t)$, only contains observed signals which are delaed more than k steps the elements in $\underline{\phi}(t)$ and $\underline{e}(t)$ are uncorrelated.

For $k=1$, the noise sequence $\underline{e}(t)$ is equal to $e(t)$ and is consequently white. It is then possible to use the detectors directly. For $k>1$ the noise signal $\underline{e}(t)$ is correlated and the detector has to be modified in order to handle this situation.

In this work, two methods have been investigated. One method consists of estimating the coefficients in the E – polynomial and bases the detection upon the residuals from this estimation. Let $\underline{\varepsilon}_r(t)$ and $\underline{\varepsilon}_o(t)$ be the residual and the prediction error from the estimator designed to estimate $\underline{\theta}(t)$ in the normal mode. For correct estimate of $\underline{\theta}(t)$ the residual sequence is fulfils:

$$\underline{\varepsilon}_r(T) = \underline{E}(q^{-1})\tilde{e}(t)$$
$$= [\ 1 + \underline{e}_1 q^{-1} + \ldots + \underline{e}_{k-1} q^{1-k}\ ]\tilde{e}(t) \qquad (90)$$

and the coefficients in $\underline{E}(q^{-1})$ can be estimated e.g. by a recursive maximum likelihood method. Here $\underline{E}(q^{-1})$ is a polynomial which is obtained by mirroring the zeroes of $E(q^{-1})$

which are situated outside the stability area in the unit circle. If $E(q^{-1})$ has no zeroes outside the stability area then $\underline{E}(q^{-1})=E(q^{-1})$. The noise sequence $\tilde{e}(t)$ is a white, zero mean gaussianly distributed noise sequence which variance, $\tilde{\sigma}_o^2$, satisfies:

$$\tilde{\sigma}_o^2 = \frac{1 + e_1^2 + \ldots + e_{k-1}^2}{1 + \underline{e}_1^2 + \ldots + \underline{e}_{k-1}^2} \sigma_o^2 \qquad (\ 91\ )$$

Introducing the vectors:

$$\underline{\theta}(t) = (\ \underline{e}_1,\ \ldots\ ,\underline{e}_{k-1}\ )^T$$

$$\tilde{\phi}(t) = (\ \tilde{\varepsilon}_o(t-1),\ \ldots\ ,\tilde{\varepsilon}_o(t-k+1)\ )^T \qquad (\ 92\ )$$

$$\tilde{\psi}(t) = \tilde{\phi}(t) - \tilde{e}_1\tilde{\psi}(t-1) - \ldots - \tilde{e}_{k-1}\tilde{\psi}(t-k+1)$$

The RML estimate of $\underline{\theta}$, which in the sequel is referred to as the tail estimate, is in the normal situation given by the recursions:

$$\tilde{\varepsilon}_o(t+1) = \underline{\varepsilon}_r(t+1) - \tilde{\phi}(t+1)\hat{\underline{\theta}}(t)$$

$$\hat{\underline{\theta}}(t+1) = [\ \hat{\underline{\theta}}(t) + \tilde{P}(t+1)\tilde{\psi}(t+1)\tilde{\varepsilon}_o(t+1)\ ]_{stab}$$

$$\tilde{P}^{-1}(t+1) = \tilde{\lambda}(t+1)\tilde{P}^{-1}(t) + \tilde{\psi}(t+1)\tilde{\psi}^T(t+1) \qquad (\ 93\ )$$

$$\tilde{\lambda}(t+1) = \tilde{\lambda}_o\tilde{\lambda}(t) + (1-\tilde{\lambda}_o)\tilde{\lambda}_1$$

The estimate $\hat{\underline{\theta}}(t)$ is at every instant of time projected into the stabibity area by the recursion:

$$\hat{\underline{\theta}}(t+1) = \hat{\underline{\theta}}(t) + \underline{\alpha}\ \tilde{P}(t+1)\tilde{\psi}(t+1)\tilde{\varepsilon}_o(t+1) \qquad (\ 94\ )$$

where $\underline{\alpha}$ is chosen in order to fulfil the stability requirement on $\hat{\underline{E}}(q^{-1})$. This can e.g. be done by a binear

search. The residual:

$$\tilde{\varepsilon}_r(t+1) = \tilde{\varepsilon}_o(t+1) - \tilde{\phi}(t+1)\hat{\underline{\theta}}(t+1)$$

$$= \frac{\tilde{\varepsilon}_o(t+1)}{1 + \tilde{\phi}^T(t+1)\tilde{P}(t)\tilde{\phi}(t+1)} \qquad (\ 95\ )$$

is under S_o an estimate of $\tilde{e}(t)$ and hence it is possible to apply the same arguments as in the preceding sections and to use the detectors on this signal.

The exponential forgetting in the recursions has been introduced in order to reduce the influence from the startup phase in which the estimates $\hat{\theta}(t)$ and $\hat{\underline{\theta}}(t)$ and therefore also the gradient, $\tilde{\psi}(t)$, are erroneous.

It should be noted that the branch estimation also must be performed in this estimator if an abrupt change in the dynamic parameters has been detected. The branch estimation of $\underline{\theta}(t)$ is performed just as to the branch estimation of $\theta(t)$. Furhtermore, the tail estimator can be supplied with any other method from chapter 4 for tracking regular evolution in the plant parameters. If the computional burden is a problem then the estimation of $\underline{\theta}$ and the detection can be based on the prediction error instead.

The tail estimation method has been used in simulations in which the load disturbance had abrupt changes. The method was applicable when the innovation was gaussianly distributed. Since the prediction error, $\tilde{\varepsilon}_o(t)$, enters in the regressor vector, $\tilde{\phi}(t)$ and $\tilde{\psi}(t)$, the method is rather sensitive to outliers. The estimators can, just as when k=1, be robustified with respect to deviations between the datagenerating distribution and the assumed gaussian distribution i.e. also with respect to outliers. Then, the prediction error enters the estimate of $E(q^{-1})$ with bounded

influence. However, a better performance can be obtained by using another method which is to be described in the sequel.

A signal which is obtained as a subsampling of $\underline{e}(t)$ will be white if the length of the sampling interval is greater than k-1, because $\underline{e}(t)$ is a moving average process of order k-1. The second method which has been investigated in this thesis is simply to base the detection on a subsampled version of $\underline{\varepsilon}_r(t)$. It is important that the length of the sampling interval is equal to (or greater than) k in order to obtain a white sequence. Instead of using one sequence, which is obtained by simple subsampling, the detection can be based on k subsequences. Then the detection of a change will be quicker.

If the modifications are introduced in the ML detector, then only the recursions, (64), for the test quantities are modified and become:

$$(1-q^{-k})Z(t) = (1-q^{-Nk})\underline{\varepsilon}_r^2(t)$$

$$(1-q^{-k})M(t) = (1-q^{-Nk})\underline{\varepsilon}_r(t)$$

(96)

For the Cusum detector the modification consists in modifying the recursions in (79) - (81) which then become:

$$Z(t) = \text{Max}[\ Z(t-k) + \frac{\underline{\varepsilon}_r^2(t)}{\sigma_o^2} - \frac{\kappa}{\kappa-1}\log(\kappa)\ ,\ 0\ ]$$

$$M_z(t) = [\ M_z(t-k) + \frac{\underline{\varepsilon}_r(t)}{\sigma_o}\ ]\ \text{sign}(Z(t))$$

(97)

$$N_z(t) = [\ N_z(t-k)+k\ ]\ \text{sign}(Z(t))$$

$$M_+(t) = \text{Max}[\ M_+(t-k) + \frac{\underline{\varepsilon}_r(t)}{\sigma_o} - \frac{\gamma}{2}\ ,\ 0\ ]$$

(98)

$$N_+(t) = [\ N_+(t-k)+k\ ]\ \text{sign}(M_+(t))$$

$$M\_(t) = \text{Max}[\ M\_(t-k) - \frac{\varepsilon_r(t)}{\sigma_0} - \frac{\gamma}{2}\ ,\ 0\ ]$$

$$N\_(t) = [\ N\_(t-k)+k\ ]\ \text{sign}(M\_(t))$$

(99)

The recursions in the simple detectors described in section
5.4.1 - 5.4.3 are embodied in the modified recursions
mentioned above and will not be stated here.

If the computional burden is a problem then the detection can
be based on only one sequence obtained by subsampling (with
an interval length equal to k) of the residual, $\varepsilon_r(t)$ or
the prediction error, $\varepsilon_0(t)$.

5.5 The Exceptional Mode.

The problem which is to be adressed in this chapter is to find methods for estimating the magnitude the parameters changes. Thus assume that the detector has alarmed for an abrupt change in the plant parameters at T_c, say, and let the result of the discrimination be the situation S_m, where $m \in \{1,2,3\}$. The procedure which is to be described here has to perform the branch estimation on the stored data which has been recorded in the period from T_c up to current time, t. In general the goal for this procedure is to determine the point $[\hat{\theta}_m(t), P_m(t)]$ in which the normal operation mode is restarted. In this presentation the normal operation mode is designed to handle constant parameters, but in practice any method described in chapter 4 to handle regulary varying parameters can be applied as well.

The goal of determining the starting point $[\hat{\theta}_m(t), P_m(t)]$ can be achieved by several methods including off line methods. Here the kalmanfilter bank approach which has been outlined in section 5.3 will be followed. If the detector alarms for a change in the parameter at T_c, then it is known that the process noise, v(t) in (2.2:12) has taken a non-zero value in the time interval from T_c to current time t. A change in the parameter can occur due to a single non-zero value of v(t), but also due to non-zero values in a period. Let us thus assume that the following model is valid within this period.

$$\theta(s+1) = \theta(s) + v(s+1)$$

$$T_c \leqslant s \leqslant t \qquad (1)$$

$$y(s) = \phi^T(s)\theta(s) + e(s)$$

where

$$v(s) \in N(0, R_m(s)\sigma_o^2); \qquad m = 1,2,3 \qquad (2)$$

For known R_m the solution to the estimation problem is simply the kalmanfilter which, for $T_c-1 \leqslant s \leqslant t-1$, is given by the recursions:

$$\varepsilon_m(s+1) = y(s+1) - \phi^T(s+1)\hat{\theta}_m(s)$$

$$\hat{\theta}_m(s+1) = \hat{\theta}_m(s) + P_m(s+1)\phi(s+1)\varepsilon_m(s+1)$$

$$P_m(s+1|s) = P_m(s) + R_m(s+1) \tag{3}$$

$$P_m^{-1}(s+1) = P_m^{-1}(s+1|s) + \phi^T(s+1)\phi(s+1)$$

where:

$$P_m(T_c-1) = P_0(T_c-1)$$

$$\hat{\theta}_m(T_c-1) = \hat{\theta}_0(T_c-1) \tag{4}$$

The kalmanfilter designed to handle situation S_m is started in that point which the kalmanfilter designed for the normal situation has passed just before the change in the plant parameters has occured. When the kalmanfilter designed for the actual exceptional situation has performed the branch estimation, then the normal mode of the estimator is restarted in the resulting point, i.e. in:

$$\hat{\theta}_0(t) = \hat{\theta}_m(t)$$

$$P_0(t) = P_m(t) \tag{5}$$

If some a'priori information about the size and directions of the change exists then it is possible to predefine the matrix R_m and use the elements in R_m as tuning parameters. Simulations in which only the load disturbance is changing (see Poulsen (1982) and Poulsen and Holst (1982)) have shown that the method is rather robust with respect to this choice.

It just have to be large enough. If, as assumed here, several parameters might have an abrupt change the problem is more complex and R_m has to be estimated. If furthermore the algorithm in the exceptional mode has to perform a discrimination due to the choice of detector it is again necessary to estimate the elements in R_m. Thus, focus on the second problem in the optimal solution mentioned in section 5.2, namely that the timevariations in $R_m(t)$ are unknown and hence must be estimated.

Assume that $R_m(s)$ is a diagonal matrix, i.e. the 3 exceptional situations are characterized by the 3 matrices:

$$R_1 = \begin{Bmatrix} 0 & & & O \\ & \cdot & & \\ & & \cdot & \\ O & & 0 & \\ & & & q_1^n \end{Bmatrix}$$

$$(6)$$

$$R_2 = \begin{Bmatrix} q_2^1 & & O \\ & \cdot & \\ & & \cdot & q_2^{n-1} \\ O & & & 0 \end{Bmatrix}, \qquad R_3 = \begin{Bmatrix} q_3^1 & & O \\ & \cdot & \\ & & \cdot & \\ O & & & q_3^n \end{Bmatrix}$$

with unknown elements corresponding to the 4 situations. If it, furthermore, is assumed that the timevariations in $R_m(s)$ are relatively slow within the period $T_c - t$ then it is possible to estimate the elements in $R_m(s)$ with a noise adaptive kalmanfilter supplied with exponential forgetting.

Introduce the vectors:

$$\underline{x}(s+1) = (y^2(s), y^2(s-1), \ldots , u^2(s), \ldots , 1)^T \qquad (7)$$

and:

$$q_m = (q_m^1, \ldots, q_m^n)^T \qquad\qquad (\ 8\ )$$

in which the nonnegative diagonal elements, q_m^i, in the i'th row of R_m corresponds to the detected situation S_m.

There exist several methods to estimate the statistical properties of a Markow process as (1). The innovation variance, σ_o^2, is assumed to be constant and can then be estimated in the normal operation mode. With the introduced notation it can be shown that:

$$E\{\varepsilon_o^2(s+1)|Y_s, R_m\} = [\ \beta(s+1) + \underline{x}^T(s+1)q_m\ ]\sigma_o^2 \qquad (\ 9\ )$$

where:

$$\beta(s+1) = [\ 1 + \phi^T(s+1)P_m(s+1|s)\phi(s+1)\ ] \qquad (\ 10\ )$$

In the covariance matching approach which will be applied here the estimate $\hat{q}_m$ is determined as the argument which minimizes the criterion:

$$J(s,q_m) = \sum_{i=1}^{N} [\ \varepsilon_o^2(s+1-i) - E\{\varepsilon_o^2(s+1-i)|Y_{s-i}, R_m\}\ ]^2 \qquad (\ 11\ )$$

However, the assumption that R_m is constant within the interval $T_c - t$ is not always very good and the estimator is accordingly supplied with exponential forgetting.

Since the elements in q_m are nonnegative it is not possible just directly to apply an ordinary least squares estimator. In the presented algoritm the estimate $\hat{q}_m$ is obtained by projecting the ordinary least squares estimate into the nonnegative part of the parameter space, as indicated in figure 1.

If the definitions:

$$
\underline{Y}(s) =
\begin{array}{c}
(\sqrt{\bar{\lambda}})^{s-T_c} \dfrac{\varepsilon_o^2(T_c)}{\sigma_o^2} - \beta(T_c) \\
\cdot \\
\cdot \\
\cdot \\
\dfrac{\varepsilon_o^2(s)}{\sigma_o^2} - \beta(s)
\end{array}
\qquad (\ 12\ )
$$

$$
\Phi(s) =
\begin{array}{c}
(\sqrt{\bar{\lambda}})^{s-T_c} \underline{x}^T(T_c) \\
\cdot \\
\cdot \\
\cdot \\
\underline{x}^T(s)
\end{array}
\qquad (\ 13\ )
$$

are introduced then the ordinary least square estimate $\hat{\underline{q}}_m$ minimizes the criterium:

$$
J(s,\underline{q}_m) = \Xi^T(s)\Xi(s) \qquad (\ 14\ )
$$

where:

$$
\underline{Y}(s) = \Phi(s)\hat{\underline{q}}_m(s) + \Xi(s) \qquad (\ 15\ )
$$

The least squares estimate, $\hat{\underline{q}}_m(s)$, is obtained by the recursions:

$$
\varepsilon(s+1) = \frac{\varepsilon_m^2(s+1)}{\sigma_o^2} - \beta(s+1) - \underline{x}^T(s+1)\hat{\underline{q}}_m(s)
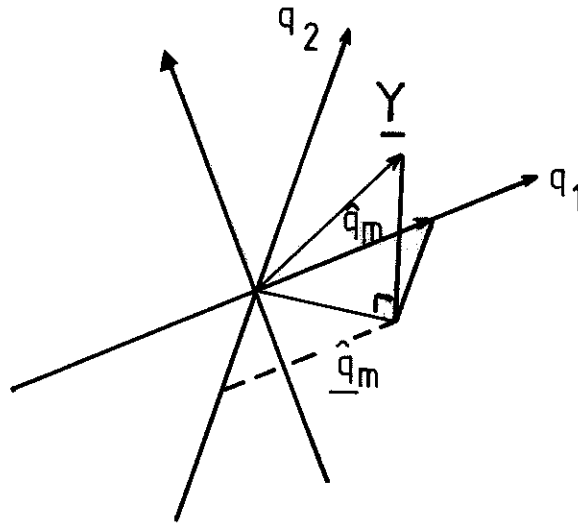$$

Figure 1: The sample spaced spanned by the columns in Φ.

$$\beta(s+1) = 1 + \phi^T(s+1)P_m(s+1|s)\phi(s+1)$$

$$\hat{q}_m(s+1) = \hat{q}_m(s) + \underline{P}_m(s+1)\underline{x}(s+1)\underline{\varepsilon}(s+1)$$

$$\underline{P}_m^{-1}(s+1) = \underline{\lambda}(s+1)\ \underline{P}_m^{-1}(s) + \underline{x}^T(s+1)\underline{x}(s+1)$$

$$\underline{\lambda}(s+1) = \underline{\lambda}^o\underline{\lambda}(s) + \underline{\lambda}^1(1-\underline{\lambda}^o)$$

(16)

In the simulations, which results are shown later in the thesis the initial values:

$$\underline{P}_1(T_c-1) = \kappa_o \left\{ \begin{array}{ccc} 0 & & \\ & \cdot & O \\ & & \cdot \\ O & & 0 \\ & & & 1 \end{array} \right\}$$

(17)

$$\underline{P}_2(T_c-1) = \kappa_o \left\{ \begin{array}{ccc} 1 & & \\ & \ddots & O \\ & & \ddots \\ O & & 1 \\ & & & 0 \end{array} \right\} \; , \; \underline{P}_3(T_c-1) = \kappa_o \left\{ \begin{array}{ccc} 1 & & \\ & \ddots & O \\ & & \ddots \\ O & & \ddots \\ & & & 1 \end{array} \right\}$$

$$\hat{\underline{q}}_m(T_c-1) = (\;0,\;\ldots\;,0)^T, \quad \kappa_o = 1$$

$$\underline{\lambda}(T_c-1) = 0.9, \; \underline{\lambda}^o = 0.9, \; \underline{\lambda}^1 = 0.99$$

are used. It can be seen that $\underline{P}_m(T_c-1)$ contains only nonzero diagonal elements corresponding to the detected situation. The nonzero elements in $\underline{P}_m(T_c-1)$ are put equal to 1.

According to the definition of $\underline{x}(s)$ the elements in the columns of $\Phi(s)$ are nonnegative, i.e. the angle between a pair of column vectors is strictly less than 90°. Since the range space of Φ is spanned by these column vectors, it is not a good idea just to zero the negative elements in $\hat{\underline{q}}_m$ in order to obtain the estimate. The estimate have to be obtained with respect to the angle between the spanning vectors, i.e. the least squaeres estimate, $\hat{\underline{q}}_m$, have to be projected into the positive part of the parameterspace:

$$\hat{\underline{q}}_m(s) = \hat{\underline{q}}_m(s)|_{pos} \tag{18}$$

In the proposed algorithm the projection is simply numerically obtained by inspecting the least squares estimate at every instant of time and reducing the model (15) according to the negative elements in $\hat{\underline{q}}_m(s)$. In appendix 5 a method for obtaining the least squares estimate for a reduced model is described.

5.6 The Variance Estimator.

In several parts of the algorithm a knowledge of the innovation variance, σ_o^2, is needed. However, in connection with adaptive control it is against the basic ideas to assume such a knowledge; the variance has to be estimated.

