1988

Online Detection and Estimation of Parameter Jumps.

Fahmida Nilufar Chowdhury
Louisiana State University and Agricultural & Mechanical College

Order Number 8917807

On-line detection and estimation of parameter jumps

Chowdhury, Fahmida Nilufar, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1988

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

ON-LINE DETECTION AND ESTIMATION OF PARAMETER JUMPS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
In partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Electrical and Computer Engineering

by
Fahmida Nilufar Chowdhury
M.Sc., Moscow Power Engineering Institute, 1980
December 1988

Acknowledgements

This work is dedicated to the memory of my grandfather, who was my earliest source of inspiration.

I am grateful to the staff of the Department of Electrical and Computer Engineering for their friendliness and cooperation. Special thanks to Mr. Steve Schmeckpepper for providing technical assistance.

This work was developed under the guidence of Dr. Jorge Aravena, who freely gave advice and read the manuscript countless times. I consider myself fortunate to have worked with such a perfectionist, and thank him deeply. Thanks are also due to Dr. Jimmie Lawson, Dr. Hong Lee, Dr. Ali Mirbod and Dr. Mine Yucel for reviewing the manuscript. Dr. William Porter, who has been a continuous source of inspiration throughout my stay here, helped improve the style of this manuscript considerably. I want to express my deep appreciation for his advice and help.

Heartfelt thanks go to my husband Prosanta Chowdhury, without whose sacrifice and understanding this work would not have been possible. I must also thankfully mention the name of Ms. Fahmida Malik, the greatest teacher I have ever had. Finally, I want to express eternal gratitude to my parents, Mr. Zahurul Islam and Mrs. Shams-Un-Nahar Islam, for their unconditional love and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

In many dynamic systems, the parameters are subject to jumps at unknown points in time. The jumps may be the result of some change in the operating system, or plant failure. In this work a method has been developed for fast detection of these jumps and estimation of the post-jump values. The proposed method is called the Combined Filter Algorithm. It is based on the conventional Kalman filter. Single-sample hypothesis test is used to determine the presence or absence of jumps. Small jumps, missed by the hypothesis test, are traced by a Random Walk Model Kalman filter. This is made possible by the introduction of a new kind of decision rule, called the Combined Decision Rule. The properties of this filter under correct and incorrect decisions are studied, and mathematical proofs are presented.

The intended main application of the algorithm is the detection of power system faults and estimation of steady-state voltages and currents. The proposed algorithm is tested for ARMA models and discrete state models. The algorithm is also tested on a continuous-time system, where a jump occurs in one parameter. The algorithm is applied to detect and estimate the resulting changes produced in an equivalent ARMA model. Simulation results are presented.

Chapter I

INTRODUCTION

Summary: This chapter introduces the general problem of parameter estimation, existing methods, current interests and associated difficulties. Possible applications are mentioned. Justification for using the Kalman filter as the basic tool is given, and an overview of the subsequent chapters is included.

I.1 The Problem of Parameter Estimation

Parameter estimation is the name given to a class of problems where unknown quantities are to be estimated based on some available information. Usually this information is obtained from the measured output of the system. This class is part of the general problem of extracting information from received or observed signals. The signals are, in general, corrupted by various disturbances or noise. The noise can be the effect of various causes, such as thermal, electrical, atmospheric, channel, or in some cases even intentional interferences. The areas of application include geophysics, remote sensing, acoustics, meteorology, medical imaging systems, electric power systems, electronic warfare, navigation, radar, sonar and telephone communications systems.

Some of the main difficulties that are usually encountered in the parameter estimation process are:

i) lack of suitable statistical descriptions for the signal and noise processes,

ii) incomplete or incorrect mathematical models,

iii) unobservabilities in the system,

iv) unpredicted sudden changes in the state variables,

v) unpredicted sudden changes in the noise process (hidden nonstationarity),

vi) errors in the detection of a change

and *vii)* delay in the detection of a change.

This dissertation deals mainly with the special case of fast detection of abruptly changing parameters, and the estimation of their new values. Out of the many estimation techniques available, the Kalman filter is chosen as the principal tool. The justification for this is given in the next section where an overview of the general estimation techniques is also presented.