Since it has been assumed that the innovation variance is almost constant it can be estimated in the normal situation. Consequently, assume that the model:

$$\theta(t+1) = \theta(t) + v_o(t+1)$$

$$y(t) = \phi^T(t)\theta(t) + e(t)$$

$$(1)$$

is correct. Additionally, assume that the white noise sequences:

$$v_o(t) \ \varepsilon \ N(0, R_o \sigma_o^2)$$

$$e(t) \ \varepsilon \ N(0, \sigma_o^2)$$

$$(2)$$

are mutually independent and that R_o is known. Since σ_o^2 is constant but unknown it is possible to assume that σ_o^2 is a stochastic variable and that:

$$\sigma_o^2 \ \varepsilon \ R\Gamma(\alpha(0), \gamma(0))$$

$$(3)$$

where $R\Gamma(\alpha, \gamma)$ is the reciprocal gamma distribution with parameters α and γ. Furthermore, assume that:

$$\theta(0)|\sigma_o^2 \ \varepsilon \ N(\hat{\theta}_o(0), P_o(0)\sigma_o^2)$$

$$(4)$$

Then it will be shown by induction that:

$$\sigma_o^2 | Y_t \ \varepsilon \ R\Gamma(\alpha(t),\gamma(t)) \tag{5}$$

$$\theta(t)|\sigma_o^2, Y_t \ \varepsilon \ N(\hat\theta_o(t), P_o(t)\sigma_o^2) \tag{6}$$

The predictive distribution is determined by the process equation (1) which gives that:

$$\theta(t+1)|\sigma_o^2, Y_t \ \varepsilon \ N(\hat\theta_o(t+1|t), P_o(t+1|t)\sigma_o^2) \tag{7}$$

where

$$\hat\theta_o(t+1|t) = \hat\theta_o(t)$$

$$P_o(t+1|t) = P_o(t) + R_o \tag{8}$$

Since

$$y(t+1)|\theta(t+1),\sigma_o^2, Y_t \ \varepsilon \ N(\phi^T(t+1)\theta(t+1),\sigma_o^2) \tag{9}$$

the conditional distribution of y(t+1) is given by:

$$y(t+1)|\sigma_o^2, Y_t \ \varepsilon \ N(\phi^T(t+1)\hat\theta_o(t+1|t), \beta(t+1)\sigma_o^2) \tag{10}$$

where

$$\beta(t+1) = \eta^2(t+1) = 1 + \phi^T(t+1)P_o(t+1|t)\phi(t+1) \tag{11}$$

Furthermore, is:

$$\frac{y(t+1)}{\eta(t+1)}|\sigma_o^2, Y_t \ \varepsilon \ N\{ \frac{\phi^T(t+1)\hat\theta_o(t+1|t)}{\eta(t+1)} , \sigma_o^2 \} \tag{12}$$

Using Bayes theorem it can be shown that (see appendix 9):

$$\sigma_o^2|Y_{t+1} = \sigma_o^2|y(t+1), Y_t$$

$$\varepsilon \ R\Gamma(\alpha(t+1),\gamma(t+1)) \tag{13}$$

where:

$$\alpha(t+1) = \alpha(t) + \frac{1}{2}$$

$$\gamma(t+1) = \gamma(t) + \frac{1}{2} \frac{[y(t+1)-\phi^T(t+1)\hat{\theta}_o(t+1|t)]^2}{\beta(t+1)}$$

(14)

As in chapter 4.2.2 Bayes theorem applied on (9) gives that:

$$\theta(t+1)|Y_{t+1},\sigma_o^2 = \theta(t+1)|y(t+1),\sigma_o^2,Y_t$$

$$\varepsilon N(\hat{\theta}_o(t+1),P_o(t+1)\sigma_o^2)$$

(15)

where

$$\varepsilon(t+1) = \varepsilon(t+1|t) = y(t+1) - \phi(t+1)^T\hat{\theta}_o(t)$$

$$\hat{\theta}_o(t+1) = \hat{\theta}_o(t) + P_o(t+1)\phi(t+1)\varepsilon(t+1)$$

(16)

$$P_o^{-1}(t+1) = P_o^{-1}(t+1|t) + \phi(t+1)\phi^T(t+1)$$

Since (13) and (15) have the same structure as the origin
(3) - (4) and (5) - (6) the assumptions (5) - (6)
are verified by induction. Applying Bayes theorem on (15)
and (13) the result:

$$\theta(t+1)|Y_{t+1} \varepsilon T[2\alpha(t+1),\hat{\theta}_o(t+1),P_o(t+1)\frac{\gamma(t+1)}{\alpha(t+1)}]$$

(17)

is obtained (see appendix 9): Here $T(x_1,x_2,x_3)$ denotes
the T - distribution with parameters, x_1, x_2 and x_3. The
properties of the T - distribution give the result:

$$E\{\theta(t+1)|Y_{t+1}\} = \hat{\theta}_o(t+1)$$

(18)

Hence, assuming (3) and applying the bayesian approach it

can be concluded that the estimate of σ_o^2 can be performed as a superstructure of an ordinary kalmanfilter. Now, the estimate, $r(t)$, of σ_o^2 is according to the properties of the reciprocal gamma distribution:

$$r(t+1) = E(\sigma_o^2 | Y_{t+1}) = \frac{\gamma(t+1)}{\alpha(t+1) - 1} \qquad (19)$$

Equation, (17) and (14), give the result:

$$r(t+1) = r(t) + \frac{1}{2\alpha(t)-1} \left\{ \frac{\varepsilon_o^2(t+1)}{\beta(t+1)} - r(t) \right\} \qquad (20)$$

For $\alpha(0) \to 1$ and $\gamma(0)=0$ this recursion can be written as follows,

$$r(t) = \frac{1}{t} \sum_{i=1}^{t} \frac{\varepsilon_o^2(t-i+1)}{\beta(t-i+1)} \qquad (21)$$

If another method is used for estimating the parameters in the normal situation, it must be required that $P_o(t+1|t)$ is a reasonable estimate of the variance of $\theta(t+1)|Y_t$.

Partly in order to defeat the influence from the start up period and partly to trace slow variations in σ_o^2 the estimator is supplied with exponential forgetting. Then the estimator for σ_o^2 is given by the recursions:

$$r(t+1) = r(t) + \frac{1}{\Sigma(t+1)} \left[\frac{\varepsilon_o^2(t+1)}{\beta(t+1)} - r(t) \right]$$

$$\beta(t+1) = 1 + \phi^T(t+1)P_o(t+1|t)\phi(t+1)$$

$$(22)$$

$$\Sigma(t+1) = \lambda_r(t+1)\Sigma(t) + 1$$

$$\lambda_r(t+1) = \lambda_r^o\lambda_r(t) + \lambda_r^1(1-\lambda_r^o)$$

where the quantity $\Sigma(t)$ can be interpreted as a common counter. The estimator in (22) can also for $\lambda_r = 1$ be derived as a recursive maximum likelihood estimator for σ_o^2. Notice that this variance estimator, (22), is not supplied with a device for avoiding the influence from the parameter changes. In the earlier proposed versions, Poulsen (1982), Poulsen and Holst (1982) and Holst and Poulsen (1984), such a device has been used successfully.

Example 11: The Variance Estimate.

The following figure shows the estimate of the innovation variance, obtained in two simulations with two different algorithms. The two algorithms are the detector algorithm (see example 13) and the variable forgetting method (see chapter 4.1 and example 8). Plant 2 (see example 2) was used in both simulations and the innovation variance was equal to 0.1. In the simulations shown in this thesis the values:

$$\lambda_r(0) = 0.8$$

$$\lambda_r^o = 0.97$$

$$\lambda_r^1 = 0.999$$

were used in order to defeat the influence from the startup period more rapidly.

In the simulation with the variable forgetting method only the load disturbance has abrupt changes at t=100. In the simulation with the detector algorithm, also the plant zero has an abrupt change at t=200. Due to the insufficient tracking ability of the variable forgetting factor method, the estimate of σ_o^2 gets very wrong after the change. Also in the simulation with the detector algorithm the variance estimate gets too large, because erroneous prediction errors are used in the estimate of σ_o^2. However, the problem is quickly
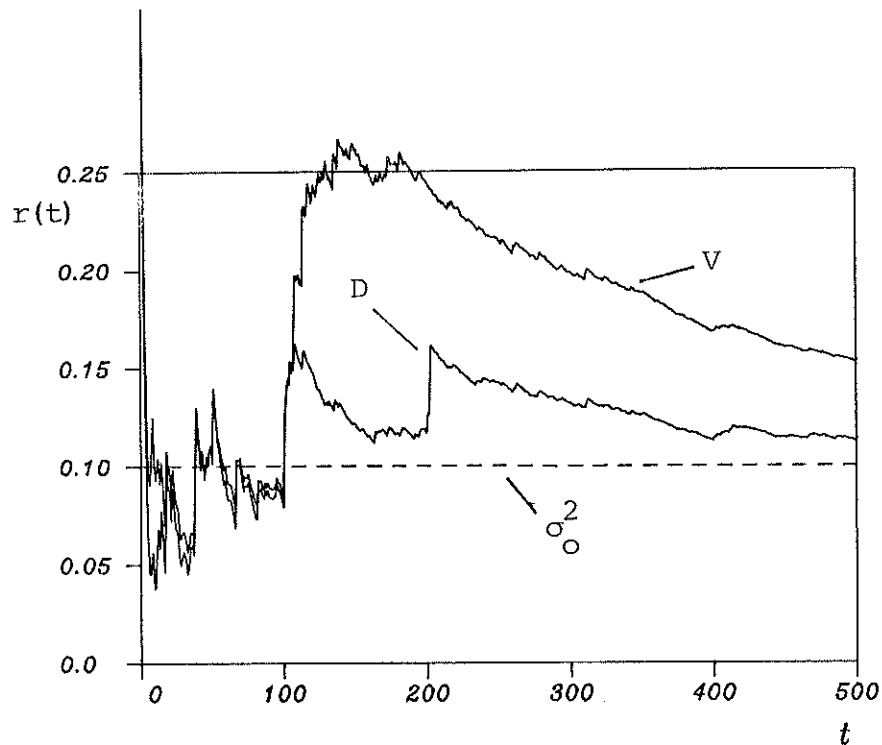
Figure 1: The variance estimate obtained in simulations with the detector algorithm (D) and the variable forgetting factor method (V).

defeated here because of a quick correction of the parameter estimate and the exponential forgetting in the estimation of σ_o^2. If the abrupt changes are more frequent then the erroneous prediction errors must be expelled from the estimation of σ_o^2.

Under adverse circumstances, e.g. when $\Sigma(t)$ is relatively small or when the changes are frequent, a change in the plant parameters may caurse a large increase in the estimate of the innovation variance, σ_o^2. This error in the estimate may even result in an undetected change. One way to avoid this remedy, is simply to exclude the erroneous prediction errors from the estimation of σ_o^2. If data, on which the estimation is based, is delayed, then it is possible only to include the observation in the variance estimate, when it has been accepted that the observation was recorded under the normal

situation. If the used detector has a fixed horizon, N, then the recursion for r(t) is interchanged with:

$$r(t+1) = r(t) + \frac{1}{\Sigma(t+1)} \left[\frac{\varepsilon_\Omega^2(t+1-N)}{\beta(t+1-N)} - r(t) \right] \qquad (23)$$

and $\Sigma(t)$ adjusted accordingly. For a detector based on the Cusum techniques, the signal $\varepsilon^2(t)/\beta(t)$ can be delayed N_{max} steps or less.

5.7 Numerical Examples.

In order to illustrate the properties of the detector algorithm two simulations with plant 1 and 2 will be given here in this chapter.

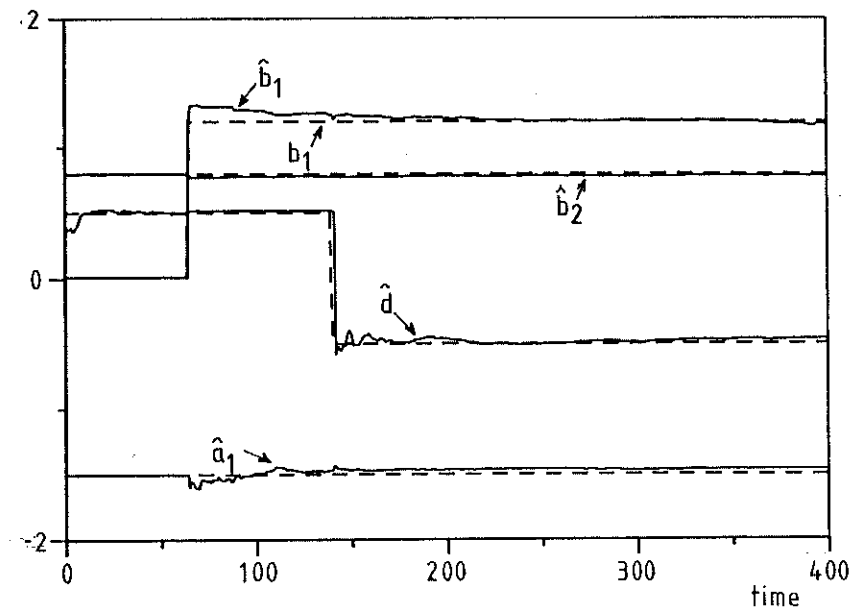Example 12: Simulation with the LQG - controller and plant 1.

Figure 1: The estimated (solid curve) and correct (dashed curve) plant parameters.

The results shown in this example are obtained in a simulation with the detector algorithm and with LQG - control of plant 1 (see example 1 and 3). The Cusum detector, CD, is used and the exceptional modes are constituted by noise adaptive kalmanfilters as described in chapter 5.5. The innovation variance is estimated with an estimator described in chapter 5.6. The tuning parameters in the detector, the variance estimator and the noise adaptive kalmanfilters are chosen as recommended in the respective chapters.

The plant parameters and their estimates are shown in figure 1. Notice, how quick the changes are detected and corrected.
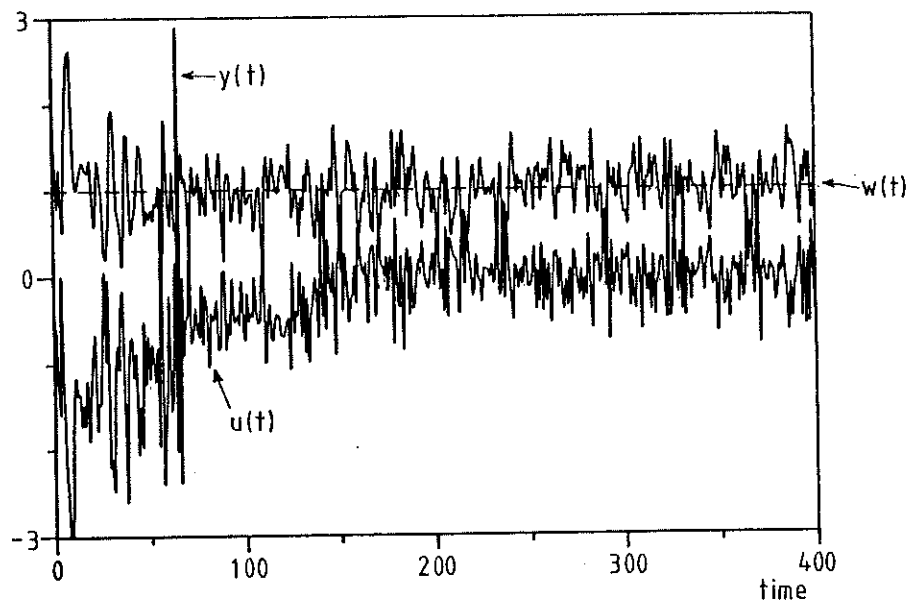
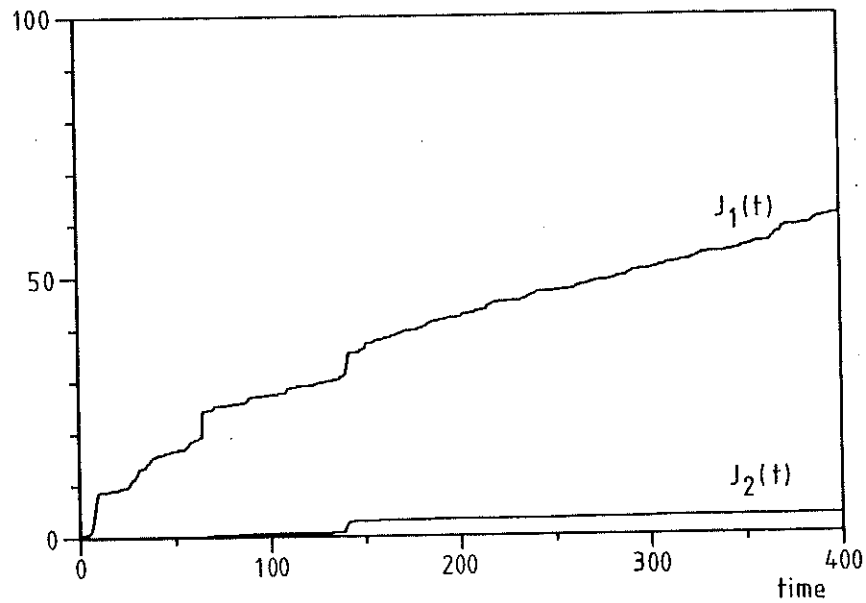Figure 2: The signals, y(t), u(t) and w(t).

Figure 3: The loss functions $J_1(t)$ and $J_2(t)$.

Also notice, that a change in one parameter is only slightly effecting the estimates of the other parameters. The properties of the algorithm are illustrated by the loss functions, J_1 and J_2, defined by:
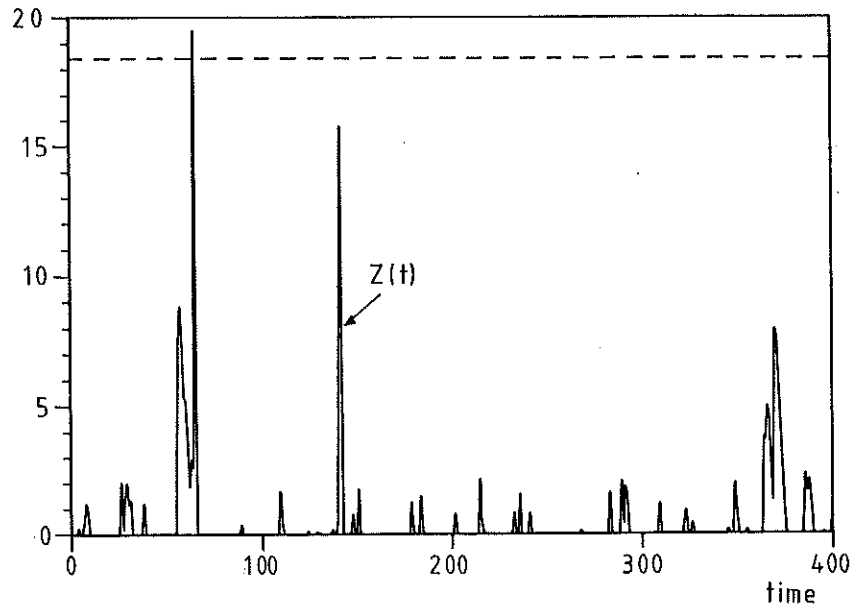
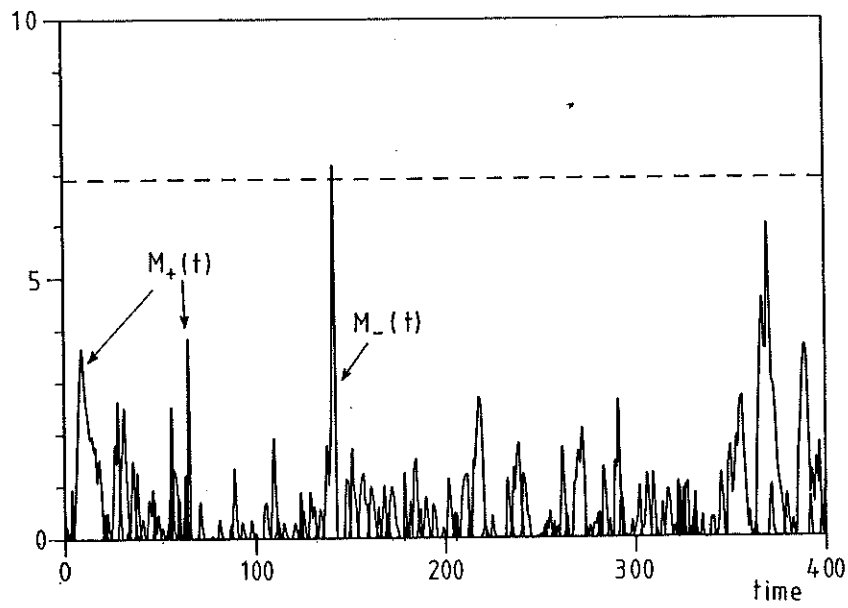**Figure 4:** The detector signal, Z(t) and its threshold limit.

Figure 5: The detector signals, $M_+(t)$ and $M_-(t)$, and their thresholds limits.

$$(1-q^{-1})J_1(t) = [\ y(t) - w(t)\ ]^2$$

and:

$$(1-q^{-1})J_2(t) = [\ \theta(t) - \hat{\theta}(t)\ ]^T [\ \theta(t) - \hat{\theta}(t)\ ]$$

These two loss functions are shown in figure 3 and the signals, y(t), u(t) and w(t) are shown in figure 2.

Firstly, the cornors in the progress of $J_1(t)$ are rather sharp. This indicates that the parameter estimatesare quickly following the abrupte changes when it has been detected. Secondly, the slope of $J_1(t)$ for constant plant parameters are close to the theoretical, i.e. the performance are good. Finally, the direct losses associated with the parameter changes are relatively small. This is mainly caursed by a quick detection. Since the exceptional operation modes are constructed as noise adaptive kalmanfilers and since the detector performs a discrimination between the different situations, mainly the estimates of the changing parameters are effected by the change. The branch estimation is done on stored data and is also reducing the losses. As also the estimated parameters indicates, the algorithm is quickly recovered from an abrupt change.

The detector signals, $Z(t)$, M_+, M_- and their threshold limits are shown in figure 4 and 5.

Example 13: Simulation with plant 2 and the poleplacement controller.

The signals and curves, which are shown in the figures, 6 - 11, are obtained in a simulation with the detector algorithm in connection with the poleplacement controller and plant 2. In the simulation both the load disturbance and the plant zero is abruptly changing. The Cusum detector was used in the simulation. Also in this simulation the tuning parameter in the detector, noise adaptive kalmanfilters and variance estimator are chosen as stated in chapter 5.3 - 5.6.

The shown detector signal, $\tilde{Z}(t)$, $\tilde{M}_+(t)$ and $\tilde{M}_-(t)$ in figure 9 - 11 are normalized with respect to the respective
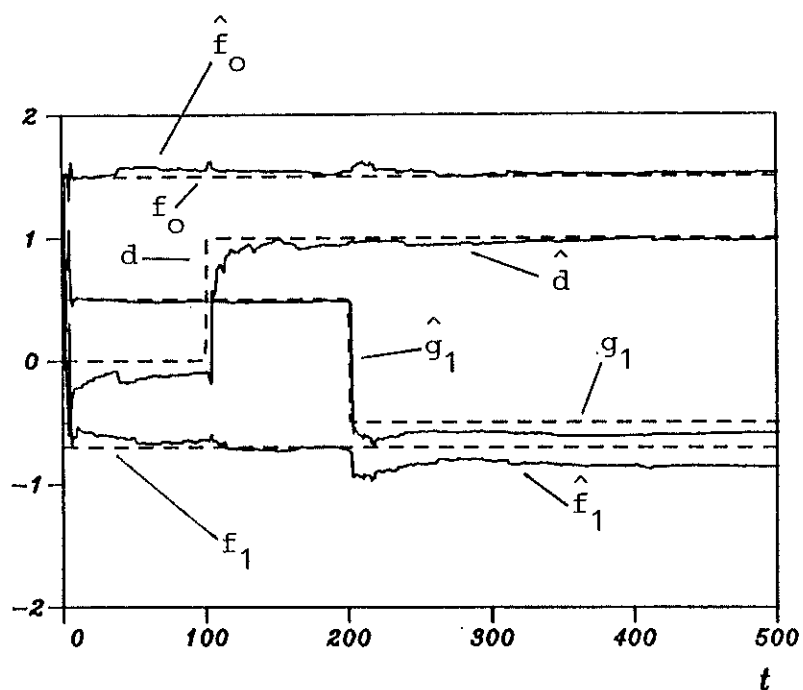
Figure 6: The estimated (solid curve) and correct (dashed curve) parameters.
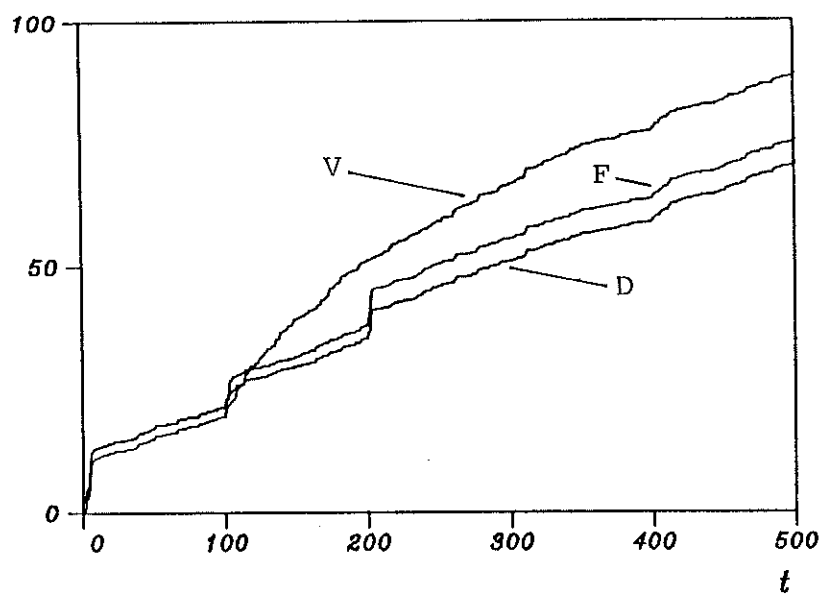
Figure 7: The loss function for the detector algorithm (D), the Filterbank algorithm (F), and the variable forgetting factor method (V).

threshold limit in the Cusum detector.