The principal area of application is chosen to be electric power systems, where the variables to be estimated are voltages and currents. They can change abruptly as the result of faults (short circuits) in the system. In the case of a fault it is desirable to estimate, as quickly as possible, the post-fault steady-state values of the voltages and currents. These estimates are used to pinpoint the fault location. Thus, fast detection and estimation techniques are required in computer relaying of power systems. In chapter V this application is elaborated, and results of computer simulation are given.

I.2 Basic Estimation Techniques

The three most widely used estimation techniques are Bayesian, maximum likelihood, and least squares methods. The Bayesian estimator is developed by minimizing "Bayes' risk", which is the expected cost of an error in estimation. The Bayesian estimator can be subdivided into maximum aposteriori (MAP) and minimum variance of error (MV) estimators. They minimize different cost functions and the estimation equations are different. However, it has been established that for a very large class of estimation problems the optimal estimate is unique, regardless of which meaningful error criterion is used. Reference [1] presents a comprehensive account of these estimators.

The main obstacle in using the Bayesian technique is that it requires apriori knowledge of the probabilistic description of the unknown quantities. If Z is the observation and X is the quantity to be estimated, then the probability distributions Pr(Z/X) and Pr(X) must be known in order to construct a Bayesian estimator.

When Pr(X) is unknown, but Pr(Z/X) is known, then the maximum likelihood (ML) estimator can be constructed. It maximizes the likelihood function, which is the aposteriori conditional probability of the observation given the estimate of the parameter. Theoretically, the ML estimator is in general inferior to the Bayesian estimator. However, it is well known ([1],[2]) that if the parameters and noises are Gaussian, then the ML estimate is the same as the MAP and MV estimates.

If the requirement of the knowledge of Pr(Z/X) is removed, then the linear minimum variance (LMV) estimator can be designed. However, it must be assumed that the first two moments, i.e. the mean μ_x and

variance σ_x^2, are known. Since in the Gaussian case the first two moments give the complete probabilistic description, the LMV estimator coincides with both versions of the Bayesian estimator. Also, if it is assumed that $\mu_x = 0$ and $\sigma_x^{-1} = 0$, then the MV estimator degenerates into the ML estimator. It is interesting to note [49] that for a large class of (linear, Gaussian) problems, all the estimators discussed so far are essentially the same.

The least squares approach is the final step in the continual relaxation of statistical requirements. In the Bayesian method the complete probabilistic structure and costs must be known. In the least squares method the estimation problem is treated as a deterministic optimization problem.

For the linear observation

$$Z = H^t X + V$$

with more measurements than unknowns, and the cost function

$$J(\hat{X}) = \frac{1}{2}[\, Z - H^t\hat{X}]^t\, R^{-1}[\, Z - H^t\hat{X}],$$

the least squares estimate is given by

$$\hat{X} = [\, H\, R^{-1}H^t]^{-1}H\, R^{-1}Z.$$

In [2] it is shown that the least squares and LMV estimators will be identical if *i)* the matrix R in the cost function stands for the covariance matrix of the measurement noise, and *ii)* in the LMV estimator it is assumed that $\mu_x = 0$ and $\sigma^{-1} = 0$, which simply states that nothing is known about X. Thus, the least squares estimator can be an optimum Bayesian estimator.

The next step in this discussion is the transition from batch least squares to recursive least squares. Recursive methods are convenient for

an on-line implementation of the estimator. The application of interest in this dissertation is real time parameter estimation; therefore the recursive least squares method is chosen. In this method the variances of the estimation error are minimized at each step using the previous best estimates and the previous covariance of error matrix. The estimates are unbiased, and approach the true values as more and more data are processed.

In this dissertation the name "filter" is used as a synonym for "estimation technique". The least squares filter has been constrained to be linear in the estimates. It is known that the optimal linear filter is not necessarily the best filter among the class of all nonlinear functions. However, in [2] Sage and White show that for a general class of criteria, including the minimum error variance criterion, the expected mean conditioned on the data is the best estimate. For the Gaussian case, the expected mean conditioned on the data is linear in the data. As a result, for the Gaussian case, the best linear filter is also the best filter out of the class of all non-linear functions.

It is well known that the Kalman filter is *the* optimal recursive least squares algorithm, thus it is also the best filter among all the linear and nonlinear functions for the Gaussian case. In this dissertation a new algorithm for the case of abruptly changing parameters is developed and discussed. The measurement and process noises are assumed to be Gaussian. Hence the Kalman filter is used as the basic tool for this task.

I.3 The Kalman Filter for Estimating States and Parameters

It is well known that for a time invariant dynamic system, the Kalman filter gives an increasingly accurate estimate of the state vector as additional information becomes available. The norm of the error covariance matrix monotonically decreases with the absorption of new measurement data. The method can be extended to include joint estimation of parameters and states. Since the problems of parameter and state estimations are closely related, the algorithm can be modified to estimate parameters.

The discrete Kalman filter determines the best estimate of the state for the stochastic system

$$X_{k+1} = \Phi_k X_k + W_k, \qquad (1.1)$$

$$Z_k = H_k X_k + V_k, \qquad (1.2)$$

where

Φ_k - state transition matrix,

Z_k - measured output vector,

X_k - state vector,

H_k - observation matrix, and

V_k, W_k - vectors of normal random variables.

The vectors V_k and W_k satisfy the following:

$$E(W_k W_j^t) = S_k \delta_{k-j},$$

$$E(V_k V_j^t) = R_k \delta_{k-j},$$

$$E(W_k) = 0,$$

$$E(V_k) = 0,$$

$$\text{where } \delta_{k-j} = \begin{cases} 1, & \text{for } k = j \\ 0, & \text{for } k \neq j. \end{cases}$$

The matrices Φ, H, S and R are all assumed to be known functions of time. However, the coefficient matrices can also be known functins of past input and output variables. The same filter is still valid, since the past inputs and outputs are known at estimation time. This fact is very important in parameter estimation. It makes equation (1.1) available for describing the dynamics of parameter variations. Eq.(1.2) can then be used for describing the process dynamics. Thus, the solution algorithm is the same for both state and parameter estimation problems. However, the theoretical interpretation and practical calculation of the coefficient matrices are quite different.

Consider the following time-invariant ARMA model:

$$Z_k = \sum_{i=1}^{\nu} -A_i Z_{k-i} + \sum_{i=1}^{\mu} B_i U_{k-i} + V_k \qquad (1.3)$$

where, in general, Z is mx1, V is mx1, and the coefficients $A_1 \bullet\bullet\bullet A_\nu$, $B_1 \bullet\bullet\bullet B_\mu$ are scalars, with $\nu + \mu = n$.

This model can be reformulated to fit the usual notation. for the Kalman filter algorithm. Equation (1.3) can be shown to be equivalent to eq.(1.2), where the state vector X consists of the parameters A and B of the ARMA model. The observation vector H consists of past inputs U and outputs Z, while V stands for measurement noise. Hence,

$$X_k^t = [\ A_1 \bullet\bullet\bullet A_\nu\ B_1 \bullet\bullet\bullet B_\mu\ ],$$

$$Z_k = H_k^t X_k + V_k;$$

with

$$H_k^t = [-Z_{k-1} | \cdots | -Z_{k-\nu} | U_{k-1} | \cdots | U_{k-\mu}]. \quad (1.4)$$

The following discussion is valid regardless of whether the H matrix is the usual known observation matrix, or it comes from the ARMA model. The only requirement is that it be known at each time step of the on-line implementation.

I.3.1 Different Cases of Parameter Variation

The various ways in which system parameters can behave with time can be categorized in the following classes:

i) constant,

ii) slowly varying,

iii) abruptly changing,

iv) mixed.