Figure 8: The diagonal elements of P(t).

Figure 11: The scaled detector signal, $\tilde{M}_-(t)$, which is used for detection of a negative mean value in the prediction error, $\varepsilon_o(t)$.

The figures show that both changes are quickly defeated and
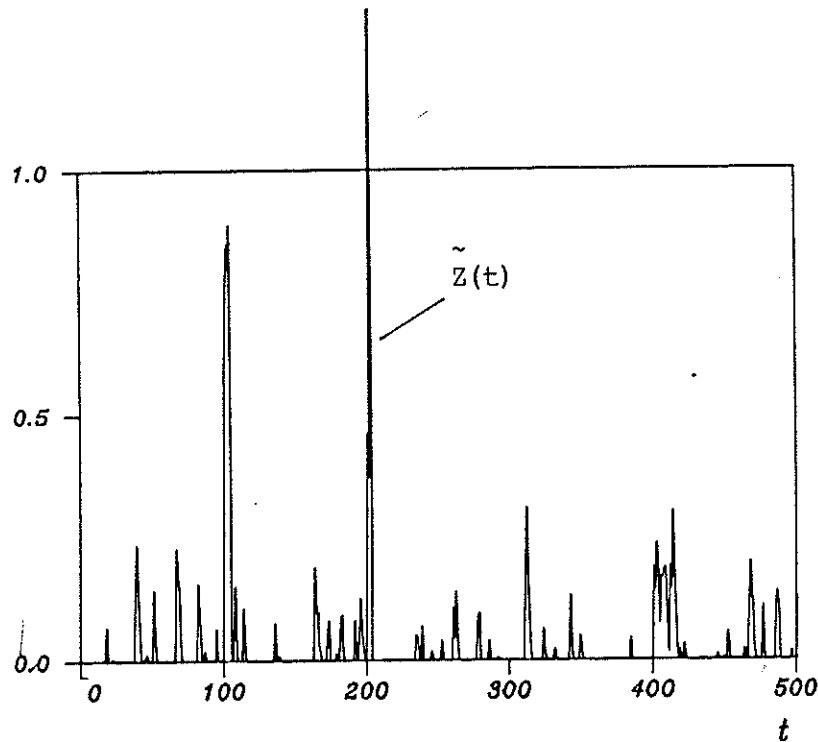
Figure 9: The scaled detector signal, $\tilde{Z}(t)$, which is used for detection of changes in the variance of $\epsilon_0(t)$.
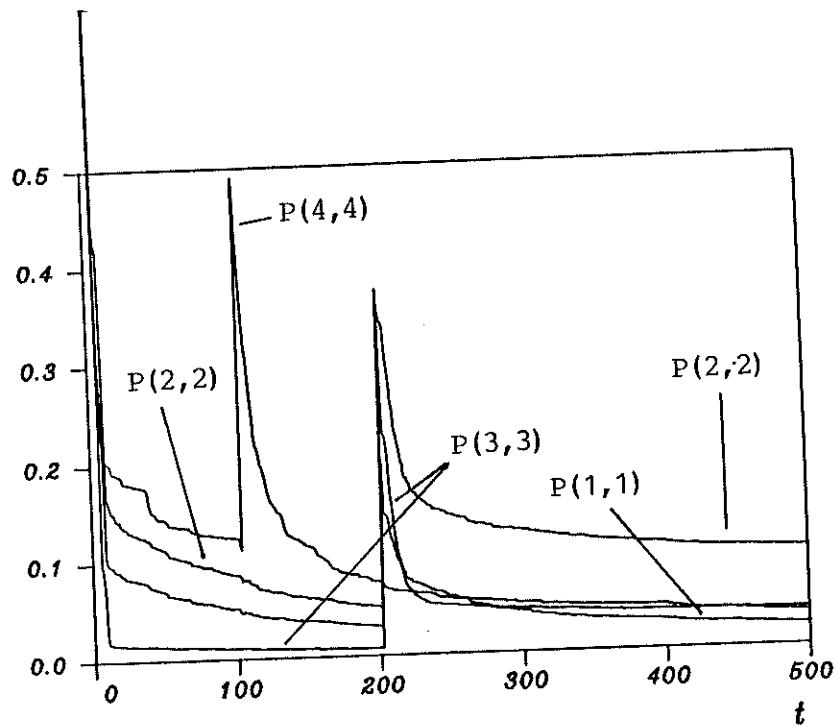
that the algorithm is able to track the changes. Especially the losses associated with the changes are small, which probably is caused by a quick detection and a relatively certain estimate of the change. The discrimination which is performed in the detector, makes it possible to achieve a reasonably certain estimate of the change. Furthermore, the estimation of the change is mainly performed on stored data, i.e. the parameter estimate, which is rather dithering during the branch estimation, is not effecting the control performance, because only the resulting estimate, $\hat{\theta}(t)$, is transferred to the design and control parts.
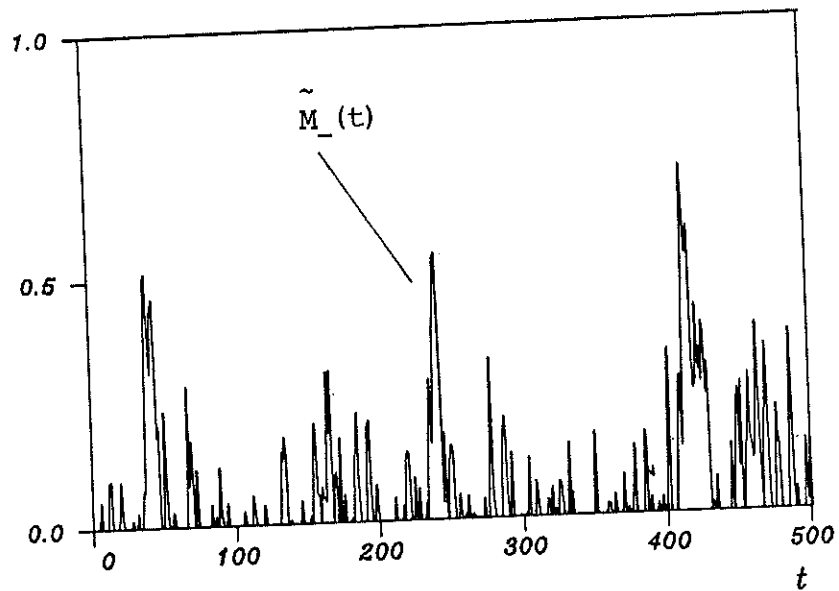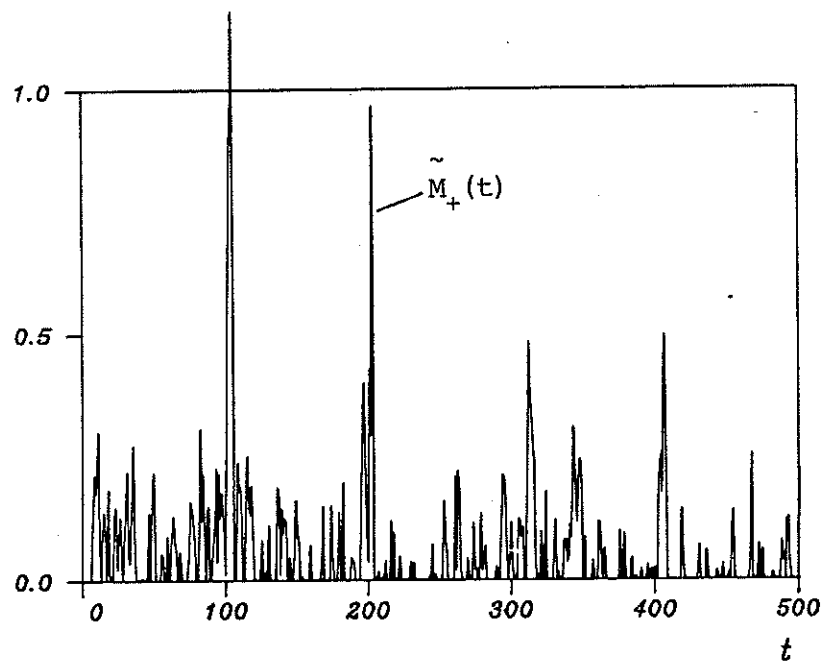
Figure 10: The scaled detector signal, $\tilde{M}_+(t)$, which is used for detection of a positive mean value in $\varepsilon_0(t)$.

5.8 Time Variations
in the
Innovation Variance.

Hitherto, it has been assumed that the innovation variance, σ_0^2 is constant. In practice this assumption is not strictly true, and in this chapter the sensitivity of the algorithm to this assumption will be discussed. Furthermore, methods for modifying the algorithm in order to handle these time variations in σ_0^2 will be discussed.

The assumption that the innovation variance, σ_0^2, is constant is rather convenient, because then σ_0^2 is not included in the estimators (or kalmanfilters). This is an advantage because the estimation performance is rather sensitive to erroneous estimate of σ_0^2. Contrary the estimators, all the detectors described in the previous chapters, involve an estimate of σ_0^2. However, the performance of these detectors is not very sensitive to an erroneous estimate of σ_0^2, because the detector typically consists of a comparison between an observation of a stochastic variable (the test quantity) and a quantile which normally corresponds to a very small probability.

As for the plant parameters a separation into situations corresponding to the type of variation in the innovation variance, σ_0^2, is appropriate.

If the timevariations of σ_0^2 are regular and slow then, it can be estimated with e.g. exponential forgetting as proposed in the previous chapter. The estimator or the kalmanfilter designed for the normal situation in practice also involve some discounting of old information in practice. If the timevariations in σ_0^2 are relatively slow compared to the regular ditto in $\theta(t)$, the detector algorithm can be used as

described previously, i.e. the estimate of σ_o^2 is estimated by exponential forgetting but is not included in the parameter estimator. If the variations in σ_o^2 are quicker then σ_o^2 can still be estimated by exponential forgetting, but the parameter estimator must be modified. The kalmanfilter then becomes:

$$\varepsilon_i(t+1) = y(t+1) - \phi^T(t+1)\hat{\theta}_i(t+1|t)$$

$$\hat{\theta}_i(t+1) = \hat{\theta}_i(t+1|t) + \frac{P_i(t+1)\phi(t+1)}{\sigma_o^2(t+1)} \varepsilon_i(t+1)$$

$$\tag{1}$$

$$P_i(t+1|t) = P_i(t|t) + R_i\sigma_o^2(t+1)$$

$$P_i^{-1}(t+1) = P_i^{-1}(t+1|t) + \frac{\phi(t+1)\phi^T(t+1)}{\sigma_o^2(t+1)}$$

Consequently, regular variations in σ_o^2 can be handled quite easily by standard methods. The problem arise if σ_o^2 might perform abrupt changes, because then it is not possible to determine if an abrupt change has occured in the plant parameters or in the innovation variance, σ_o^2. This is mainly caused by the fact that the detection is only based on the prediction error from the estimator designed for the normal situation.

It is actually possible to design a variance estimator which can track abrupt changes in the variance. Furthermore, it is also possible to design a detector which is insensitive to large variations in σ_o^2. One example of such a detectors is the signtest proposed by Hägglund (see Hägglund (1983) or chapter (6.2)). Another example is a detector, which bases the decision upon data, $r(t)$ and $\varepsilon_o(t)$, from two kalmanfilters; one designed for the normal situation and one with a very short estimation horizon. If the test is performed

in the signal:

$$\underline{Z}_q(t) = \frac{Z_0(t)}{Z_1(t)} \qquad (2)$$

where:

$$(1-q^{-1})Z_0(t) = (1-q^{-N})\varepsilon_0^2(t)$$

$$\qquad (3)$$

$$(1-q^{-1})Z_1(t) = (1-q^{-N})\tau^2(t)$$

the detector has a rather low sensivity to changes in $\sigma_0^2(t)$. Here N is a suitable detector horizon. Other types of test quantities with similar properties have been proposed in Basseville and Benveniste (1982). The only problem left is to design the exceptional mode, i.e. the noise adptive Kalmanfilters, in such manner that both $\sigma_0^2(t)$ and R_m can be estimated simultanusly. Such a kalmanfilter can not be based on the covariance matching approach, which was used previously, but must be based on another principle. However, the research has not yet resulted in an applicable method.

5.9 Summary.

In this chapter the detector algorithm has been presented as an approximation to the optimal kalmanfilter bank algorithm and it has been derived for the explicit type of selt tuning controller (LQG - controller) in connection to the simple system structure. In order to estimate the parameters in a plant which has a more general structure the detector algorithm can be augmented as indicated in appendix 1. The detector algorithm, which is a recursive estimation method, can also be used in other adaptive problems such as adaptive prediction and on line identification.

The detector algorithm consists of an estimator which has several operation modes and a detector which controls the mode. The design of the operation modes can be based on different approaches. Here, the three exceptional modes are designed as noise adaptive kalmanfilters. In the design of the normal mode any method described in chapter 4, for estimating regular varying parameters, can be applied. Another characteristic concept is the branch estimation, i.e. the estimation of the parameter change which is performed on stored data and is started in the point of time, which has been associated with the change by the detector.

The construction of the detector algorithm is based on three ideas. Firstly, to use an estimator with several operation modes controlled by a detector. This idea also includes the branch estimation. It has been introduced in order to separate the contradictary demands of having a good ability of tracking abrupt change and having certain estimates when the parameters are constant or have a regualar evolution. It is also possible to regard the idea as a solution to the first problem connected to the optimal kalmanfilter bank algorithm, i.e. the exponential growing number of filters in the bank. The second

idea is to use a noise adaptive kalmanfilter for estimating the size of the parameter change. This idea has been introduced in order to solve the second problem in the optimal kalmanfilter bank algorithm, namely that the matrices, R_i, $i=1,2,3$, are unknown. The third idea is to perform a discrimination in the detector between the different situations. Then, the exceptional operation mode can be designed to handle one specific situation through which a better estimation of the parameter changes can be gained.

The detector algorithm has been presented in connection with an explicit self tuning controller in which the plant parameters are estimated. The method can also be applied in connection to the implicit poleplacement controller. Since the residual for correct estimates is a moving average process of order k-1 the detector has to be modified. Some modifications of this nature has been described.

The derivation of the detector algorithm is based on the assumption that the innovation variance is constant. Ways to modify the algorithm in order to handle regular variations in σ_o^2 have also been described.

6

Some Related Methods.

This chapter is devoted to related methods for estimating abruptly changing plant parameters. The detector algorithm and the three methods which are to be described in the sequel have in common that they contain an estimator with different operation modes. This feature makes it possible to seperate the two demands of having a good performance in both the normal and in the exceptional situations. Another feature, which these algorithms have in common, is that the operation mode designed for the normal situation easily might be extended with a method from chapter 4 in order to trace regular parameter variations.

It is furthermore possible to regard the four mentioned algorithms as being related to the optimal filter bank algorithm. Firstly, the filterbank algorithm, i.e. an algorithm which is derived as an approximation to the optimal filter bank algorithm, will be described.

The detector algorithm, which merely is a class of algorithms, can also be regarded as an approxiamtion to the optimal solution. As mentioned in the previous chapter the detector algorithm consists of severel building blocks such as the detector and the different operation modes. In chapter 6.2 another algorithm will be described, which can be regarded as derived from the class of detector algorithms. This algorithm, which has been proposed in Hägglund (1983), also consists of a detector and an estimator with two modes; a normal and an exceptional operation mode. The main diversion from the basic principles of the detector algorithms is that this algorithm does not include a branch estimation, i.e. nothing except a change in the operation mode is carried out when a parameter change has been detected.

This part of the thesis ends with a description of an algorithm proposed in Millnert (1982). This algorithm is designed to handle a plant which dynamic properties are

performing abrupt change between a finite number of
possibilities. The design of this algorithm takes its origin
in a specific model, which reflecs the notion of abrupt change
between a finite number, N, of points in the parameter space.
The resulting algorithm also consists of a bank of
kalmanfilters, but due to the particular model, the number, N,
is not exponentially growing.

6.1 The Filterbank Algorithm.

This algorithm has been proposed in Andersson (1983) and the design takes its origin directly in the optimal kalmanfilter bank algorithm. Accordingly, the design is based on the model:

$$\theta(t+1) = \theta(t) + v(t+1)$$

$$y(t) = \phi^T(t)\theta(t) + e(t)$$

(1)

This method only involves one exceptional situation concerning the behavior of the plant parameters and the density function, $f[x]$, of $v(t)$ is a compound of two gaussian densities, i.e.:

$$f[v(t)] = p_o \Phi_N[v(t);0,0] + p_3 \Phi_N[v(t);0,R_3\sigma_o^2]$$

(2)

As mentioned in chapter 5.3 the optimal solution involves two problems. The first problem, i.e. the exponentially increasing number of filters in the bank, is also here solved by a modification of the selection rule. In order to obtain a finite number, N, of filters in the bank of kalmanfilters the following two principles are applied:

1) The branching is only initiated in the kalmanfilters which describes the observed data with largest likelihood (i.e. only $\hat{\theta}_{i_{max}}(t)$ and $P_{i_{max}}(t)$ are used as branching point. Here i_{max} is the index corresponding to the largest likelihood, i.e $\alpha_{i_{max}} \geqslant \alpha_i$ for i=1,..,N).

2) The kalmanfilter which describes the observed data with the smallest likelihood is stopped.

Let i_{min} be the index which corresponds to the smallest likelihood. With the two design principles the evolution

recursions become for i=1, ... , N.

$$\hat{\theta}_i(t+1|t) = \hat{\theta}_i(t) \qquad i \neq i_{min}$$

$$\hat{\theta}_{i_{min}}(t+1|t) = \hat{\theta}_{i_{max}}(t)$$

$$P_i(t+1|t) = P_i(t) \qquad i \neq i_{min}$$

$$P_{i_{min}}(t+1|t) = P_{i_{max}}(t) + R_3$$

$$\alpha_i(t+1|t) = C \, \alpha_i(t) \qquad i \neq i_{min}$$

$$\alpha_{i_{min}}(t+1|t) = C \, p_3$$

(3)

The inference recursions are unchanged:

$$\hat{\theta}_i(t+1) = \hat{\theta}_i(t+1|t) + P_i(t+1)\phi(t+1)\varepsilon_i(t+1)$$

$$\varepsilon_i(t+1) = y(t+1) - \phi^T(t+1)\hat{\theta}_i(t+1|t)$$

$$P_i^{-1}(t+1) = P_i^{-1}(t+1|t) + \phi(t+1)\,\phi^T(t+1)$$

(4)

$$\alpha_i(t+1) = \underline{C} \, \phi_N[\varepsilon_i(t+1);0,\beta_i] \, \alpha_i(t+1|t)$$

$$\beta_i = \{ 1 + \phi^T(t+1)P_i(t+1|t)\phi(t+1) \} \, \sigma_o^2$$

where again the quantities C and $\underline{C}$ ensure that

$$\sum_{i=1}^{N} \alpha_i(t) = 1 \qquad and \sum_{i=1}^{N} \alpha_i(t+1|t) = 1$$

(5)

In figure 1 an exampel of an evolution tree corresponding to the filterbank algorithm is shown

Compared to the original work of Andersson a minor modification of the recursions for $P_{i_{min}}(t)$ has been

introduced and all variances have been factorized by σ_o^2 due to the assumed constancy of the innovation variance. The arguments which were adduced in chapter 5.8 concerning time variations in σ_o^2 are also valid here.
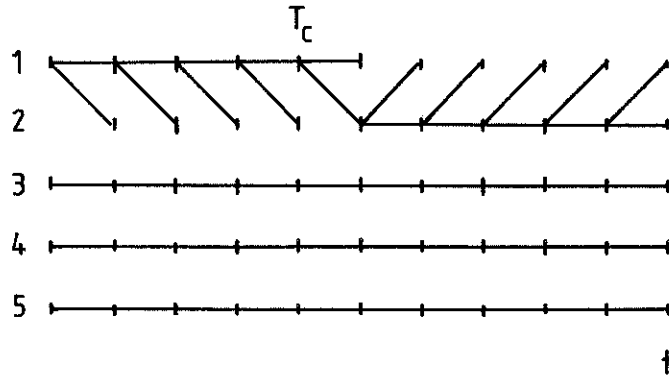
Figure 1: Tree showning the evolution of 5 kalmanfilters
when a change occurs at T_c.

In the original work of Andersson the second problem in the optimal filter bank algorithm, i.e. the unknown R_3 matrix, is handled by treating the elements in R_3 as tuning parameters.

It is also possible to modify this algorithm in order to estimate the elements in R_3. This can e.g. be done by applying a noise adaptive kalmanfilter as described in chapter 5.5. The algorithm can also be extended by increasing the number of exceptional situations. Then some of the work related to the discrimination is transfered from the noise adaptive kalmanfilters to the selection rule and the number, N, of kalmanfilters must then be increased accordingly.

Compared to the algorithm proposed here, this algorithm has 3 additional tuning parameters, namely R_3, N and p_3. Both R_3 and p_3 should be choosen in accordence to the a'priori information about the size of the changes and their probability. The number, N, of kalmanfilters in the bank should (cf. Andersson (1983)) be chosen at the same size as the number, n, of elements in the parameter vector, $\theta(t)$.