These four cases are reviewed below. The next chapter develops a combined algorithm to estimate these parameters. The main emphasis is on abruptly changing (jump) parameters.

i) Constant Parameters

If the parameters can be considered constant, then eq.(1.1) simplifies to the following:

$$X_{k+1} = X_k. \quad (1.6)$$

The estimation problem is solved efficiently with the ordinary Kalman filter algorithm. The main equations of the estimation and updating

process are:

$$\text{estimate: } \hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k(Z_k - H_k^t \hat{X}_{k-1}); \qquad (1.7)$$

where

$$\text{gain: } K_k = Q_{k-1}H_k[R_2 + H_k^t Q_{k-1}H_k]^{-1}, \qquad (1.8)$$

$$\text{updated covariance: } Q_k = (I - K_k H_k^t)Q_{k-1}, \qquad (1.9)$$

$$\text{projected estimate: } \hat{X}_{k+1/k} = \hat{X}_k,$$

Because of the constant parameter model, no separate equations for estimation and prediction are required.

ii) Slowly Varying Parameters

Suppose the parameter change is described by a Gauss-Markov stochastic difference equation:

$$X_k = \Phi_k X_{k-1} + \Gamma_k W_{k-1}, \qquad (1.10)$$

where Φ is an nxn transition matrix and Γ is an nxm input matrix. Both Φ and Γ can be time varying. If the variations can be modeled, i.e., if matrices Φ and Γ are known, then the usual Kalman filter yields efficient and consistent estimates. However, when the variations can not be modeled, one must search for other devices. This problem has been well studied, and approached with many different techniques. Recursive estimation using Kalman-type least squares filters has been widely investigated {[13], [15], [20], [21]}. The most general way of handling this problem is to shape the memory of the filter in a desired way to remove the effects of obsolete data. This is sometimes called finite

memory filtering or moving window filtering.The shape of the memory can either be rectangular or exponential.

In its simplest form, a sliding window Kalman filter can be described by the following equations (due to Peter Young, in [13]). Notations have been changed to be consistent with this work.

Receipt of new data at k^{th} instant:

$$\hat{X}_{kn} = \hat{X}_{k-1} + K_k[\, Z_k - H_k^t \hat{X}_{k/k-1}], \tag{1.11}$$

with

$$K_{kn} = Q_{k-1}H_k[\, H_k^t\, Q_{k-1}H_k + R_{2k}]^{-1} \tag{1.12}$$

and

$$Q_{kn} = [I - K_k H_k^t]Q_{k-1}. \tag{1.13}$$

Removal of old data received at $(k-s)^{th}$ instant:

$$\hat{X}_k = \hat{X}_{kn} + K_k[Z_{k-s} - H_{k-s}^t \hat{X}_{kn}], \tag{1.14}$$

with

$$K_k = Q_{kn}H_{k-s}[\, H_{k-s}^t\, Q_{kn}\, H_{k-s} - R_{2k}]^{-1} \tag{1.15}$$

and

$$Q_k = [I - K_k H_{k-s}^t]Q_{kn}. \tag{1.16}$$

If no apriori knowledge is available about the possible nature of the parameter changes,then the finite memory filter can be expected to perform reasonably well for all kinds of changes that might take place. However, the noise in the estimates increases with short memory lengths. In addition, a finite memory filter requires much more computations than

a standard one. In [13], [14] and [20] other algorithms are treated in detail. Their numerical complexity is higher than the new approach developed in this dissertation.

iii) Abruptly Changing Parameters

It is possible that the parameters are subject to jumps at unknown moments of time. This kind of situation is the main area of interest in the present work. It will be given special treatment in the next section. It is enough to mention here that the Kalman filter is applicable to this case also, although the parameter model will not match the true model at the instants of jumps. It should also be made clear that the Kalman filter, by itself, is ill equipped to handle such cases of abrupt changes. With a separate detection technique, the Kalman filter can be modified to handle the situation of abruptly changing parameters.

iv) Mixed Parameters

In some situations the behavior of the parameters may be mixed, i.e. they may change in any way at unknown instants of time, or they may remain constant for long periods. This behavior requires an algorithm capable tracking both small and large changes. In addition, the algorithm should give reasonably smooth estimates if the parameters happen to remain constant. The finite memory filter is inherently suitable for such a case although the performance may be less than satisfactory for piecewise constant parameters. In the next chapter an alternative to the finite memory filter is described. It also introduces

an algorithm which is claimed to be very efficient for tracking both small and large changes. Furthermore, the algorithm is capable of behaving as a constant memory filter if the parameters remain unchanged for a long segment of time.