Example 14: Simulation with the filterbank algorithm.
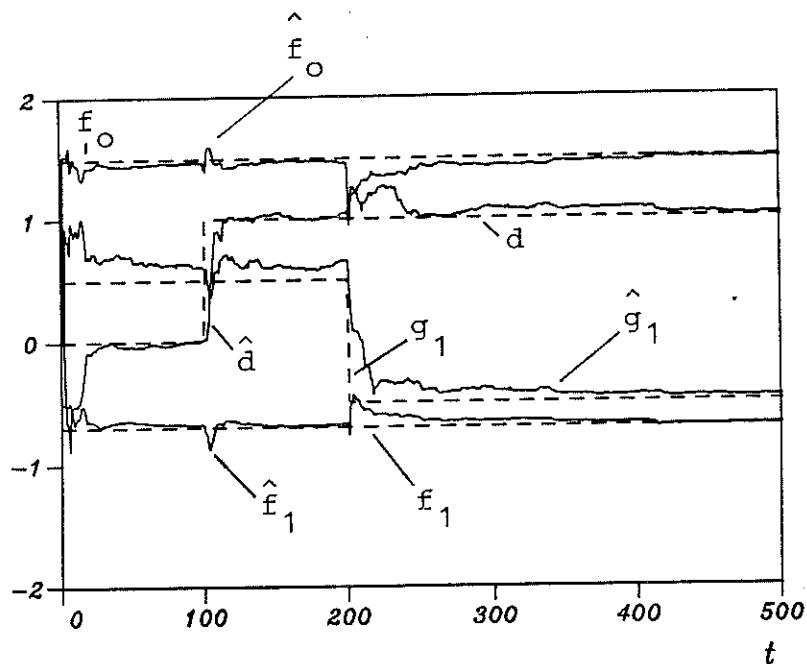
Figure 2: The estimated parameter obtained in a simulation with the filterbank algorithm in connection with plant 2 and the poleplacement controller.

Results from a simulation with plant 2 and the poleplacement controller will be given in this example. The tuning parameters were chosen as:

$$p_3 = 0.005, \quad N = 4 \text{ and } R_3 = 10 \ I_4$$

where I_4 is the unit matrix with 4 rows. The 4 kalmanfilters in the bank are equally started in:

$$\hat{\theta}_i(0) = 0 \text{ and } P_i(0) = 1000 \ I_4$$

and the diagonal elements in the P – matrix are for the restarted kalmanfilter increased by 10. The assumed probability for an abrupt parameter change is one per 200 and the innovation variance is assumed to be known. The evolution of the parameter estimates and the selection rule are shown in

figure 2 and 3, respectively.
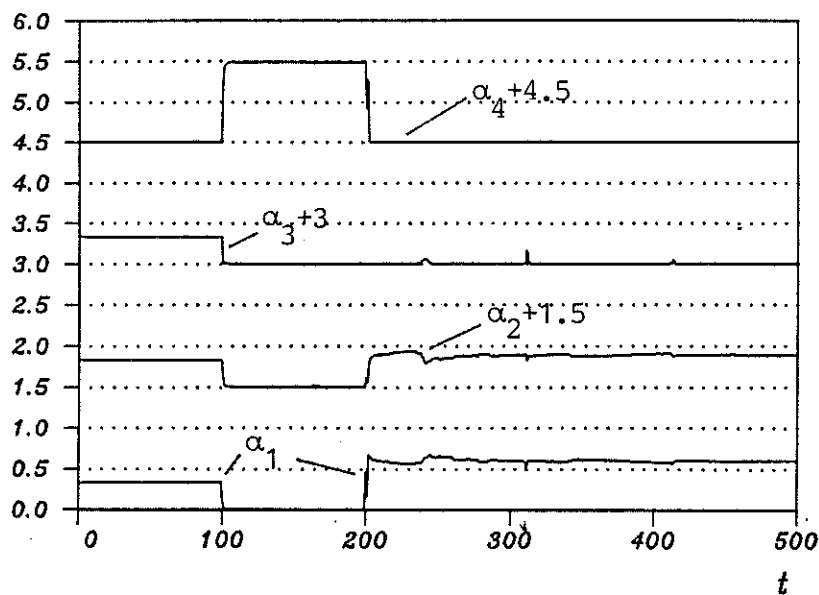
Figure 3: The selection rule, $\alpha_i(t)$.

The progress of the simulation can be divided up into 3 periods which is determined by the points of time, at which the changes take place. In the first period three filters are equally describing the data. The fourth is restarted at every instant of time. After the load disturbance has changed, the fourth kalmanfilter, which was restarted at t=100, is entirely describing the observed data. Notice how quick the changes are detected. Since no discrimination is performed, neither in the selection rule nor in the exceptional estimation mode, an abrupt change in a single plant parameter will effect the estimate of all parameters in a short period just after the change. This might be the main reason why the losses (see figure 4) associated with the parameter changes are larger than for the detector algorithm.

In figure 4 it is seen that the algorithm quickly recovers from an abrupt change in a plant parameter. (The slope of the loss function is equal to σ_o^2 in the normal situation and the corners associated with the changes are sharp). However, the losses, associated with the jumps are larger than for the
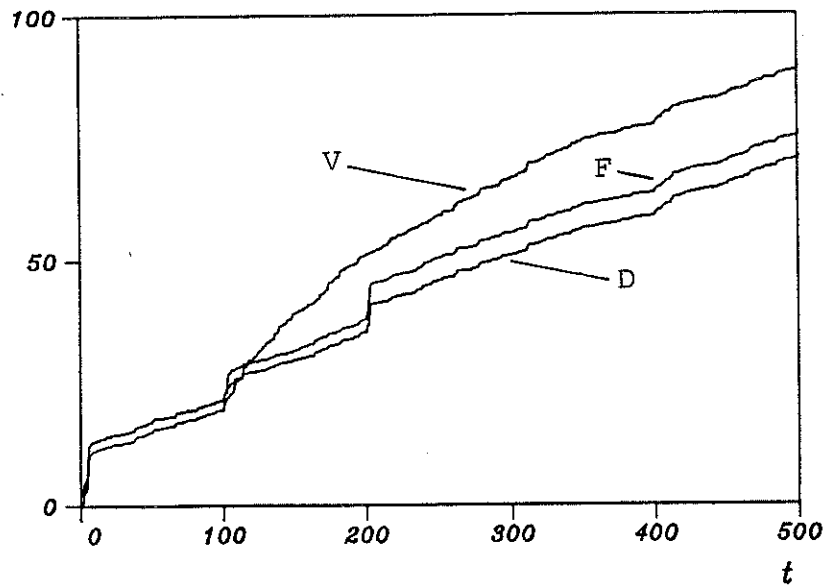
Figure 4: The loss function (defined in example 8) for simulations performed with the detector (D) the filterbank (F) and the variable forgetting factor algorithm (V).

detector algorithm.

This method handles the first problem in the optimal solution, i.e the exponential growing number of filters in the bank, in a reasonable way. The filterbank algorithm requires, however, considerable cumputations, because every N kalmanfilter are updated in every sampling point. On the other hand, the detector algorithm requires that the signal and the states of the kalmanfilter are stored in a number of steps which correspond to the detector horizon.

6.2 The Signtest Algorithm.

The signtest algorithm has been proposed in Hägglund (1983). This algorithm can be regarded as derived from the class of detector algorithms.

The normal situation is handled by a forgetting method as e.g. the exponential forgetting (see section 4.1.3) or the directional forgetting method (see sevtion 4.1.6). Based on data from this estimator a test is carried out. If a change has been detected the operation mode is changed, but no branch estimation is performed. This implies that the estimation of the parameter change is performed on line and the resulting estimates, which could have a rather large variation just after a detected change, enter in the control law. Apart from the different building blocks, the lack of branch estimation is the major difference between this algorithm and the algorithm proposed in chapter 5.3.

The detector might be interpreted as a kind of sign test in the increments of the estimate. More precisely, the detector consist of the recursions:

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

$$\underline{v}(t) = \gamma_1\underline{v}(t-1) + \Delta\hat{\theta}(t)$$

$$\underline{s}(t) = \text{Sign}[\underline{v}^T(t-1)\Delta\hat{\theta}(t)] \tag{1}$$

$$\underline{r}(t) = \gamma_2\underline{r}(t-1) + (1-\gamma_2)\underline{s}(t)$$

and the inequality:

$$\underline{r}(t) < \underline{r}_o \tag{2}$$

where:

$$0 \leqslant \gamma_1 < 1 \qquad \text{and} \qquad 0 \leqslant \gamma_2 < 1$$

Here the actual test quantity is the signal $r(t)$ which is limited to the interval, -1 to 1. For constant parameters and γ_2 close to 1 the signal $r(t)$ is approximately gaussianly distributed with zero mean and variance:

$$\sigma^2 = \frac{1-\gamma_2}{1+\gamma_2} \qquad\qquad (3)$$

The detector is based on the assumption that a parameter change results in a correction of the normal mode estimate mainly in one particular direction. Then the test signal, $r(t)$, will become positive. If $r(t)$ is larger that the threshold limit $r_0 > 0$ the exceptional action is taken.

If the directional forgetting method is applied then the estimator is given by the recursions:

$$\epsilon(t+1) = y(t+1) - \phi^T(t+1)\hat{\theta}(t)$$

$$\hat{\theta}(t+1) = \hat{\theta}(t) + P(t+1)\phi(t+1)\epsilon(t+1)$$

$$\qquad\qquad (4)$$

$$\underline{P}^{-1}(t+1) = \underline{P}^{-1}(t) + [1-\alpha(t+1)]\phi(t+1)\phi^T(t+1)$$

$$P(t+1) = \underline{P}(t+1) + \beta(t+1) I_0$$

where $\beta(t) = 0$ in the normal mode. The matrix I_0, which has to be chosen a'priori, has only non zero elements in the diagonal corresponding to those plant parameters, which might perform abrupt changes. The non zero elements are put equal to one.

If a change has been detected, i.e. the inequality (2) fails to hold, then $\beta(t+1)$ takes a positive value as long as the

test quantity, $\underline{r}(t)$, is larger than the threshold limit $\underline{r}_0$. Notice that all parameters, which might change, are treated in an equal manner.

Due to (4) the time evolution of the estimation error:

$$\tilde{\theta}(t) = \theta(t) - \hat{\theta}(t) \qquad (5)$$

is given by:

$$\tilde{\theta}(t+1) = [I-P(t+1)\phi(t+1)\phi^T(t+1)][\tilde{\theta}(t)+\omega(t+1)]$$
$$(6)$$
$$- P(t+1)\phi(t+1)e(t+1)$$

if the parameter variation follows:

$$\theta(t+1) = \theta(t) + \omega(t+1) \qquad (7)$$

Now, the transition matrix for the evolution of $\tilde{\theta}(t)$:

$$F(t+1) = I - P(t+1)\phi(t+1)\phi^T(t+1) \qquad (8)$$

has n-1 eigenvalues equal to 1 and one eigenvalue equal to:

$$\nu_1(t+1) = 1 - \phi^T(t+1)P(t+1)\phi(t+1)$$
$$(9)$$
$$= \nu_0(t+1) - \beta(t+1)\phi^T(t+1)I_0\phi(t+1)$$

Here $\nu_0(t+1)$ is the eigenvalue corresponding to the normal operation mode. The eigenvalue, $\nu_1(t+1)$ corresponds to the eigenvector, $P(t+1)\phi(t+1)$, which is equal to the direction of the correction, and determines the stepsize of the algorithm.

In this algorithm a detected parameter change results in an increased stepsize for the estimator. This is obtained by adjusting $\beta(t+1)$. Introduce the quantity:

$$S_1(t+1) = \phi^T(t+1)P(t)\phi(t+1) \qquad (\ 10\ )$$

then:

$$\nu_o = 1 - \frac{S_1(t+1)}{1 + [1-\alpha(t+1)]S_1(t+1)} \qquad (\ 11\ )$$

for the directional forgetting method and:

$$\nu_o = 1 - \frac{S_1(t+1)}{\lambda(t+1) + S_1(t+1)} \qquad (\ 12\ )$$

for exponential forgetting. In Hägglund (1983) it is proposed that the eigenvalue, ν_1, is linear reduced if $\underline{r}(t) > \underline{r}_o$, i.e.

$$\nu_1(t+1) = \nu_o(t+1)\text{Max}[\ 1;\ \frac{1-\underline{r}(t)}{1-\underline{r}_o}\ ] \qquad (\ 13\ )$$

Then the increased steplength in the parameter estimation is accompliced by:

$$\beta(t+1) = \frac{\nu_o(t) - \nu_1(t)}{\phi^T(t+1)I_o\phi(t+1)} \qquad (\ 14\ )$$

Due to the nature of the detector an estimate of σ_o^2 is not necessary if σ_o^2 is constant. If the innovation variance is timevarying the algorithm must be modified as described in chapter 5.8 and σ_o^2 must be estimated by e.g. exponential forgetting. In Hägglund (1983) the estimator:

$$r(t+1) = \begin{cases} \gamma_3 r(t) + (1-\gamma_3)\varepsilon^2(t+1-N) & \text{for } \underline{r}(t) < \tilde{\underline{r}}_o \\ r(t) & \text{for } \underline{r}(t) > \underline{r}_o \end{cases} \qquad (\ 15\ )$$

is proposed. Here $\tilde{\underline{r}}_o$ ($\tilde{\underline{r}}_o < 1$) is a threshold limit. The tuning parameter, N, is a suitably chosen delay in order

to ensure insensivity with respect to parameter changes. For undetected changes this estimator is a low pass filter for the square of the prediction error, $\varepsilon(t)$. The pole, r_3, has to be chosen in accordance to the time variations of σ_o^2.

Except for the lack of branch estimation it is possible to regard this algorithm as being a detector algorithm. This cource can, however, be continued. If the detector only consists of a comparison between the square of the prediction error, $\varepsilon(t)$, and its expected value and if the exceptional action is just a decreased forgetting factor then the variable forgetting method (4.1.4) proposed by Fortescue et al. (1981) appears. The main difference between this algorithm and the signtest algorithm (and the detector algorithm) is that the forgetting factor in Fortescue et al. (1981) is a differentiable function of the testsignal, i.e. not a decision in the strict sense. Hence, there is actually no separation in a normal and an exceptional operation mode. Finally, the test in the variable forgetting method is only based on one observation.

Example 15: A simulation with plant 2.

In this example the figures show the results, obtained in a simulation with the signtest algorithm in connection with the poleplacement controller and plant 2. Both the load disturbance and the plant zero have abrupt changes at t=100 and at t=200, respectively.

The estimator, designed to handle the normal situation, is supplied with directional forgetting (see chapter 4.1.5) in order to track slow variations in the parameters. The tuning parameter, a, was chosen to be equal 0.05. Figure 1 shows the evolution of $\alpha(t)$ in the recursion (4.1.6:20) for P(t).

The design variable in the detector, i.e the threshold limit, $\underline{r}_o$, and the poles, r_1 and r_2, in the low pass filters
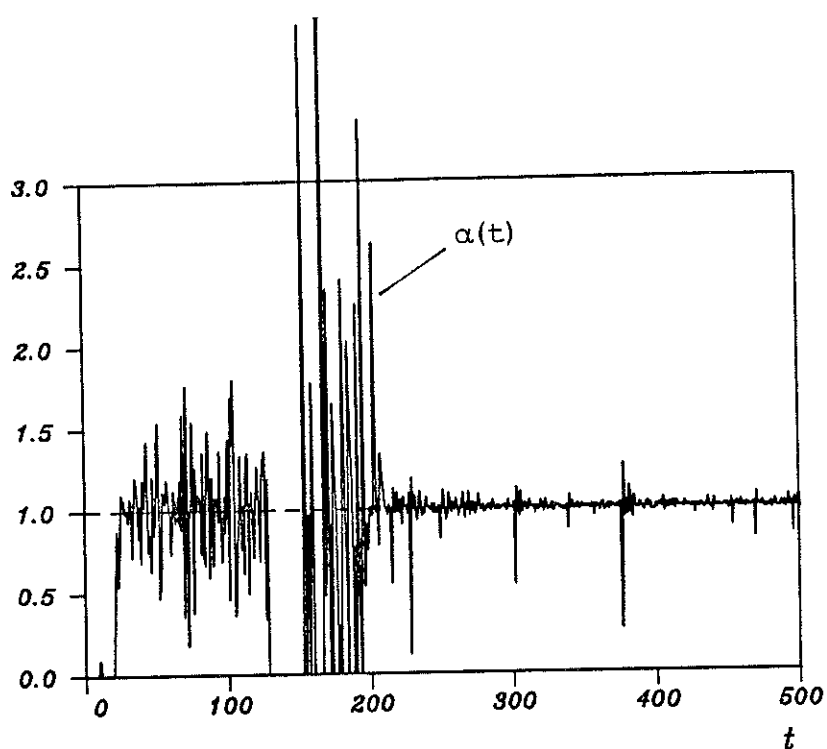
Figure 1: The directional forgetting factor, $\alpha(t)$
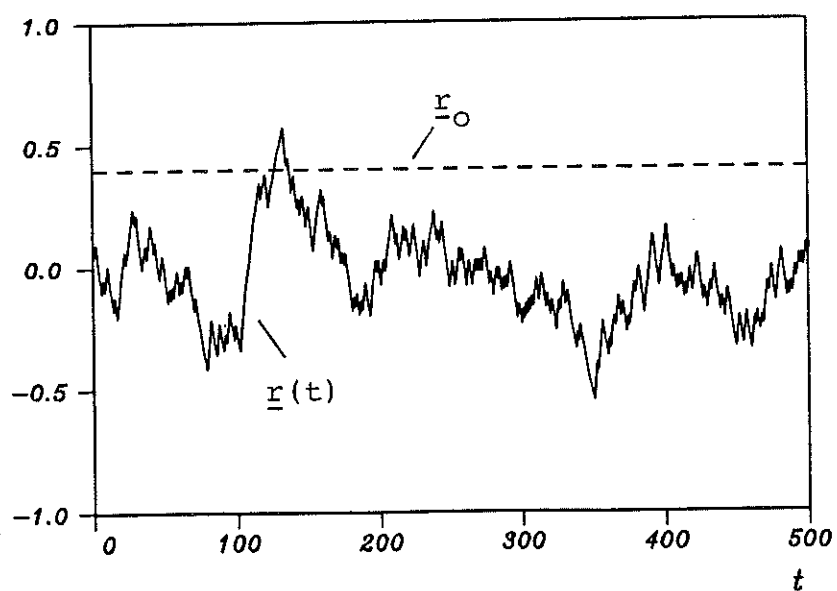
Figure 2: The test quantity, $\underline{r}(t)$, and the threshold limit, $\underline{r}_o$, in the signtest detector.

were chosen to be:
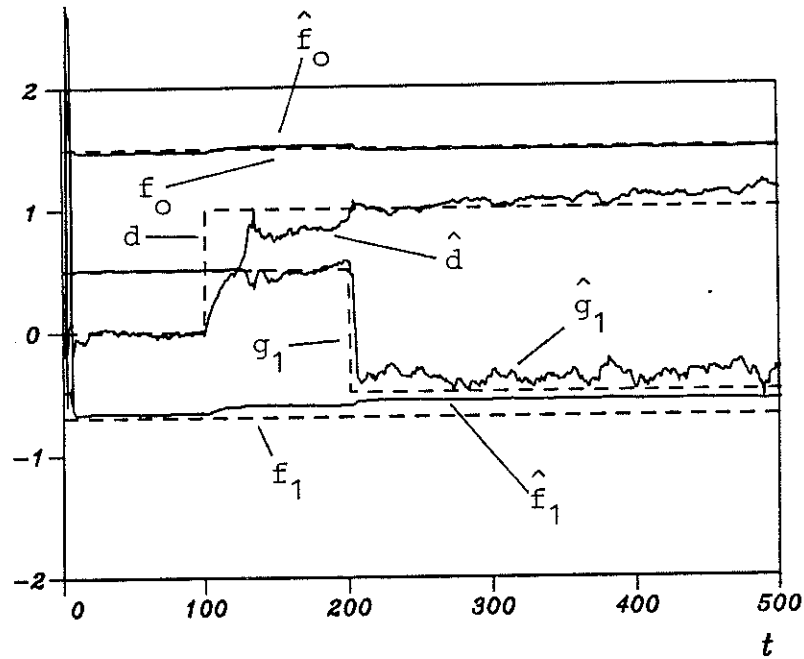
Figure 3: The estimated parameters obtained with the signtest algorithm.
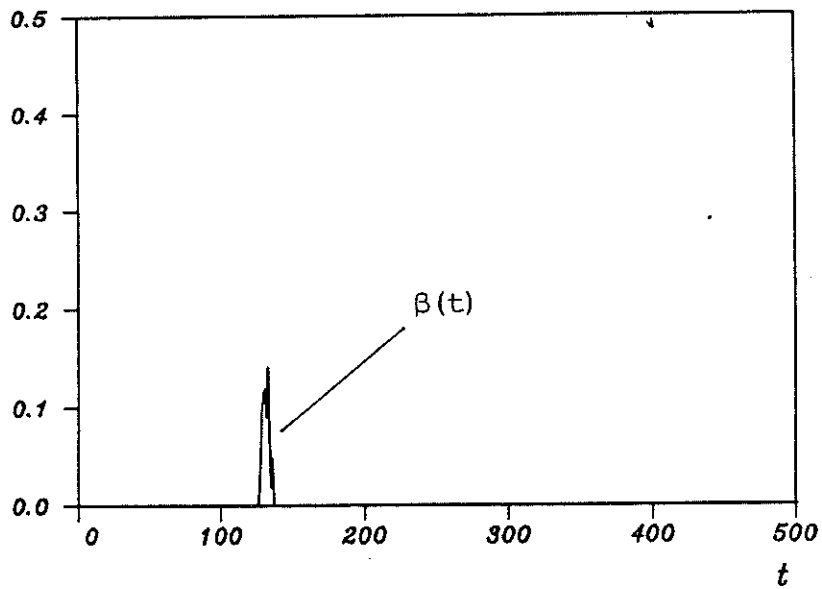
Figure 4: The signal, $\beta(t)$, is added to those diagonal elements in P(t) which correspond to the load disturbance and g_1, i.e. P(3,3) and P(4,4).
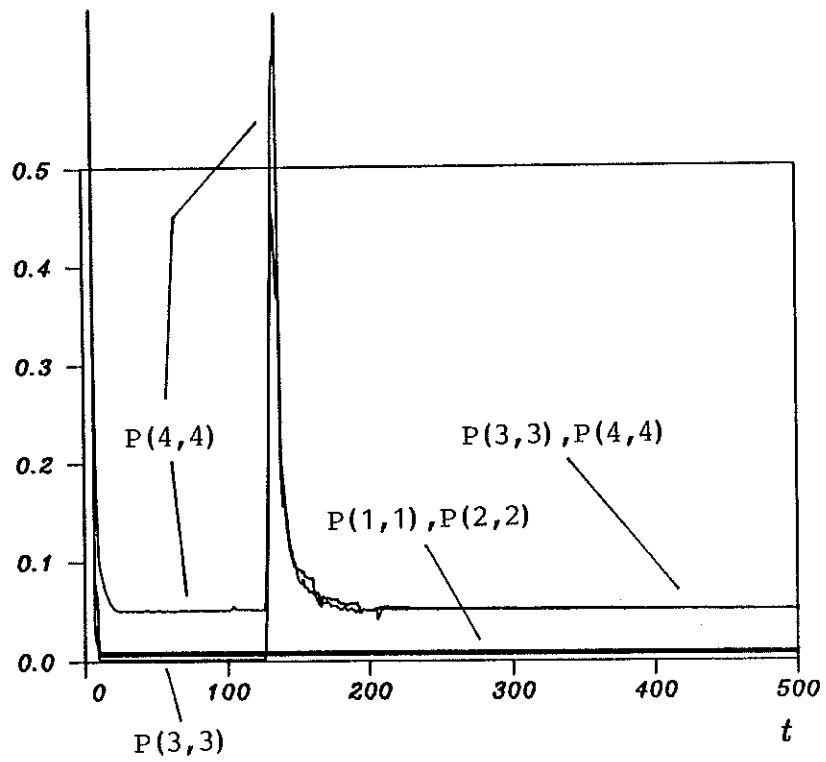
Figure 5: The diagonal element in P(t).
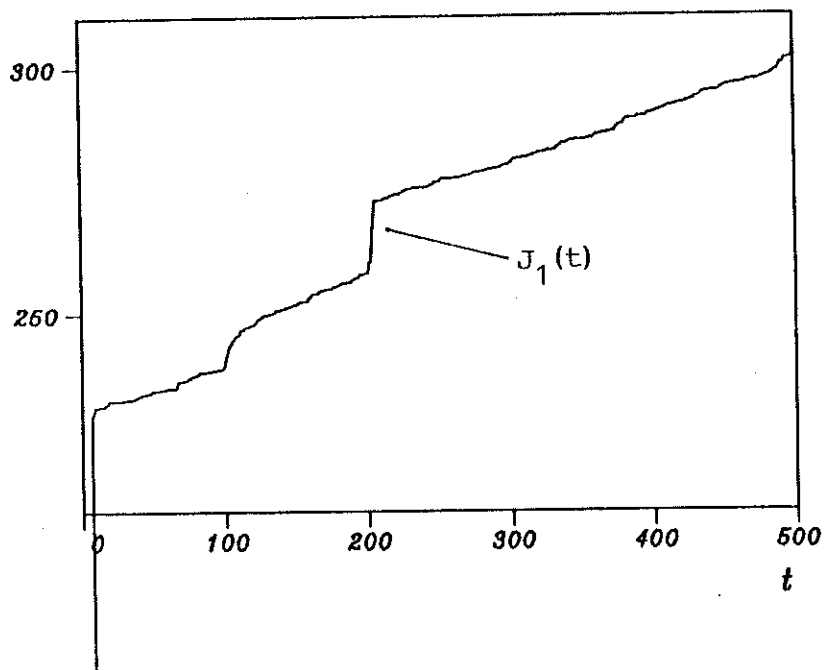
Figure 6: The loss function, $J_1(t)$.

$r_0 = 0.4$, $r_1 = 0.8$ and $r_2 = 0.95$

The test quantity, $\underline{r}(t)$, and the threshold limit, $\underline{r}_o$, are shown in figure 2. In this figure it is seen that the detection of the abrupt change in the load disturbance is not very quick. The change in the plant zero is not detected at all.

The reason why the change in the plant zero is not detected, is probably caused by the change being entirely handled by the normal operation mode, i.e. with directional forgetting. The objective in the directional forgetting method is to obtain a P(t) - matrix which is proportional to the unit matrix. If the P(t) - matrix is actually proportional to the unit matrix, the estimation algorithm becomes a gradient algorithm. Then the direction in which the estimates are changed is only determined by the gradient of the prediction error and the influence from old observations is consequently highly reduced. Hence, compared to other forgetting methods, the directional forgetting method might have good properties with respect to abrupt changing parameters.

In the design of the exceptional mode it is assumed that only the plant zero and the load disturbance might perform abrupt changes. In the figures, 5 and 4, the diagonal elements in P(t) and the signal, $\beta(t)$, are shown.

The loss function, $J_1(t)$, which is defined in example 8, is shown in figure 6. Except for the startup losses this simulation can be compared to simulations performed with other algorithms as e.g. the filterbank, the detector algorithm and the variable forgetting method (figure 6.1:4). Such a comparison shows that in spite of the larger amount of a'priori information this algorithm slowly recovers from the change in the load disturbance. The loss associated with the abrupt change in the plant zero is bigger than obtained in simulations with both the filterbank and the detector algorithm.

6.3 The Regime
Variable Algorithm.

Hitherto, algorithms which are designed to estimate the plant parameters when they might change in an unknown and random direction, have been investigated. These methods are of course also applicable when the plant description might perform abrupt changes between a finite number of models. However, the methods described in the previous chapters are discharging too much information compared to an algorithm which is designed for this specific problem, such as the method described in this chapter. This method is proposed in Millnert (1982).

The design takes its origin in the state space model:

$$\theta(t) = \sum_{i=1}^{N} \delta_{\zeta(t),i} \theta_i$$

$$(1)$$

$$y(t) = \phi^T(t)\theta(t) + e(t)$$

where $\zeta(t)$ is a N state markov chain. Introduce the vector:

$$\underline{\alpha}(t) = (\alpha_1, \ldots , \alpha_N)^T \qquad (2)$$

where $\alpha_i(t)$ is the probability of $\zeta(t) = i$, i.e.:

$$\theta(t) = \theta_i \text{ where } \theta_i \varepsilon \{ \theta_1, \ldots , \theta_N \} \qquad (3)$$

It is assumed that the characteristics of $\zeta(t)$ satisfies the equation:

$$\underline{\alpha}(t+1) = Q \underline{\alpha}(t) \qquad (4)$$

where the elements in the transition matrix, Q, are the jump

probabilities, q_{ij}.

Now, the regime variable algorithm consists for $i=1,\ldots,N$ of the evolution recursion:

$$\alpha_i(t+1|t) = \sum_{j=1}^{N} q_{ij}\, \alpha_j(t) \qquad\qquad (5)$$

and the inference recursions:

$$\hat{\theta}_i(t+1) = \hat{\theta}_i(t) + P_i(t+1)\phi(t+1)\alpha_i(t+1)\varepsilon_i(t+1)$$

$$\varepsilon_i(t+1) = y(t+1) - \phi^T(t+1)\hat{\theta}_i(t)$$

$$P_i(t+1) = P_i(t) - P_i(t)\phi(t+1)\phi^T(t+1)P_i(t)\frac{\alpha_i(t+1)}{\beta_i(t+1)} \qquad (6)$$

$$\alpha_i(t+1) = \underline{C}\ \phi_N[\varepsilon_i(t+1);0,\sigma_o^2]\ \alpha_i(t+1|t)$$

$$\beta_i(t+1) = \{\ 1 + \phi^T(t+1)P_i(t)\phi(t+1)\ \}\ \sigma_o^2$$

Also here the likelihood of the whole kalmanfilter bank, $\underline{C}^{-1}$, is independent of $\theta(t)$ and insures that:

$$\sum_{i=1}^{N} \alpha_i(t) = 1. \qquad\qquad (7)$$

The estimate is given by:

$$\hat{\theta}(t) = \sum_{i=1}^{N} \alpha_i(t)\hat{\theta}_i(t) \qquad\qquad (8)$$

Notice, that the estimator consists of a bank of subestimators which number, N, is fixed. Each subestimator is designed with respect to one specific subsystem and is updated in accordance to the likelihood, $\alpha_i(t)$, of discribing the observed data.

The additional tuning parameters is in this algorithm the jump probabilities in the transition matrix, Q, and the number, N, of subestimators. The initial values of r the likelihoods, $\alpha_i(t)$, must be chosen with respect to the a'priori probability of:

$$\theta(0) \ \varepsilon \ N[\hat{\theta}_i(0), P_i(0)\sigma_o^2] \qquad\qquad (9)$$

The parameters in the subsystems may be timevarying. Then the subestimators must be designed as kalmanfilters or supplied with any other forgetting method described in chapter 4. Also the innovation variance, σ_o^2, may be timevarying and the extentions from chapter 5.8 can be applied.

The basic principles in using a bank of kalmanfilters, each designed to handle one specific situation, for estimating unknown properties of a plant in an unknown situation, can also be found in Lainiotis (1976a-b). However, in these papers the problem is to estimate the states in a plant with unknown parameters. In this approach the paramter space is approximated with set of finite order. Then the state estimator consists of a bank of kalmanfilters each designed for one specific member of the parameter set, i.e. with respect to one specific system. The result from the appropriate kalmanfilter is chosen by a selection rule.

In Wittenmark (1979), the problem of estimating plant parameters is treated by using a two-level estimator consisting of a fine and a coarse part. The coarse part is built upon the basic principles described in Lainiotis (1976), i.e. consisting of a bank of submodels, which parameters are a'priori determined and are not updated. The appropriate submodel is chosen by a selection rule and the result is transferd to the fine estimator, which in the paper by Wittenmark is a kalmanfilter.

6.4 Summary.

In this chapter three related algorithms, which are designed to handle abruptly changing parameters, are described.

Firstly, another approximation to the optimal kalamnfilter bank algorithm is described. Also this algorithm has excelent properties with respect to abruptly changing parameters. Compared to the detector algorithm this algorithm has a heavier computional burden, but has a smaller storage requirement. The filterbank algorithm is not performing a discrimination between different situations which can be usefull e.g. if the load disturbance is frequently changing. It would be interesting to introduce noise adaptive kalmanfilters and a discrimination feature in this algorithm.

Secondly, an estimator with two operation modes and a detector is described. In the detector algorithm the detector and the different operation modes are regarded as building blocks, which can be designed in various manners. Except for the lack of branch estimation, this method can be regarded as a detector algorithm. However, the design of the detector and the exceptional operation mode is different from the design given in chapter 5 in connection with the detector algorithm.

If the plant description is performing abrupt changes between a finite number of models the just mentioned methods are inherently discounting too much information. The regime variable method, which is proposed in Millnert (1982), is designed to estimate the parameters if they are performing abrupt changes between a finite number of points in the parameter space. Also this algorithm consists of a bank of kalmanfilters.

7

Conclusion.

The present thesis is addressed to methods for robustifying self tuning controllers with respect to abrupt changes in the plant parameters.

A detector algorithm, which is designed to handle abruptly changing plant parameters, has been presented in the thesis. The presented algorithm is also able to handle abrupt changes in the time delay through the plant. In appendix 6 a comprehensive description of the algorithm can be found. The properties of the detector algorithm has been investigated by simulations. These simulations show that the detector algorithm has excelent properties. The simualtions also show that the detctor algorithm is superior compared to other methods designed for estimating abruptly changing parameters.

A model of the parameter variations can be found in chapter 2. The detector algorithm, and the other described methods, have been presented in connection with the explicit LQG - controller and the implicit poleplacement controller, but can be applied as the estimation part in any abitrary adaptive controller. The basic design principles for these adaptive controllers are shortly given in chapter 3. The mentioned estimation methods are also useable in other adaptive problems such as on line identification and adaptive prediction.

Chapter 4 contains a comparative description of different algorithms designed to estimate regulary varying plant parameters. Such regular variations can be slow but can also be rather quick compared to the sampling interval. It is a problem to estimation methods, which aim is to estimate regular varying parameters, that they cannot seperate the two contradictary demands of having a good performance when the parameter variations are regular, and having a good ability to track abrupt changes in plant parameters. The problem is mainly caused by the fact that they have only one operation mode.

In general, the proposed detector algorithm, which is described in chapter 5, is characterized by three properties, namely:

1)   it has several operation modes, each designed to handle one specific situation.

2)   it has a detector which control the operation mode.

3)   it performs a branch estimation, i.e. the estimation in the exceptional mode is done on stored data.

Since the algorithm has several operation modes, it is possible to design each operation mode to handle one specific type of variation in the plant parameters. Thus, it is possible to separate the two contradictory demands of having a good performance when the plant parameters have a regular variation and having a good ability to track abrupt changes in the plant parameters. The specific algorithm used in the numerical examples have four operation modes, one normal and three exceptional. The normal operation mode is designed to handle constant or regulary varying plant parameters and can be any of the forgetting or model estimator methods described in chapter 4. In the shown simulations the plant parameters are constant in the normal situation and, consequently, the normal operation mode is an ordinary least squares scheme.

It is likely that the loaddisturbance might change more frequently than the other plant parameters. Hence, the detector algorithm has one exceptional operation mode solely designed to handle abrupt changes in the load disturbance. Another exceptional operation mode is designed to handle abrupt changes in an unknown number of dynamic plant parameters. Finally, the detector algorithm has an exceptional mode designed to handle abrupt changes among all plant parameters.

In the detector algorithm the exceptional modes are constructed as noise adaptive kalmanfilters (see chapter 5.7) which both estimates the abruptly changing plant parameters and the variance, $R_m\sigma_o^2$, of the process noice. Hence, the detector algorithm is in one of the exceptional mode able both to determine which of the parameters have changed and to determine the magnitude of the change.

The detector, which control the operation mode, has three tasks:

it has to detect the abrupt changes.

it has to discriminate between the situations, if a change has occured.

it has to estimate the point of time at which the change has occured.

In chapter 5.6 several detectors have been proposed and investigated. The most general detector, the Cusum detector, which is described in chapter 5.4.6, discriminations between the exceptional situations, if an abrupt change has been detected. Additionally, the detector also estimates the point of time at which the changes have occured.

If an abrupt change has been detected a noise adaptive kalmanfileter which is designed to handle the actual detected situation, is started in the point associated with the change by the detector and estimates the parameter change on stored data. Since this branch estimation is done on stored data the abrupt change has a highly reduced influence on furure control actions.

For the sake of simplicity the detector algorithm has been derived in connection to the simpel model structure embodied

in (2:5). It is, as indicated in appendix 1, possible to apply the detector algorithm on the general model structure.

In chapter 5.8 some methods for handling regular variations in the innovation variance have been described. However, the problem which arises if the innovation variance has abrupt changes can not yet be solved with the described methods, because the research has not resulted in a method for simultaneously estimation of the innovation variance and the matrices, R_i, i=1,2,3.

In order to robustifying the self tuning controllers there has in this thesis been focussed on the assumption that the plant parameters are constant or have a regular variation. An algorithm, which is robust with respect to the mentioned assumption has been proposed in this thesis. Additionally, some related methods have been described.

However, the design of self tuning controllers is not only based on the assumed regular parameter variations. It is also assumed that the innovation signal is gaussianly distributed. In general this assumption is only a mathematical convenient rationalization which in practice is a good approximation. If the innovation distribution is heavier tailed than the gaussian, the ordinary estimators based on the gaussian assumption or the least squares principles might give problems. Outliers or gross errors, i.e. extreme large noise pulses can damage the parameter and the variance estimates. This is especially a problem in algorithms designed to estimate varying parameters. Methods for robustifying self tuning controllers with respect to deviations between the datagenerating distribution and the assumed gaussian have been proposed in Poulsen (1982) and in Poulsen and Holst (1982). These methods are based on the work in Huber (1964) and in Poljak and Tsypkin (1980). If these methods are introduced in the detector algorithm a very low sensivity with respect to outliers is obtained.

The design of self tuning controllers is also based on several structural assumptions. In the poleplacement controllers it is assumed that the time delay through the plant is known. This assumption is avoided in the LQG – controller in which abrupt change in the time delay can be handled by the detector algorithm. In the design of the self tuning controllers it is also assumed that the process can be described by a linear model of finite order. Future research might be directed towards methods for further robustifying self tuning controllers with respect to this assumption.

8

Appendices.

A1. THE GENERAL ESTIMATION ALGORITHM.

The methods for estimating timevarying plant parameters were presented in connection to the simple system structure embodied in the model, (2:5). The methods are, however, also applicable in the general case. The two algorithms which are to be described in this appendix have the same structure as the simple estimator (3.2:43). A description of the two general types of algorithms can for d=0 be found in Ljung and Söderström (1983).

Assume the plant is given by:

$$A(q^{-1})y(t) = \frac{B(q^{-1})}{F(q^{-1})} u(t) + \frac{C(q^{-1})}{H(q^{-1})} e(t) + d \qquad (1)$$

where

$$
\begin{aligned}
A(q^{-1}) &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \\
B(q^{-1}) &= \quad\; b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \\
C(q^{-1}) &= 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c} \\
H(q^{-1}) &= 1 + h_1 q^{-1} + \ldots + h_{n_h} q^{-n_h} \\
F(q^{-1}) &= 1 + f_1 q^{-1} + \ldots + f_{n_f} q^{-n_f}
\end{aligned}
\qquad (2)
$$

and where e(t) is white noise sequence. The plant parameters are assumed to be constant. Introduce the signals:

$$\nu(t) = A(q^{-1})y(t) - \omega(t) - d \qquad (3)$$

$$\omega(t) = \frac{B(q^{-1})}{\underline{F}(q^{-1})} u(t) \qquad \eta(t) = \frac{H(q^{-1})}{C(q^{-1})} \nu(t) \qquad (\ 4\ )$$

Then the difference equation, (1), can be written:

$$y(t) = [1-A(q^{-1})]y(t) + B(q^{-1})u(t) + [1-\underline{F}(q^{-1})]\omega(t) +$$

$$+ [C(q^{-1})-1]e(t) + [1-H(q^{-1})]\nu(t) + d + e(t)$$
$$(\ 5\ )$$

$$= \phi^{T}(t)\theta + e(t)$$

where

$$\theta = [a_1, \ \ldots \ ,b_1, \ \ldots \ ,\underline{f}_1, \ \ldots \ ,c_1, \ \ldots \ ,h_1, \ \ldots \ ,d]^{T}$$
$$(\ 6\ )$$

$$\phi(t+1) = [-y(t), \ \ldots \ ,u(t), \ \ldots \ ,-\omega(t), \ \ldots$$

$$,\eta(t), \ \ldots, \ -\nu(t), \ \ldots \ ,1]^{T}$$

Now, define the signals:

$$\alpha(t) = - \frac{H(q^{-1})}{C(q^{-1})} y(t) \qquad\qquad \beta(t) = \frac{H(q^{-1})}{C(q^{-1})\underline{F}(q^{-1})} u(t)$$

$$\delta = \frac{H(q^{-1})}{C(q^{-1})} 1 \qquad\qquad \gamma(t) = - \frac{H(q^{-1})}{C(q^{-1})\underline{F}(q^{-1})} \omega(t)$$
$$(\ 7\ )$$

$$\kappa(t) = \frac{1}{C(q^{-1})} \eta(t) \qquad\qquad \xi(t) = - \frac{1}{C(q^{-1})} \nu(t)$$

The one step prediction:

$$\hat{y}(t) = \hat{y}(t|t-1) = \phi^T(t)\theta \qquad (8)$$

can also be written as a recursion:

$$C(q^{-1})\underline{F}(q^{-1})\hat{y}(t) = \underline{F}(q^{-1})[C(q^{-1})-A(q^{-1})H(q^{-1})]y(t) +$$
$$\qquad (9)$$
$$B(q^{-1})H(q^{-1})u(t) + \underline{F}(q^{-1})H(q^{-1})d$$

This recursion may be used for determining the gradient of the prediction error:

$$\varepsilon(t) = y(t) - \hat{y}(t) \qquad (10)$$

It can be shown (see Ljung and söderström (1983) for details) that:

$$\nabla_\theta \varepsilon(t) = - \psi(t) \qquad (11)$$

where:

$$\psi(t+1) = [\alpha(t), \ldots ,\beta(t), \ldots ,\gamma(t), \ldots$$
$$\qquad (12)$$
$$,\kappa(t), \ldots ,\zeta(t), \ldots ,\delta]^T$$

The coefficients in the polynomials, A, B, C, D, F, i.e the elements in θ, can be estimated by the recursive prediction error method which is given by the recursions:

$$\varepsilon(t+1) = y(t+1) - \hat{\phi}(t+1)^T \hat{\theta}(t)$$

$$\hat{\theta}(t+1) = \hat{\theta}(t) + P(t+1)\hat{\psi}(t+1)\varepsilon(t+1) \qquad (13)$$

$$P^{-1}(t+1) = P^{-1}(t) + \hat{\psi}(t+1)\hat{\psi}^T(t+1)$$

The vectors, $\hat{\phi}$ and $\hat{\psi}$, are obtained by substitution of the unknown coeffients in the filters, (3), (4) and (7)

by their estimates. If a recursive pseudo linear regression method is applied the recursion becomes:

$$\varepsilon(t+1) = y(t+1) - \hat{\phi}(t+1)^T \hat{\theta}(t)$$

$$\hat{\theta}(t+1) = \hat{\theta}(t) + P(t+1)\hat{\phi}(t+1)\varepsilon(t+1) \qquad\qquad (14)$$

$$P^{-1}(t+1) = P^{-1}(t) + \hat{\phi}(t+1)\hat{\phi}^T(t+1)$$

Notice, except for the definition of ϕ and ψ, that the recursions in (13) and (14) are identical with the recursions (3.2:43) for the simple case. Consequently it is possible to extend the methods described in this thesis and make them applicable in the general case.

A2. NEWTON AND GRADIENT RECURSIONS.

In this appendix the basic recursions, on which the recursive estimators are based, are to be described briefly. The appendix ends with a recursion for the evolution of the criterium evaluated in the actual estimate.

Consider the following problem of determining the decision variable, $\theta(t)$, in such a manner that:

$$I[t,\theta(t)] = 0 \qquad\qquad\qquad (\ 1\ )$$

where:

$$I(s+1,\theta) = \lambda(s+1)I(s,\theta) + \psi(y(s+1),\theta)$$
$$\qquad\qquad\qquad (\ 2\ )$$

$$0 \leqslant s < t, \quad I(0,\theta) = 0$$

Forgetting has been introduced in order to discount old information.

The problem in finding the solution to (2) might have its origin as the necessary condition in the unbounded minimizing problem:

$$\hat{\theta}(t) = \arg \underset{\theta}{\text{Min}}\ J(t,\theta) \qquad\qquad\qquad (\ 3\ )$$

where:

$$J(s+1,\theta) = \lambda(s+1)J(s,\theta) + \rho(y(s+1),\theta)$$
$$\qquad\qquad\qquad (\ 4\ )$$

$$0 \leqslant s < t, \quad J(0,\theta) = 0$$

and where $\rho(t,\theta)$ is a scalar (pseudo) concav function of θ. The connection between the two problems excist if:

$$I(s,\theta) = \nabla_\theta \; J(s,\theta)$$

$$\psi(y(s),\theta) = \nabla_\theta \; \rho(y(s),\theta)$$

(5)

If the plant parameters, θ, are to be estimated by a least squares scheme then:

$$\rho(s,\theta) = \frac{1}{2} [\; y(s) - \phi^T(s)\theta \;]^2$$

(6)

and

$$\psi(y(s),\theta) = -\; \phi(s)[\; y(s) - \phi^T(s)\theta \;]$$

(7)

This choise for ρ is also used in the maximum likelihood case if the distribution of the innovation is gaussian.

Newton Recursions.

The Newton method is also denoted as the adaptive Newton-Raphson method. The deriviation of the algorithm has its origin in the non recursive Newton method. Now, define:

$$A(y(s),\theta) = \nabla_\theta \psi(y(s),\theta)$$

(8)

The non recursive iterations for determination of $\theta(t+1)$ is then:

$$\theta_{k+1} = \theta_k - M(t+1,\theta_k)^{-1} I(t+1,\theta_k)$$

(9)

where

$$I(s+1,\theta) = \lambda(s+1)I(s,\theta) + \psi(y(s+1),\theta)$$

$$M(s+1,\theta) = \lambda(s+1)M(s,\theta) + A(y(s+1),\theta)$$

(10)

$$0 \leqslant s < t+1, \qquad I(0,\theta) = 0, \qquad M(0,\theta) = 0$$

In the non recursive method all t+1 observations are directly used in the determination of $\hat{\theta}$(t+1). In the recursive method a data reduction is introduced and the decision variable, x(t+1) in the recursive method, is only directly determined by the last observation, y(t+1), and the last estimate, x(t). The solution, θ(t) and the approximative solution, x(t), are only identical in special cases. Shortly, the recursive Newton algorithm can be regarded as a method in which one iteration is performed for every new observation. This corresponds to approximate the criterium, J(t+1,θ), with a quadratic criterium.

If the criterium, J(t+1,θ), is quadratic in θ, a Newton method will give the stationary point in one iteration. If this is not the case the following two approximations have to be assumed.

1) It assumed that the starting point, x(t), for this single iteration is satisfying:

$$I(t,x(t)) = 0 \qquad\qquad (\ 11\ )$$

Then:

$$x(t+1) = x(t) - M(t+1,x(t))^{-1}\psi(y(t+1),x(t)) \quad (\ 12\ )$$

If $\psi(t,\theta)$ is linear in θ, ($\rho(t,\theta)$ is quadratic in θ), then this assumption is true. Additionally, the assumption is reasonable if the solution, x(t), is close to the exact solution, θ(t) (and I(t,θ) is continuous).

2) The quantity, M(t,x(t)) is involved in the determination of M(t+1,x(t)) but only M(t,x(t-1)) is known. The matrix, M(t+1,x(t)) is then approximated by the sequence, H(t+1) given by the recursion:

$$H(t+1) = \lambda(t+1)H(t) + A(y(t+1),x(t)) \qquad (\ 13 \ )$$

This approximation is valid if the variations in x(t) are small or if the dependence of θ in A(t+1,θ) is small.

With the mentioned approximations the Newton recursions become:

$$x(t+1) = x(t) - H(t+1)^{-1}\psi(y(t+1),x(t))$$

$$(\ 14 \ )$$

$$H(t+1) = \lambda(t+1)H(t) + A(y(t+1),x(t))$$

Gradient Recursions.

The stationary point, θ(t+1), can also be determined by a gradient method which in the non recursive version can be obtained by the iterations:

$$\theta_{k+1} = \theta_k - c_k I(t+1,\theta_k) \qquad (\ 15 \ )$$

where c_k is a decreasing positive sequence. The gradient recursions can be derived in the same manner as the Newton method. If the approximation, 1), is valid the recursions are:

$$x(t+1) = x(t) - c(t+1)\psi(y(t+1),x(t)) \qquad (\ 16 \ )$$

where the choice:

$$c(t) = \frac{1}{t} \qquad (\ 17 \ )$$

is frequently seen. Exponential forgetting can be introduced by choosing:

$$C(t) = \frac{1}{T(t)} \qquad (\ 18 \ )$$

where

$$T(t+1) = \lambda(t+1)T(t) + 1 \qquad\qquad (19)$$

With this choice the gradient recursions become.

$$x(t+1) = x(t) - \frac{\psi(y(t+1),x(t))}{T(t+1)}$$

$$(20)$$

$$T(t+1) = \lambda(t+1)T(t) + 1$$

The Evolution of the Criterium.

Now focus on the evolution of the criterium when it is evaluated in the approximative stationary point, $x(t)$. It is assumed that $x(t)$ is obtained by the Newton recursions and that $J(t+1,\theta)$ and $\rho(t+1,\theta)$ are quadratic in θ. The recursions which are to be stated in the sequel are consequently valid under the same assumptions as the Newton recursions.

Accordingly, let:

$$\rho(t,\theta) = \underline{c}(t) + \underline{b}^T(t)\theta + \frac{1}{2}\theta^T A(t)\theta$$

$$(21)$$

$$J(t,\theta) = \underline{J}(t) + J_1^T(t)\theta + \frac{1}{2}\theta^T H(t)\theta$$

where $H(t)$ is symmetric and positive definite and where $A(t)$ is symmetric and positive semidefinite. Since $x(t)$ is a stationary point:

$$I(t,x(t)) = J_1 + H(t)x(t) = 0 \qquad\qquad (22)$$

and

$$J(t,x(t)) = \underline{J}(t) - \frac{1}{2}x^T(t)H(t)x(t) \qquad\qquad (23)$$

Now:

$$\underline{J}(t+1) = \lambda(t+1)\underline{J}(t) + \underline{c}(t+1) \qquad (\ 24\ )$$

Using the Newton recursions the result:

$$-\frac{1}{2}\ x^T(t+1)H(t+1)x(t+1) = -\frac{1}{2}\ \lambda(t+1)x^T(t)H(t)x(t)$$

$$+\ \frac{1}{2}\ x^T(t)A(t+1)x(t) \qquad (\ 25\ )$$

$$+\ \frac{1}{2}\ \psi^T(t+1,x(t))H^{-1}(t+1)\psi(t+1,x(t))$$

$$-\ x^T(t)\psi(t+1,x(t))$$

is obtained. Since $\rho(t,\theta)$ is given by (21) and:

$$\psi(t,\theta) = \underline{b}(t) + A(t)\theta \qquad (\ 26\ )$$

the criterium has the evolution given by:

$$J(t+1,x(t+1)) = \lambda(t+1)J(t,x(t)) + \rho(t+1,x(t))$$

$$(\ 27\ )$$

$$-\ \frac{1}{2}\ \psi^T(t+1,x(t))H^{-1}(t+1)\psi(t+1,x(t))$$

A3 THE VARIABLE FORGETTING METHOD.

The basic idea in the variable forgetting method (see section 4.1.4) which is proposed in Fortescue et al. (1981) is to determine the forgetting factor in such a manner that the criterium, evaluated in the estimate, is constant, i.e.:

$$J[t+1,\hat{\theta}(t+1)] = J[t,\hat{\theta}(t)] = J_o \qquad (1)$$

Now apply the least squares criterium which is given by:

$$J(t+1,\theta) = \lambda(t+1)J(t,\theta) + \rho(t+1,\theta)$$

$$\rho(t,\theta) = \frac{1}{2} [y(t) - \phi^T(t)\theta]^2 \qquad (2)$$

Since $\rho(t,\theta)$ is quadratic in θ the results obtained in appendix 2 can be applied here directly. The recursions for the estimate, $\hat{\theta}(t)$, are:

$$\varepsilon(t+1) = \varepsilon(t+1|t) = y(t+1) - \phi(t+1)^T\hat{\theta}(t) \qquad (3)$$

$$\hat{\theta}(t+1) = \hat{\theta}(t) + P(t+1)\phi(t+1)\varepsilon(t+1) \qquad (4)$$

$$P^{-1}(t+1) = \lambda(t+1)P^{-1}(t) + \phi(t+1)\phi^T(t+1) \qquad (5)$$

and the evolution of the criterum evaluated in $\hat{\theta}(t)$ is according to (A2:27) given by:

$$J[t+1,\hat{\theta}(t+1)] = \lambda(t+1)J[t,\hat{\theta}(t)]$$

$$\qquad (6)$$

$$+ \frac{1}{2} [1 - \phi^T(t+1)P(t+1)\phi(t+1)]\varepsilon^2(t+1)$$

Introduce the quantities:

$$\mu = \phi^{T}(t+1)P(t)\phi(t+1)$$

$$\gamma = \frac{1}{2} \frac{\varepsilon^{2}(t+1)}{J_{0}}$$
(7)

Due to (6) the design objective (1) gives that:

$$\lambda^{2}(t+1) - [1-\mu-\gamma] \lambda(t+1) - \mu = 0$$
(8)

or:

$$\lambda(t+1) = \frac{1}{2} \{ 1-\mu-\gamma + \sqrt{(1-\mu-\gamma)^{2} + 4\mu} \}$$
(9)

Since:

$$\lambda(t+1) = 1 \qquad \text{for } \gamma = 0$$

$$\nabla_{\gamma} \lambda(t+1) = -\frac{1}{1+\mu} \qquad \text{for } \gamma = 0$$
(10)

the expression for $\lambda(t+1)$ can for small values of γ ($\varepsilon^{2}(t+1) \ll 2J_{0}$) be approximated with:

$$\lambda(t+1) \simeq 1 - \frac{\gamma}{1+\mu} \qquad \text{for } \gamma \ll 1$$
(11)

The design variable J_{0} can be transformed into:

$$J_{0} = \frac{1}{2} N_{0}\sigma_{0}^{2}$$
(12)

where σ_{0}^{2} is the innovation variance and N_{0} is an asymptotic horizon which becomes the new tuning parameter. Then the expression, (11) becomes:

$$\lambda(t+1) \simeq 1 - \frac{1}{N_0} \frac{\varepsilon^2(t+1)}{[1+\phi^T(t+1)P(t)\phi(t+1)]\sigma_o^2} \qquad (13)$$

as stated in chapter 4.1.4.

A4. THE DETECTOR LIMITS.

In this appendix it will be investigated in which way the test principle used in chapter 5.4 is dividing the sample space into subsets. In order to compare with the detectors which are based on the wellknown maximum likelihood ratio test, the investigation is also caried out for the mean value and the variance detectors described in section 5.4.1 - 5.4.3.

Let ξ_i, i=1, ..., N, be N independent and identically distributed stochasitc variables and let:

$$\xi_i \ \varepsilon \ N(\mu,\sigma^2)$$

$$(\mu,\sigma) \ \varepsilon \ \Omega \ = \ \{ \ \mu,\sigma \ | \ \sigma > \sigma_0 \ \}$$

(1)

Define the vector:

$$\underline{\xi} = (\ \xi_1, \ ... \ ,\xi_n \ )^T$$

(2)

Due to the assumption on independency the joint density function, $f(\underline{\xi};\mu,\sigma)$, is given by:

$$f(\underline{\xi};\mu,\sigma) = \prod_{i=1}^{N} \phi_N[\xi_i;\mu,\sigma]$$

(3)

where ϕ_N is the Gaussian density function. The aim of the detectors is to dicriminated between the 4 situations defined by:

$$S_0: \ (\mu,\sigma) \ \varepsilon \ \Omega_0$$
$$S_1: \ (\mu,\sigma) \ \varepsilon \ \Omega_1$$
$$S_2: \ (\mu,\sigma) \ \varepsilon \ \Omega_2$$
$$S_3: \ (\mu,\sigma) \ \varepsilon \ \Omega_3$$

(4)

where:

$$\Omega_0 = \{\ \mu,\sigma\ |\ \mu=0,\ \sigma=\sigma_0\ \}$$
$$\Omega_1 = \{\ \mu,\sigma\ |\ \mu\neq0,\ \sigma=\sigma_0\ \}$$
$$\Omega_2 = \{\ \mu,\sigma\ |\ \mu=0,\ \sigma>\sigma_0\ \}$$
$$\Omega_3 = \{\ \mu,\sigma\ |\ \mu\neq0,\ \sigma>\sigma_0\ \}$$

(5)

Let $f_i(\underline{\xi})$ be the principal element from the set of density functions which correspond to Ω_i and let:

$$\underline{L}_i(\underline{\xi}) = \log[\ f_i(\underline{\xi})\ ] + \frac{N}{2}\log(2\pi) + K_i$$

(6)

The test principle, (5.4.4:44) states that the hypothesis, H_i, which corresponds to the situation, S_i, is accepted if the inequality:

$$\underline{L}_i(\underline{\xi}) - \underline{L}_j(\underline{\xi}) \geqslant C > 0$$

(7)

holds for all values of $j\neq i$. Introduce the signals:

$$\underline{M} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\sigma_0} \qquad\qquad \underline{\underline{M}} = N\ \underline{M}$$

$$\underline{Z} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i^2}{\sigma_0^2} \qquad\qquad \underline{\underline{Z}} = N\ \underline{Z}$$

(8)

and the quantity:

$$\gamma = \frac{\mu}{\sigma_0}$$

then $\log[f(\underline{\xi};\gamma,\sigma)]$ are:

$$\log[f(\underline{\xi};\gamma,\sigma)] = -\frac{N}{2}\log[2\pi\sigma^2] - \frac{N}{2}[\underline{Z}-2\gamma\underline{M}+\gamma^2]\frac{\sigma^2}{\sigma_0^2}$$

(9)

Notice, that $\underline{Z} \geqslant \underline{M}^2$.

The ML Detector.

This detector is based on a discriminator which is characterised by the following choise of principal likelihood function.

$$f_i(\underline{\xi}) = \sup_{(\mu,\sigma)\varepsilon\Omega_i} f(\underline{\xi};\mu,\sigma) \qquad (\ 10\ )$$

Elementary maximaztions techniques give that:

$$\underline{L}_0(\underline{\xi}) = -\frac{N}{2}\log(\sigma_o^2) - \frac{N}{2}\underline{Z} + K_0 \qquad (\ 11\ )$$

$$\underline{L}_1(\underline{\xi}) = -\frac{N}{2}\log(\sigma_o^2) - \frac{N}{2}[\underline{Z}-\underline{M}^2] + K_1 \qquad (\ 12\ )$$

$$\underline{L}_2(\underline{\xi}) = \begin{cases} -\frac{N}{2}\log(\underline{Z}\sigma_o^2) - \frac{N}{2} + K_2 & \text{for } \underline{Z} \geqslant 1 \\[2ex] -\frac{N}{2}\log(\sigma_o^2) - \frac{N}{2}\underline{Z} + K_2 & \text{for } \underline{Z} < 1 \end{cases} \qquad (\ 13\ )$$

$$\underline{L}_3(\underline{\xi}) = \begin{cases} -\frac{N}{2}\log[(\underline{Z}-\underline{M}^2)\sigma_o^2] - \frac{N}{2} + K_3 & \text{for } \underline{Z} \geqslant 1+\underline{M}^2 \\[2ex] -\frac{N}{2}\log(\sigma_o^2) - \frac{N}{2}[\underline{Z}-\underline{M}^2] + K_3 & \text{for } \underline{Z} < 1+\underline{M}^2 \end{cases} \qquad (\ 14\ )$$

Now, define the function, $\underline{f}(x)$, by:

$$\underline{f}(x) = x - 1 - \log(x) \ , \ x > 0 \qquad (\ 15\ )$$

and g(x) by:

$$x = g(x) - 1 - \log(g(x)), \ g(x) \geqslant 1 \qquad (\ 16\ )$$

These two functions are ploted in figure 1.

The acceptance area for the hypothesis, H_i, can be determined
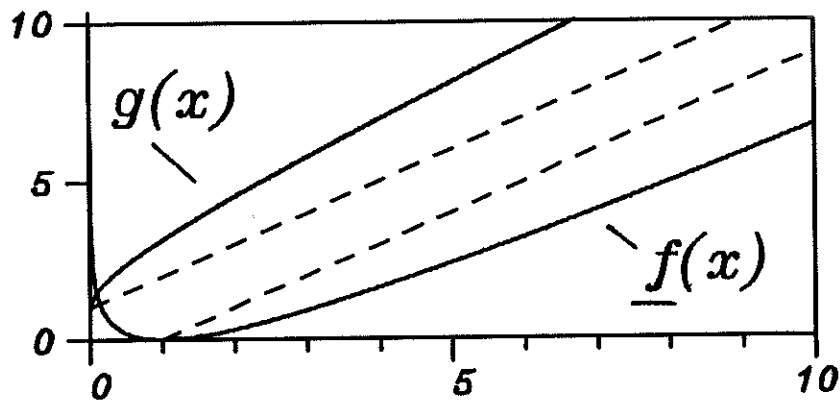
Figure 1: The function $f(x)$ and $g(x)$.

as the interception of 3 basic regions defined by:

$$\underline{L}_i(\underline{\xi}) - \underline{L}_j(\underline{\xi}) \geqslant C \qquad j \neq i$$

These basic regions are acceptance in a test between H_i and H_j. Consequently, the limits which define the acceptance areas can be determined by investigating the equation:

$$\underline{L}_i(\underline{\xi}) - \underline{L}_j(\underline{\xi}) = \pm C \qquad\qquad\qquad (17)$$

In the investigation it is assumed that the choice of K_i is proper, i.e. such that a solution to (17) exists. The investigation is refered in the following and the results are illustrated in figure 2.

$\underline{H_1 \text{ and } H_0}$

$$\underline{L}_1(\underline{\xi}) - \underline{L}_0(\underline{\xi}) = -\frac{N}{2}\underline{M}^2 + K_1 - K_0 = \pm C$$

$$(18)$$

$$\underline{M}^2 = -\frac{2}{N}[K_0 - K_1 \pm C]$$

$\underline{H}_2$ and $\underline{H}_0$

$$\underline{L}_2(\underline{\xi}) - \underline{L}_0(\underline{\xi}) = -\frac{N}{2} \underline{f}(\underline{Z}) + K_2 - K_0 = \pm C$$

$$\underline{Z} = g\{-\frac{2}{N}[K_0 - K_2 \pm C]\}$$

(19)

$\underline{H}_3$ and $\underline{H}_0$

for $\underline{Z} \geqslant 1 + \underline{M}^2$:

$$\underline{L}_3(\underline{\xi}) - \underline{L}_0(\underline{\xi}) = -\frac{N}{2} \underline{f}(\underline{Z} - \underline{M}^2) + \frac{N}{2} \underline{M}^2 + K_3 - K_0 = \pm C$$

$$\underline{Z} = g\{\frac{2}{N}[K_0 - K_3 \pm C] - \underline{M}^2\} + \underline{M}^2$$

(20)

for $\underline{Z} < 1 + \underline{M}^2$:

$$\underline{L}_3(\underline{\xi}) - \underline{L}_0(\underline{\xi}) = -\frac{N}{2} \underline{M}^2 + K_3 - K_0 = \pm C$$

$$\underline{M}^2 = -\frac{2}{N}[K_0 - K_3 \pm C]$$

(21)

$\underline{H}_2$ and $\underline{H}_1$

for $\underline{Z} \geqslant 1$:

$$\underline{L}_2(\underline{\xi}) - \underline{L}_1(\underline{\xi}) = -\frac{N}{2} \underline{f}(\underline{Z}) - \frac{N}{2} \underline{M}^2 + K_2 - K_1 = \pm C$$

$$M^2 = -\frac{2}{N}[K_2 - K_1 \mp C] + \underline{f}(\underline{Z})$$

(22)

for $\underline{Z} \geqslant 1$:

$$\underline{L}_2(\underline{\xi}) - \underline{L}_1(\underline{\xi}) = -\frac{N}{2} \underline{M}^2 + K_2 - K_1 = \mp C$$

$$M^2 = -\frac{2}{N}[K_2 - K_1 \mp C]$$

(23)

$\underline{H}_3$ and $\underline{H}_1$

$$L_3(\underline{\xi}) - L_1(\underline{\xi}) = -\frac{N}{2} f(\underline{Z} - \underline{M}^2) + K_3 - K_1 = \pm C$$

$$\underline{Z} = g\{-\frac{2}{N}[K_1 - K_3 \pm C]\} + \underline{M}^2 \tag{24}$$

$\underline{H_3}$ and $\underline{H_2}$

for $\underline{Z} \geqslant 1 + \underline{M}^2$:

$$L_3(\underline{\xi}) - L_2(\underline{\xi}) = -\frac{N}{2} \log[1 - \frac{\underline{M}^2}{\underline{Z}}] + K_3 - K_2 = \pm C \tag{25}$$

$$\underline{M}^2 = \underline{Z}[1 - \exp\{-\frac{2}{N}[K_3 - K_2 \mp C]\}]$$

for $1 \leqslant \underline{Z} \leqslant 1 + \underline{M}^2$:

$$L_3(\underline{\xi}) - L_2(\underline{\xi}) = -\frac{N}{2} f(\underline{Z}) + \frac{N}{2} \underline{M}^2 + K_3 - K_2 = \pm C \tag{26}$$

$$\underline{M}^2 = -\frac{2}{N}[K_2 - K_3 \pm C] + f(\underline{Z})$$

for $\underline{Z} < 1$:

$$L_3(\underline{\xi}) - L_2(\underline{\xi}) = -\frac{N}{2} \underline{M}^2 + K_3 - K_2 = \pm C \tag{27}$$
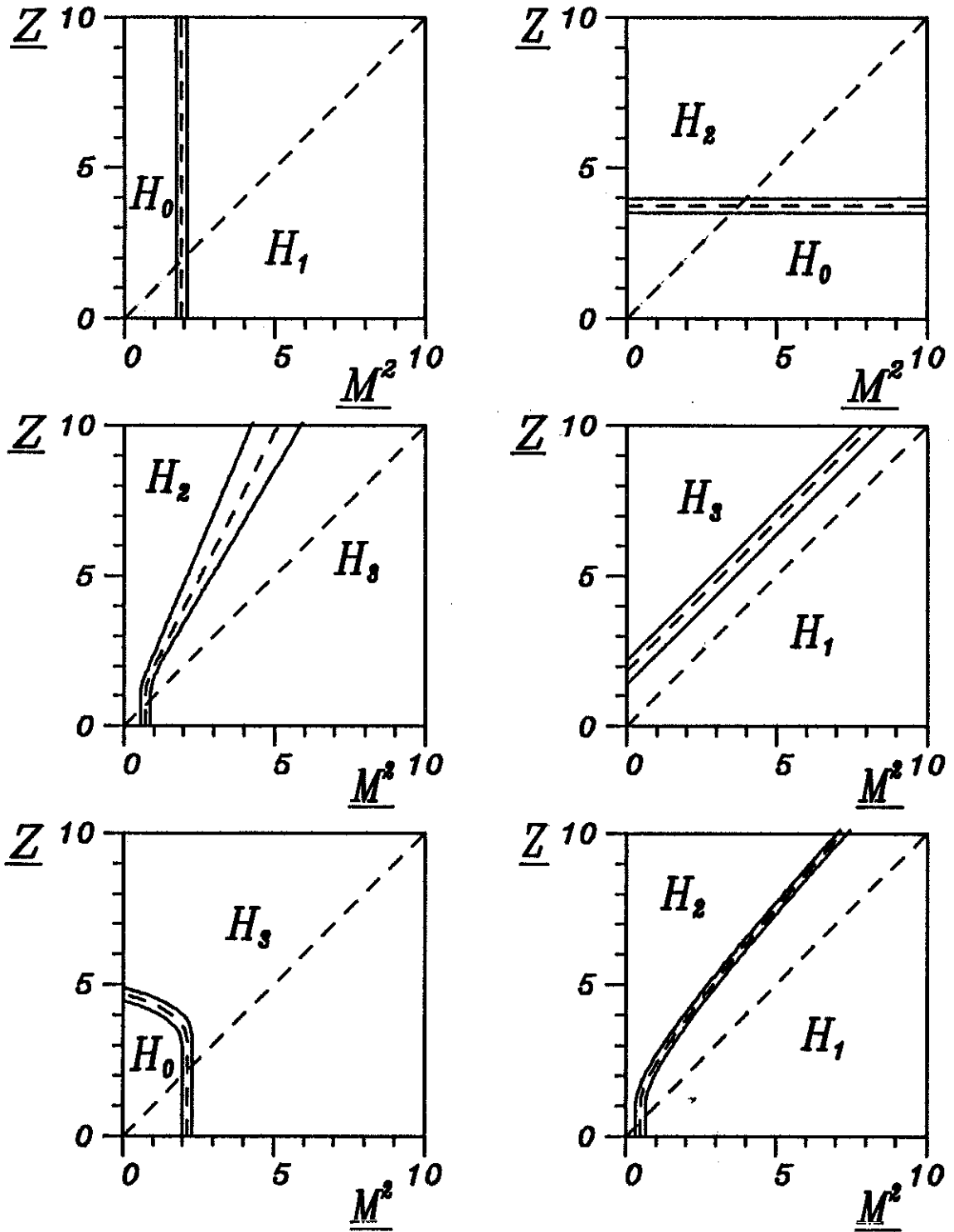
$$\underline{M}^2 = -\frac{2}{N}[K_2 - K_3 \pm C]$$

Figure 2: The detector limits (18) − (27) for N=6, C=0.5 $K_o=0$, $K_1=-5.75$, $K_2=-4.27$, $K_3=-6.42$.
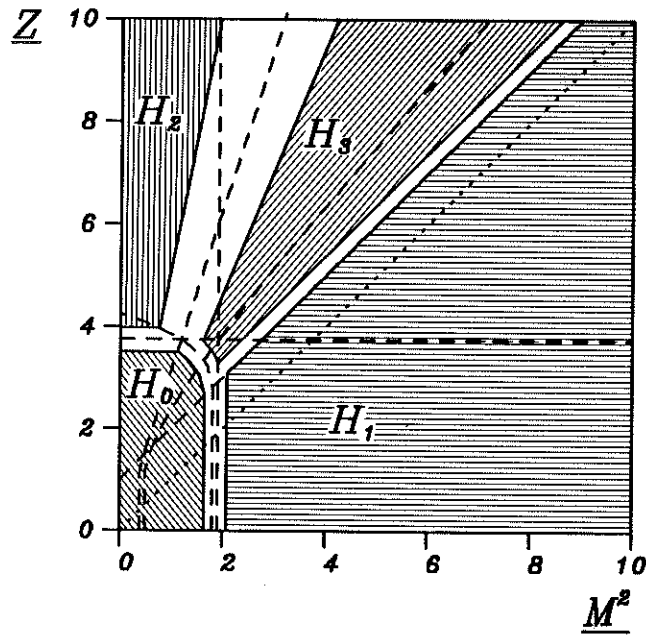
Figure 3: The acceptance area for N=6, C=0.5, K_1=-5.75 K_2=-4.27 K_3=-5.42 and K_0 = 0.

The MLR - detectors.

The maximum likelihood ratio test principle only concerns test between one hypothesis, H_i, and one alternative, H_j. This is the major difference between the MLR and the ML testprinciple in (5.3:44) and (5.3:46).

The MLR testprinciple states that the hypothesis, H_i, is accepted and the alternative, H_j, is rejected if:

$$L_i(\underline{\xi}) - \tilde{L}_j(\underline{\xi}) \geqslant \underline{K}_j \qquad (28)$$

where

$$L_i(\underline{\xi}) = \log[\sup_{(\mu,\sigma)\varepsilon\Omega_i} f(\underline{\xi};\mu,\sigma)] \qquad (29)$$

Another difference between the two maximum likelihood principle is that in the ML principle the maximation is carried out with respect to the, Ω_j, which corresponds to the alternative, H_j ($j \neq i$). In the MLR principle the maximazation is carried out with respect to the whole model set, $\tilde{\Omega}$, (which especially includes Ω_i), i.e.

$$\tilde{L}_j(\underline{\xi}) = \log[\sup_{(\mu,\sigma)\varepsilon\tilde{\Omega}_j} f(\underline{\xi};\mu,\sigma)] \qquad (30)$$

In the mean value MLR test used in section 5.4.2 it is assumed that $\sigma=\sigma_o$ and thus:

$$\tilde{\Omega}_1 = \Omega_o \cup \Omega_1$$

and:

$$L_o(\underline{\xi}) = -\frac{N}{2}\log(\sigma_o^2) - \frac{N}{2}\underline{Z} \qquad\qquad (\;31\;)$$

$$\tilde{L}_1(\underline{\xi}) = -\frac{N}{2}\log(\sigma_o^2) - \frac{N}{2}[\underline{Z}-\underline{M}^2] \qquad\qquad (\;32\;)$$

and the detector limits is given by:

$$\underline{M}^2 = -\frac{2}{N}\;\underline{K}_1 \qquad\qquad (\;33\;)$$

In the variance MLR test where it is assumed that $\mu=0$:

$$\tilde{\Omega}_2 = \Omega_o\;\cup\;\Omega_2$$

$$L_o(\underline{\xi}) = -\frac{N}{2}\log(\sigma_o^2) - \frac{N}{2}\underline{Z} \qquad\qquad (\;34\;)$$

$$\tilde{L}_2(\underline{\xi}) = -\frac{N}{2}\log(\underline{Z}\sigma_o^2) - \frac{N}{2} + K_2 \quad \text{for } \underline{Z} \geqslant 1 \qquad (\;35\;)$$

and the detector limit is given by:

$$\underline{Z} = g\{-\frac{2}{N}\underline{K}_2\} \qquad\qquad (\;36\;)$$

The detector limits (18) and (19) from the ML test can be compared to (33) and (36). The two MLR test mentioned here is conditioned on known mean and variance, respectively. If the procedure just described above is carried out for the unconditioned test, then two detector limit which are similar to (24) and (25) occur. Actually, it is for normal chioce of K_i , i=0, ... , 3, only the limits (18), (19), (24), (25), (20) and (22) which are active in the ML detector.

The Cusum Detector.

The design of this detector is based on the choice:

$$f_i(\underline{\xi}) = f(\underline{\xi}; \mu_i, \sigma_i) \qquad\qquad (37)$$

for the principal likelihood function. Here μ_i and σ_i are predefined elements in Ω_i.

With the special choice:

$$\mu_0 = 0 \qquad\qquad \mu_1 = \pm\gamma\sigma_0$$

$$\sigma_0^2 = \sigma_0^2 \qquad\qquad \sigma_1^2 = \sigma_0^2$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad (38)$$

$$\mu_0 = 0 \qquad\qquad \mu_3 = \pm\gamma\sigma_0$$

$$\sigma_2^2 = \kappa\sigma_0^2 \qquad\qquad \sigma_3^2 = \kappa\sigma_0^2$$

of tuning parameters the results:

$$\underline{L}_0 = -\frac{N}{2}\log(\sigma_0^2) - \frac{N}{2}\underline{Z} + K_0 \qquad\qquad (39)$$

$$\underline{L}_1 = -\frac{N}{2}\log(\sigma_0^2) - \frac{N}{2}(\underline{Z}-2\gamma\underline{M}+\gamma^2) + K_1 \qquad\qquad (40)$$

$$\underline{L}_2 = -\frac{N}{2}\log(\kappa\sigma_0^2) - \frac{N}{2}\frac{\underline{Z}}{\kappa} + K_2 \qquad\qquad (41)$$

$$\underline{L}_3 = -\frac{N}{2}\log(\kappa\sigma_0^2) - \frac{N}{2}\frac{\underline{Z}-2\gamma\underline{M}+\gamma^2}{\kappa} + K_3 \qquad\qquad (42)$$

are obtained. As for the ML detector the acceptance areas can be determined by investigating the equation:

$$\underline{L}_i(\underline{\xi}) - \underline{L}_j(\underline{\xi}) = \pm C \qquad\qquad (43)$$

Direct calculations give the following results and is shown in figure 4.

$\underline{H}_1 - \underline{H}_0$:

$$\underline{M} = \frac{2}{N} \cdot \frac{K_0 - K_1 \mp C}{2\gamma} + \frac{\gamma}{2} \qquad (\ 44\ )$$

$\underline{H}_3 - \underline{H}_2$:

$$\underline{Z} = \kappa \cdot \frac{2}{N} \cdot \frac{K_2 - K_3 \mp C}{2\gamma} + \frac{\gamma}{2} \qquad (\ 45\ )$$

$\underline{H}_2 - \underline{H}_0$:

$$\underline{Z} = \frac{\kappa}{\kappa - 1} \left(\frac{2}{N}(K_0 - K_2 \mp C) + \log(\kappa) \right) \qquad (\ 46\ )$$

$\underline{H}_3 - \underline{H}_1$:

$$\underline{Z} = \frac{\kappa}{\kappa - 1} \left\{ \frac{2}{N}(K_1 - K_3 \mp C) + \log(\kappa) \right\} + 2\gamma\underline{M} - \gamma^2 \qquad (\ 47\ )$$

$\underline{H}_2 - \underline{H}_1$:

$$\underline{Z} = \frac{\kappa}{\kappa - 1} \left\{ \frac{2}{N}(K_1 - K_z \mp C) + \log(\kappa) + 2\gamma\underline{M} - \gamma^2 \right\} \qquad (\ 48\ )$$

$\underline{H}_3 - \underline{H}_0$:

$$\underline{Z} = \frac{\kappa}{\kappa - 1} \left\{ \frac{2}{N}(K_0 - K_3 \mp C) + \log(\kappa) - 2\frac{\gamma}{\kappa}\underline{M} + \frac{\gamma^2}{\kappa} \right\} \qquad (\ 49\ )$$

The acceptance areas are shown in figure 5 for the same values as in figure 4.

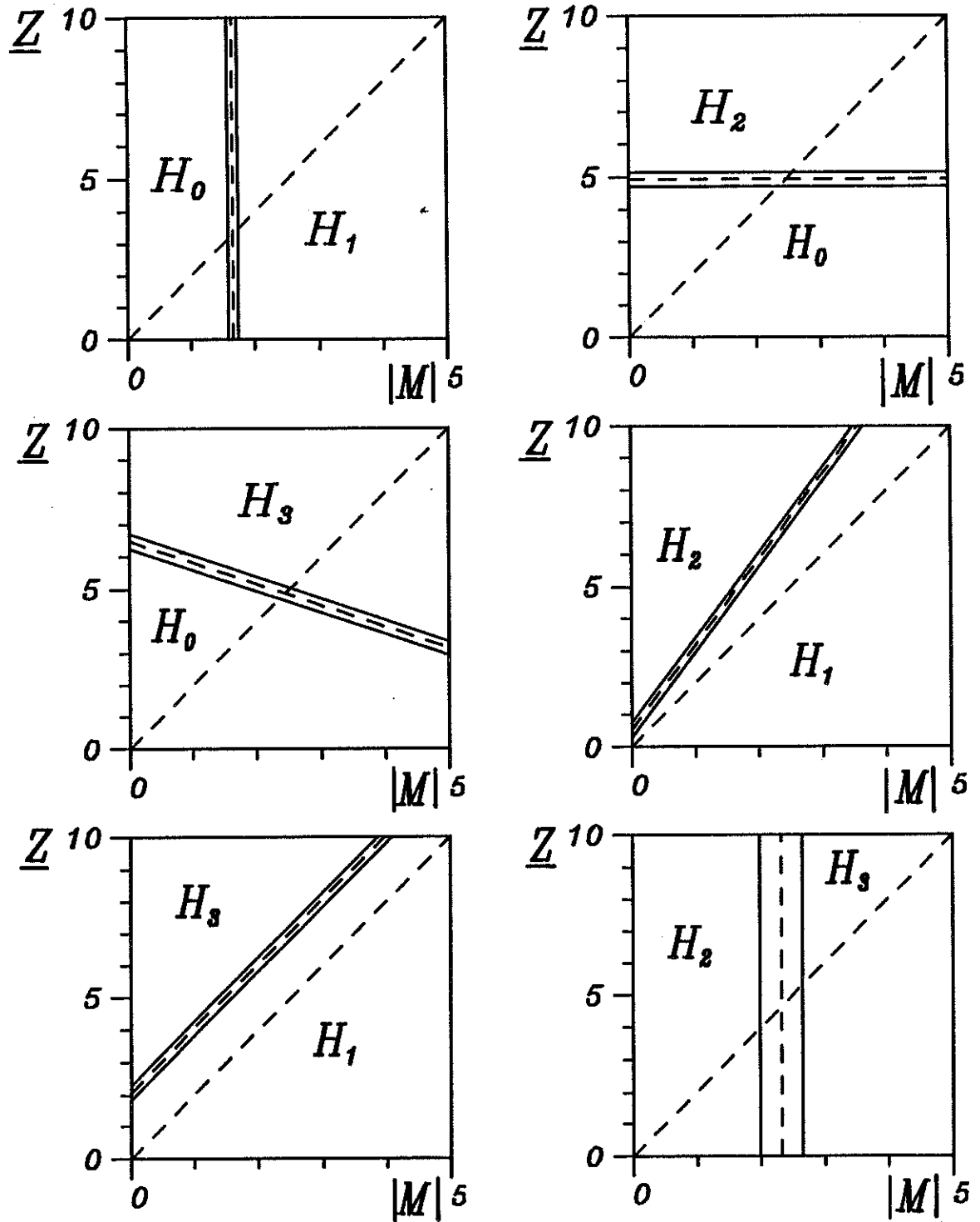**Figure 4:** The detector limits (44) - (50) for N=6, C=0.5, $K_1 = K_2 = 0$, $K_3 = -1.73$ and $K_0 = 6.91$.
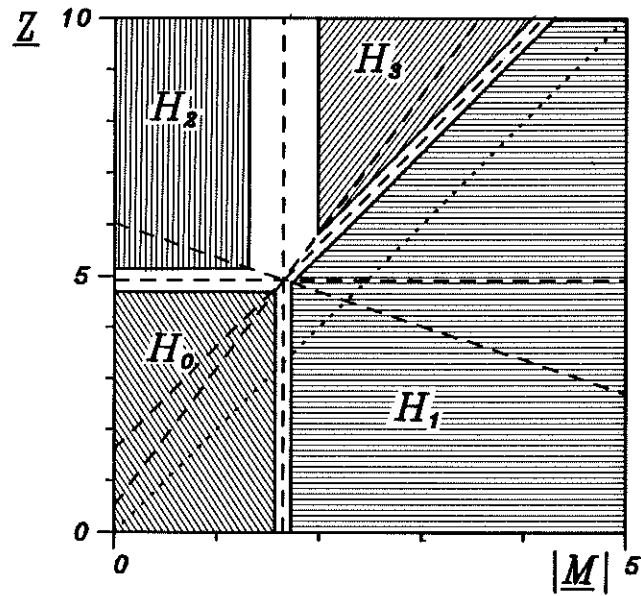
Figure 5: The acceptance areas for the same characteristic
values as in figure 4.

The Simple Cusum detectors, PRm and PRv.

The limit, (5.4:27), used in the mean value detector, PRm, is
embodied in (29) and the limit, (5.4:36), from the simple
detector, PRv, is a special case of (31).

A5. MODEL REDUCTION.

In this appendix a method for obtaining a least square estimate, $\hat{x}_o$, belonging to a model of reduced order, will be described.

Assume that the estimate, $\hat{x}$, belonging to the model:

$$Y = \Phi \hat{x} + \underline{\varepsilon} = [\; A \mid B \;] \begin{array}{c} \tilde{x}_o \\ \tilde{x}_1 \end{array} + \underline{\varepsilon} \qquad (\; 1 \;)$$

is available and that the estimates, $\hat{x}_o$, belonging to the reduced model:

$$Y = A \; \hat{x}_o + \varepsilon \qquad (\; 2 \;)$$

is desired. The two least squares estimate are given by:

$$\hat{x}_o = (A^T A)^{-1} A^T Y \qquad (\; 3 \;)$$

and:

$$\hat{x} = (\Phi^T \Phi)^{-1} \Phi^T Y$$

$$= \begin{bmatrix} A^T A & | & A^T B \\ \overline{B^T A} & | & \overline{B^T B} \end{bmatrix}^{-1} \begin{bmatrix} A^T Y \\ B^T Y \end{bmatrix} \qquad (\; 4 \;)$$

Now, let $\tilde{x}_1$ be a scalar. If the estimate $\hat{x}$ is obtained by a recursive least squares algorithm, then the matrix:

$$P(t) = (\Phi^T \Phi)^{-1} = \begin{bmatrix} Q & | & v \\ ------ \\ v^T & | & \mu \end{bmatrix} \qquad (\; 5 \;)$$

is directly available. The partitioning of $P(t)$ in Q, v and μ is done in accordance to the partitioning of $\hat{x}$ in $\hat{x}_0$ and $\tilde{x}_1$. If it is possible for the scalar case to find an equation for $\hat{x}_0$ and $(A^T A)^{-1}$ in the above terms, then a reduction can be carried out by sequentiel reduction of $\hat{x}$ by one elements at a time.

By elementary calculus (see. e.g. Kailath (1980)):

$$P(t) = \begin{bmatrix} a & | & c \\ \hline c^T & | & b \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} a^{-1} + a^{-1} c c^T a^{-1} \mu & | & -\mu a^{-1} c \\ \hline -c^T a^{-1} \mu & | & \mu \end{bmatrix}$$

(6)

where:

$$a = A^T A, \qquad b = B^T B, \qquad c = A^T B$$

(7)

$$\mu = \frac{1}{B^T [I - A(A^T A)^{-1} A^T] B}$$

Direct calculations give:

$$\hat{x}_0 = a^{-1} A^T Y$$

(8)

$$\tilde{x}_0 = \hat{x}_0 + \mu \, a^{-1} c \, [c^T a^{-1} A^T - B^T Y]$$

(9)

$$\tilde{x}_1 = \mu \, [B^T Y - c^T a^{-1} A^T Y]$$

(10)

or that:

$$\hat{x}_0 = \tilde{x}_0 + a^{-1} c \tilde{x}_1$$

(11)

In the available terms the equations become:

$$\hat{x}_0 = \tilde{x}_0 - \frac{v}{\mu} \tilde{x}_1 \tag{12}$$

and:

$$(A^T A)^{-1} = Q - \frac{v v^T}{\mu} \tag{13}$$

222

A6. THE DETECTOR ALGORITHM.

In this appendix the whole detector algorithm will be given. In the stated version the Cusum detector, described in chapter 5.4.6, has been applied. The normal situation is designed with respect to constant parameters.

In the normal situation the estimates:

$$\varepsilon_o(t+1) = \varepsilon_o(t+1|t) = y(t+1) - \phi(t+1)^T \hat{\theta}_o(t)$$

$$\hat{\theta}_o(t+1) = \hat{\theta}_o(t) + P_o(t+1)\phi(t+1)\varepsilon_o(t+1) \tag{1}$$

$$P_o^{-1}(t+1) = P_o^{-1}(t) + \phi(t+1)\phi^T(t+1)$$

$$r(t+1) = r(t) + \frac{1}{\Sigma(t+1)} \left[\frac{\varepsilon_o^2(t+1)}{\beta(t+1)} - r(t) \right]$$

$$\beta(t+1) = 1 + \phi^T(t+1)P_o(t+1|t)\phi(t+1)$$

$$\Sigma(t+1) = \lambda(t+1)\Sigma(t) + 1 \tag{2}$$

$$\lambda(t+1) = \lambda^o\lambda(t) + \lambda^1(1-\lambda^o)$$

and the detector signals:

$$Z(t+1) = \text{Max}[Z(t) + \frac{\varepsilon_r^2(t+1)}{\sigma_o^2} - \frac{\kappa}{\kappa-1} \log(\kappa) , 0]$$

$$M_z(t+1) = [M_z(t) + \frac{\varepsilon_r(t+1)}{\sigma_o}] \text{sign}(Z(t+1)) \tag{3}$$

$$N_z(t+1) = [N_z(t)+1] \; sign(Z(t+1))$$

$$M_+(t+1) = Max[M_+(t) + \frac{\varepsilon_r(t+1)}{\sigma_o} - \frac{\gamma}{2} , 0] \qquad (4)$$

$$N_+(t+1) = [N_+(t)+1] \; sign(M_+(t+1))$$

$$M_-(t+1) = Max[M_-(t) - \frac{\varepsilon_r(t+1)}{\sigma_o} - \frac{\gamma}{2} , 0] \qquad (5)$$

$$N_-(t+1) = [N_-(t)+1] \; sign(M_-(t+1))$$

are updated. Dependent on the applied method for estimating the regular varying plant parameters $P_o(t+1|t)$ is obtained from $P_o(t)$. If the the kalmfilter method is applied then $P_o(t+1|t)$ is given by (4.2.2:21). If all inequalitites:

$$M_+(t) < \frac{\log(B)}{\gamma} \qquad (6)$$

$$M_-(t) < \frac{\log(B)}{\gamma} \qquad (7)$$

$$Z(t) < 2 \frac{\kappa}{\kappa-1} \log(B) \qquad (8)$$

are satisfied, then H_o is accepted and nothing exceptional is carried out.

If only one of the inequalities (8) or (6) [or (7)] is violated, the hypothesis H_2 or H_1 are accepted, respectively. If two inequalities are violated, a discrimination between H_1 and H_3 has to be performed. If the inequality:

$$Z(t)-2\gamma|M_z(t)|+N_z(t)\gamma^2 < 2 \; \frac{1}{\kappa-1} \; \log(B) \qquad (9)$$

is satisfied then the hypothesis, H_1, is accepted. The estimate of the point of time, T_c, in which the change has occured is given by:

$$\hat{T}_c = t - N(t) \qquad (10)$$

where $N(t)$ is equal to $N_z(t)$, $N_+(t)$ or $N_-(t)$ corresponding to the accepted hypothesis. When one of the inequalities (6) - (8) are violated the corresponding states in the detector are put to zero.

If a change has been detected, i.e. the hypothesis H_m, $m=1,2,3$, has been accepted, a noise adaptive kalmanfilter which for $\hat{T}_c-1 \le s < t$ is given by the recursions:

$$\varepsilon_m(s+1) = y(s+1) - \phi^T(s+1)\hat{\theta}_m(s)$$

$$\hat{\theta}_m(s+1) = \hat{\theta}_m(s) + P_m(s+1)\varepsilon_m(s+1)$$

$$\qquad (11)$$

$$P_m(s+1|s) = P_m(s) + R_m(s+1)$$

$$P_m^{-1}(s+1) = P_m^{-1}(s+1|s) + \phi^T(s+1)\phi(s+1)$$

$$\underline{\varepsilon}(s+1) = \frac{\varepsilon_m^2(s+1)}{\sigma_o^2} - \beta(s+1) - \underline{x}^T(s+1)\hat{\underline{q}}_m(s)$$

$$\beta(s+1) = 1 + \phi^T(s+1)P_m(s+1|s)\phi(s+1)$$

$$\qquad (12)$$

$$\hat{\underline{q}}_m(s+1) = \hat{\underline{q}}_m(s) + \underline{P}_m(s+1)\underline{x}(s+1)\underline{\varepsilon}(s+1)$$

$$\underline{P}_m^{-1}(s+1) = \lambda(s+1) \; \underline{P}_m^{-1}(s) + \underline{x}^T(s+1)\underline{x}(s+1)$$

$$\underline{\lambda}(s+1) = \underline{\lambda}^o \underline{\lambda}(s) + \underline{\lambda}^1 (1-\underline{\lambda}^o)$$

is started in:

$$P_m(T_c-1) = P_o(T_c-1)$$

$$\hat{\theta}_m(T_c-1) = \hat{\theta}_o(T_c-1)$$

$$\hat{q}_m(T_c-1) = (\ 0,\ \ldots\ ,0)^T, \qquad \kappa_o = 1$$

$$\underline{\lambda}(T_c-1) = 0.9,\ \underline{\lambda}^o = 0.9,\ \underline{\lambda}^1 = 0.99 \qquad\qquad (\ 13\ )$$

$$\underline{P}_1(T_c-1) = \kappa_o \left\{ \begin{matrix} 0 & & & \\ & \cdot & & O \\ & & \cdot & \\ O & & 0 & \\ & & & 1 \end{matrix} \right\}$$

$$\underline{P}_2(T_c-1) = \kappa_o \left\{ \begin{matrix} 1 & & & \\ & \cdot & & O \\ & & \cdot & \\ O & & 1 & \\ & & & 0 \end{matrix} \right\}, \underline{P}_3(T_c-1) = \kappa_o \left\{ \begin{matrix} 1 & & & \\ & \cdot & & O \\ & & \cdot & \\ O & & \cdot & \\ & & & 1 \end{matrix} \right\}$$

i.e. in the result obtained by the normal mode just before the parameter change.

When the branch estimation, which is performed on stored data is finished (s=t), the algorithm is changed back to the normal mode and starts in the results obtained in the actual exceptional mode, i.e.

$$\hat{\theta}_o(t) = \hat{\theta}_m(t)$$

$$\qquad\qquad (\ 14\ )$$

$$P_o(t) = P_m(t)$$

A7. THE CONSTANT TRACE ALGORITHM.

In the constant trace algorithms the recursion for $P(t)$ is given by:

$$P^{-1}(t+1) = \lambda_1(t+1)P^{-1}(t) + \lambda_2(t+1)\phi(t+1)\phi^T(t+1) \qquad (1)$$

or:

$$P(t+1) = \left\{ P(t) - \frac{P(t)\phi(t+1)\phi^T(t+1)P(t)}{\gamma + \phi^T(t+1)P(t)\phi(t+1)} \right\} \frac{1}{\lambda_1} \qquad (2)$$

where the quantity:

$$\gamma(t+1) = \frac{\lambda_1(t+1)}{\lambda_2(t+1)} \qquad (3)$$

has been introduced. Direct caluculations give that:

$$L(t+1) = tr[P(t+1)]$$

$$= \left\{ L(t) - \frac{\phi^T(t+1)P^2(t)\phi(t+1)}{\gamma+\phi^T(t+1)P(t)\phi(t+1)} \right\} \frac{1}{\lambda_1} \qquad (4)$$

The design objective:

$$L(t+1) \geqslant L_0 \qquad (5)$$

results in the equation:

$$\lambda_1(t+1) = Min\left[\frac{L(t)-\beta(t+1)}{L_0}, 1 \right] \qquad (6)$$

$$\underline{g}(t+1) = \frac{\phi^T(t+1)P^2(t)\phi(t+1)}{\gamma + \phi^T(t+1)P(t)\phi(t+1)} \qquad\qquad (7)$$

which has been stated in (4.1:12).

A8. THE POWER FUNCTIONS.

The power functions of the ML detector were shown in figure 5.4.5:7-10. In this appendix the power functions will be given.

Firstly, the quantities, $\underline{M}$ and $\underline{Z}$, are transformed into the stochastic variable, $\tilde{M}$ and $\tilde{Z}$ given by:

$$\tilde{M}^2 = N\,\underline{M}^2\,r \qquad\qquad \tilde{Z} = N\,[\,\underline{Z} - \underline{M}^2\,]\,r \qquad\qquad (1)$$

where

$$r = \frac{\sigma_o^2}{\sigma^2} \qquad\qquad (2)$$

The stochastic variable, $\tilde{M}$ and $\tilde{Z}$ are mutually independent and:

$$\tilde{M}\ \varepsilon\ N[\ \frac{\sqrt{N}}{\sigma}\mu,\ 1\ ]$$

$$(3)$$

$$\tilde{Z}\ \varepsilon\ \chi^2(N-1)$$

Let $F(x)$ and $f(x)$ be the distribution and the density function for $\tilde{M}$ and let $G(x)$ and $g(x)$ be the distribution and density function for $\tilde{Z}$.

In the transformed variable the 4 important (for $K_i = \tilde{K}_i$, $C=0$) detector limits (5.4.5:49-52) are:

$$\tilde{M}^2 = \nu^2\,r \qquad\qquad (4)$$

$$\tilde{Z} + \tilde{M} = \beta\,r \qquad\qquad (5)$$

$$\tilde{Z} = (\beta - \nu^2) \ r \qquad\qquad (6)$$

$$\tilde{M}^2 = \frac{\nu^2}{\beta} \ \tilde{Z} \qquad\qquad (7)$$

which is illustrated in figure 1.
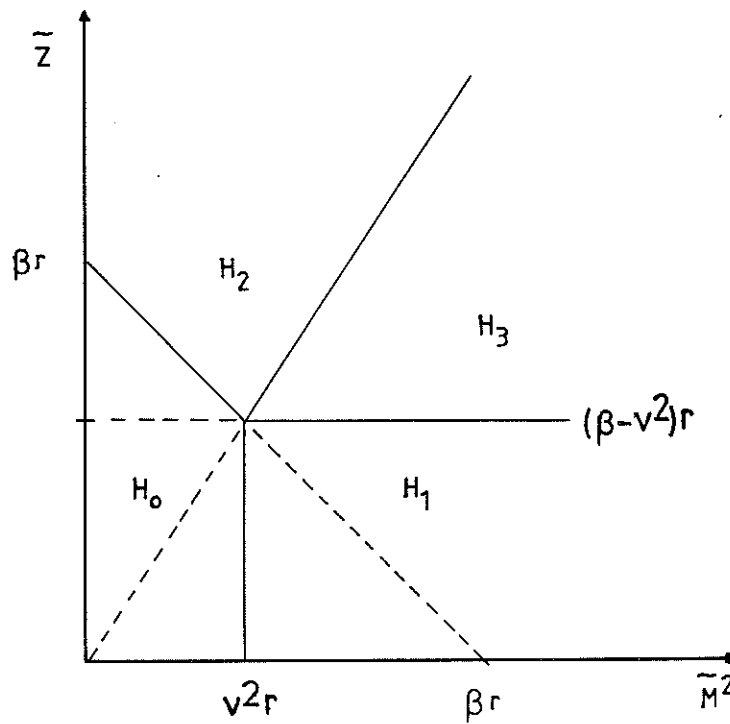
Figure 1: The acceptance areas in the transformed variables.

The probability, $p(H_i)$, of accepting H_i can be determined as a function of σ and μ.

$$p(H_1) = \{ \ F(-\nu\sqrt{r}) + 1 - F(\nu\sqrt{r}) \ \} G[(\beta - \nu^2)r] \qquad\qquad (8)$$

$$p(H_0) = \int_{-\nu\sqrt{r}}^{\nu\sqrt{r}} f(x) G(\beta r - x^2) dx \qquad\qquad (9)$$

$$p(H_2) = \int_{-\infty}^{\infty} f(x)[1 - G(\delta(x))] dx \qquad\qquad (10)$$

where

$$\delta(x) = Max\{ \beta r - x^2, \frac{\beta}{\nu^2} x^2 \}$$ (11)

and:

$$p(H_3) = 1 - p(H_0) - p(H_1) - p(H_2)$$ (12)

A9. AN APPLICATION OF BAYES THEOREM.

Assume that the scalars, y and r, and the vector θ are stochastic variables and that:

$$y|r \ \varepsilon \ N(\mu,r) \tag{1}$$

$$\theta|r \ \varepsilon \ N(\theta_o,Pr) \tag{2}$$

$$r \ \varepsilon \ R\Gamma(\alpha,\gamma) \tag{3}$$

where $N(\mu,r)$ is the gaussian distribution with mean and variance equal to μ and r, respectively. P is a known matrix. $R\Gamma(\alpha,\gamma)$ is the reciproc gamma distribution with the parameters, α and γ. It can be shown (see e.g. Thyregod (1978)) that

$$y \ \varepsilon \ T(2\alpha,\mu,-\frac{\gamma}{\alpha}) \tag{4}$$

$$\theta \ \varepsilon \ T(2\alpha,\theta_o,P\frac{\gamma}{\alpha}) \tag{5}$$

and:

$$r|y \ \varepsilon \ R\Gamma[\ \alpha + \frac{1}{2}\ ,\ \gamma + \frac{1}{2}\ (y-\mu)^2\ ] \tag{6}$$

where $T(a,b,c)$ is the T – distribution with the location paramter, b, and the scale parameter, c. Furthermore, due to the T – distribution:

$$E\{\ \theta\ \} = \theta_o \tag{7}$$

and the reciprocal gamma distribution:

$$E\{ r \} = \frac{r}{\alpha - 1} \qquad\qquad (8)$$

9. References.

==

Albert, A. and Sittler, R. W. (1966): A method for computing
    least squares estimators that keep up with the data.
    SIAM journal on Control, ser. A., 3, No. 3,
    pp. 384-417.

Andersson, P. (1983): Adaptive Forgetting in Recursive
    Identification through Multiple Models. Report
    LiTH-ISY-0638, Department of Electical Engineering,
    Linköping University, Linköping, Sweden.

Basseville, M. and Benveniste, A. (1983): Sequential Detection
    of Abrupt Changes in Spectral Characteristics of
    Digital Signals. IEEE trandactions on Information
    Theory, 29, No.5, pp. 709-724, 1983

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

Clarke, D.W and Gawthrop, B.A. (1975): Self tuning controller.
    IEE proc., 122, pp. 929-934.

Dumont, G. (1982): Self tuning control of a chip refiner moter
    load. Automatica, 18, pp. 307-314.

Egardt, B. (1979): Stability of adaptive controllers.
    Springer, Berlin.

Evans, R. J. and Betz, R. E. (1982): New Results and
    Applications of Adaptive Control to Classes of
    Nonlinear Systems.
    Preprints pp. 281-310, Workshop on Adaptive Control,
    Oct. 27-29, 1982, Florence, Italy.

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

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

Hinkley, D. V. (1971): Inference About the Change-point from
        Cumulative Sum Test.
        Biometrica, 58, 3, pp. 509-523

Holst J. and Poulsen N. K. (1984): Self Tuning Control of
        Plants with Abrupt Changes.
        Preprint vol. VII pp. 144-149, 9th World Congress of
        the International Federation of Automatic Control,
        Budapest, Hungary, 1984.

Huber, P.T. (1964): Robust Statistics. Wiley, New York

Hägglund, T. (1983): New Estimation Techniques for Adaptive
        Control.
        Ph. D. Dissertation: LUTFD2/(TFRT-1025)/1-120/(1983),
        Department of Automatic Control, Lund Instistute of
        Technology, Lund, Sweden.

Isermann, R. (1977): Digitale Regelsysteme.
        Springer, New York.

Jazwinski, A. H. (1970): Stochastic Processes and Filtering
        Theory.
        Academic Press, New York.

Kailath, T. (1980): Linear Systems. Printice Hall, Englewood
        Cliff.

Kemp, K. W. (1962): The Use of Cumulative Sums for Sampling
         Inspection Schemes.
         Applied Statistics, 11, pp. 16-31

Kucera, V. (1979): Discrete Linear Control.
         Wiley, New York.

Lainiotis, D. G. (1976a): Partitioning: A unifying framework
         for Adaptive Systems I : Estimation.
         Proc. IEEE 64, pp. 1126-1143

Lainiotis, D. G. (1976b): Partitioning: A unifying framework
         for Adaptive Systems II : Control.
         Proc. IEEE 64, pp. 1182-1198

Landau, I. D. (1981): Unification of Discrete Time explicit
         Model Reference Adaptive Control design.
         Automatica, 17, pp. 593-611

Landau, I. D. (1982): Deterministic and Stochastic adaptive
         pole - zero placement for minimum - phase Systems.
         Preprints pp. 207-244, Workshop on Adaptive Control,
         Oct. 27-29, 1982, Florence, Italy.

Ljung, L. (1977): Asymptotic behavior of the Extended Kalman
         filter as a parameter Estimator for Linear Systems.
         IEEE Transactions on Automatic Control, 24, pp 36-50

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

Lozano, R. (1982): Independent Tracking and Regulation Adaptive
         Control with Forgetting Factor.
         Automatica, 18, pp. 455-459, 1982.

Millnert, M. (1982): Identification and Control of Sytems
        Subject to Abrupt Changes. Ph. D. Dissertation, Dept
        of Electrical Engineering, Linköping University,
        Linköping, Sweden.

Poljak, B.T. and Tsypkin, J.Z. (1980): Robust Identification.
        Automatica, 16, pp.53-63.

Poulsen N. K. (1982): Adaptive Control of Non stationary
        Processes (in Danish). IMSOR, The Technical
        University of Denmark

Poulsen N. K. and Holst J. (1982): Robust Self tuning
        Controllers in Nonstationary Situations.
        Preprints pp. 257-280, Workshop on Adaptive Control,
        Oct. 27-29, 1982, Florence, Italy.

Rasmussen H. and Pedersen T. S. (1983): Self Tuning Controllers
        Applied to Steam Superheaters.
        Proceedings pp. 18.1-18.5, IASTED Symposium, Applied
        Control and Identification, 1983, Copenhagen, Denmark

Thyregod, P. (1978): Statistical Decision Theory.
        IMSOR, The Technical University of Denmark,
        Lyngby, Denmark.

Wellstead, P. E. and Sanoff, S. P. (1981): Extended Self
        tuning Algorithm.
        International Journal on Control, 34.3, pp. 433-455

Willsky, A. S. (1976): A survey of design methods for failure
        detection in Dynamic Systems.
        Automatica, 12, pp. 601-611

Wittenmark, B. (1979): A two-level estimator for time varying

parameters.
Automatica, 15, pp.85-89.

Xianya, X. and Evans, R. J. (1984): Discrete-time Adaptive
Control for Deterministic Time-varying Systems.
Automatica, 20, pp. 309-319

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

Aström, K. J. (1980): Design Principle for Self Tuning
Regulators.
Methods and Application in Adaptive Control.
Lecture Notes in Control and Information Science,
pp. 1 - 20. Springer, Berlin.

Aström, K. J. and Zhou Zhao-ying (1982): A Linear Quadratic
Gaussian Self Tuner.
Preprint pp. 1-20, Workshop on Adaptive Control,
Oct. 27-29, 1982, Florence, Italy.

Aström, K. J. (1983): Theory and Application of adaptive
Control - A survey.
Automatica, 19, pp. 471-486

Licentiatafhandlinger ved IMSOR

Nr. 1. SIGVALDASON, HELGI: "Beslutningsproblemer ved et Hydro-Termisk Elforsyningssystem". 1963. 92 pp.

Nr. 2. NYGAARD, JØRGEN: "Behandling af et dimensioneringsproblem i telefonien". 1966. 157 pp.

Nr. 3. KRARUP, JAKOB: "Fixed-cost and other network flow problems". 1967. 159 pp.

Nr. 4. HANSEN, NIELS HERMAN: "Problemer ved forudsigelse af lydhastighed i danske farvande". 1967. I+II, 199 pp.

Nr. 5. LARSEN, MOGENS E.: "Statistisk analyse af elementære kybernetiske systemer". 1968. 210 pp.

Nr. 6. LAL, AMRIT PUNHANI: "Decision Problems in Connection with Atomic Power Plants". 1968. 133 pp.

Nr. 7. CLAUSEN, SVEND: "Kybernetik - Systemer og Modeller". 1969. 206 pp.

Nr. 8. VIDAL, R.V. VALQUI: "Operations Research in Production Planning. Interconnections between Production and Demand". 1970. I+II, 322 pp.

Nr. 9. BILDE, OLE: "Nonlinear and Discrete Programming in Transportation, Location and Road Design". 1970. I+II, 291 pp.

Nr. 10. RASMUSEN, HANS JØRGEN: "En decentraliseret planlægningsmodel". 1972. 185 pp.

Nr. 11. DYRBERG, CHR.: "Tilbudsgivning i en entreprenørvirksomhed". 1973. 158 pp.

Nr. 12. MADSEN, OLI B.G.: "Dekomposition og Matematisk Programmering". 1973. 271 pp.

Nr. 13. DAHLGAARD, PETER: "Statistical aspects of tide prediction". 1973. I+II, 385 pp.

Nr. 14. SPLIID, HENRIK: "En statistisk model for stormflodsvarsling". 1973. 205 pp.

Nr. 15. PINOCHET, MARIO: "Operations Research in Strategic Transportation Planning. The Decision Process in a Multiharbour System". 1973. 388 pp.

Nr. 16. CHRISTENSEN, TORBEN: "Om Semi-Markov processer". 1973. 240 pp.

Nr. 17. JACOBSEN, SØREN KRUSE: "Om lokaliseringsproblemer: Modeller og løsninger". 1973. 360 pp.

Nr. 18. MARQVARDSEN, HANS: "Skemalægning ved numerisk simulation". 1973. 222 pp.

Nr. 19. MORTENSEN, JENS HALD: "Interregionale Godstransporter". 1974. 223 pp.

Nr. 20. MELO, JUAN SEVERIN: "Introduction to Operations Research in Systems Synthesis". 1974. 249 pp.

Nr. 21 BUNDGAARD-JØRGENSEN, UFFE
og 22 SPLIID, IBEN: "Skitse til en procedure for kommunalplanlægning". 1974. 544 pp.

Licentiatafhandlinger ved IMSOR

Nr. 23 MOSGAARD, CHRISTIAN: "International Planning in Disaster Situations". 1975. 187 pp.

Nr. 24 HOLM, JAN: "En optimeringsmodel for kollektiv trafik". 1975. 246 pp.

Nr. 25 JENSSON, PÁLL: "Stokastisk Programmering. Del I: Modeller. Del II: Metodologiske Overvejelser og Anvendelser". 1975. 340 pp.

Nr. 26 IVERSEN, VILLY BÆK: "On the Accuracy in Measurements of Time Intervals and Traffic Intensities with Application to Teletraffic and Simulation". 1976. 206 pp.

Nr. 27 DRUD, ARNE: "Methods for control of complex dynamic systems". 1976. 209 pp.

Nr. 28 TOGSVERD, TOM: "Koordinering af kommunernes ressourceforbrug. 1976. 275 pp.

Nr. 29 JENSEN, OLAV HOLST: "Om planlægning af kollektiv trafik". 1976. 320 pp.

Nr. 30 BEYER, JAN E.: "Ecosystems. An operational research approach. 1976. 315 pp.

Nr. 31 BILLE, THOMAS BASTHOLM: "Vurdering af Egnsudviklingsprojekter". 1977. 264 pp.

Nr. 32 HOLST, ERIK: "En statistisk undersøgelse af tabletserier". 1979. 316 pp.

Nr. 33 AAGAARD-SVENDSEN, ROLF: "Econometric Methods and Kalman Filtering". 1979. 300 pp.

Nr. 34 HANSEN, STEEN: "Project Control by Quantitative Methods". 1979. 232 pp.

Nr. 35 SCHEUFENS, ERNST EDVARD: "Statistisk analyse og kontrol af tidsafhængige vandkvalitetsdata". 1980. 152 pp.

Nr. 36 LYNGVIG, JYTTE: "Samfundsøkonomisk planlægning". 1981. 252 pp.

Nr. 37 TROELSGÅRD, BIRGITTE: "Statistisk bestemmelse af modeller for rumlufttemperatur". 1981. 213 pp.

Nr. 38 RAFT, OLE: "Delivery Planning by Modular Algorithms". 1981. 220 pp.

Nr. 39 JENSEN, SIGRID M.: "Analyse af interregionale togrejser". 1981. 216 pp.

JENSEN, SIGRID M.: "Analyse af interregionale togrejser", figurer og appendices. 1981. 176 pp.

240

Licentiatafhandlinger ved IMSOR

Nr. 40.   RAVN, HANS: "Technology and Underdevelopment". The case of Mexico. 1982. 380 pp.

Nr. 41.   HANSEN, STEN: " Phase-type Distributions in Queueing Theory". 1983. 209 pp.

Nr. 42.   FERREIRA, José A.S.: "Optimal Control of Discrete-Time Systems with Applications". 1984. 252 pp.

Nr. 43.   BEHRENS, JENS CHRISTIAN: "Mathematical Modelling of Aquatic Ecosystems applied to Biological Waste Water Treatment. 1985.

           BEHRENS, JENS CHR. Mathematical Modelling of Aquatic Ecosystems applied to Biological Waste Water Treatment. Appendix I. 392 pp. Appendix II. 172 pp.

Nr. 44.   POULSEN, NIELS KJØLSTAD: "Robust Self Tuning Controllers. 1985. 225 pp.