I.3.2 Special Case: Piecewise Constant Parameters

In this work, special attention is given to piecewise constant parameters. A large class of problems belongs to this category. For example, a typical case of plant failure may involve one set of constant parameters before the failure, and another set of constant parameters after it. It is necessary to estimate the parameters on line, be prepared to detect any sudden changes in them (jumps), and then be able to estimate the new values as quickly as possible.

This dissertation develops an algorithm which is capable of detecting jumps very quickly. The jump detection is used to re-initialize a conventional Kalman filter with constant parameter model. Since the parameters are known to be piecewise constant, this approach yields much better estimates of the parameters compared to other techniques, such as the finite memory filter.

In the subsequent chapters it is shown that for the case of piecewise constant parameters it is very efficient to use the constant parameter filter with a one step detector. It is also shown that certain properties of the detector (test of hypothesis) can be improved by introducing a different kind of decision rule.

I.4 Brief Survey of Jump Detection Techniques

If one or more of the system parameters change abruptly, then the output, or the measured quantity can also be expected to show some abrupt change. If the measured signal is corrupted by noise, the problem is one of identification of changes in the mean. This classic problem has been widely studied in statistical and control literature. References [16] and [17] include detailed discussion of material concentrating on detection of abrupt changes. Also, in [3], [4], [5], [6], [7] and [8] this problem has been investigated, and various solutions have been suggested. The most widely used detection techniques are the following:

a. Likelihood ratio test,

b. Cumulative sum test,

c. Weighted sum square test,

d. Hinkly's detector,

e. Hine's detector.

These methods are briefly described below. First, the general theory behind detection techniques is summarized. The residual of the recursive filter, under the assumption of parametric invariance, is known to be {[12], [29]} a sequence of independent, zero-mean random variables. Any jump in the system parameters is expected to produce a shift in the mean of the resisual, which is to be detected by a hypothesis test designed on one of the detectors mentioned above. Actually, all of these detectors are based on the same principles, although they differ in the details of design and application. Each one has certain advantages and disadvantages, some are more suitable for off-line application, others

are better for on-line. Some of these detectors, which were originally designed for one type of systems, can be adapted to work for a different type of system. In particular, Hinkly's detector has been efficiently modified by Nikiforov [7] for use in more complex systems, including multidimensional ARMA models.

The likelihood ratio test,in its original form, must assume the jump magnitude to be known. In other words, one has to perform the likelihood ratio test between $H_0 : \mu_k=\mu_0$ for $k\leq n$ and $H_1 : \mu_k=\mu_1$ for $k\leq n$, *backward*, until H_1 is decided. A.N. Shirayev [32] and G. Lorden [33] have shown, independently of each other, that this procedure is optimal with respect to detection delays. It minimizes the detection delay for a fixed mean time between false alarms. The technique is, however, cumbersome when extended to the case of completely unknown jump magnitudes. There exist other detection techniques which are efficient for off-line application, but not for recursive filtering. It is enough to say here that extensive research has been done and strong theoretical framework has been established for the case when one is able to write explicit likelihoods (for example for Markovian models) and to expand them.

Michèle Bashville in [4], [5], [16] and [17] has explored a number of approaches for detection of abrupt changes in a scalar signal. She considered a "one model approach" in [17], which looks for a significant deviation from the model before change, and also a "two model approach" which is a simplification of the likelihood ratio test (GLR). She uses a CUSUM (cumulative sum) test in various autoregressive frameworks, and shows that the two model approach yields better performance because it

can take into account the direction of the jump. However, in the opinion of this author, if a Kalman filter is being used as the estimator, then it is not necessary to detect the direction of the jump.

This dissertation develops an algorithm in which the only job of the detector is to detect the occurrence of a jump as fast as possible, on the basis of a single sample (suitably normalized). Estimating the direction and magnitude of the jump is left to the filter after the jump has been identified. Such an algorithm is described in chapter II. It is called the "Combined Filter Algorithm (CFA)". In chapter III, the properties of this filter are presented together with the mathematical proofs.

Chapter IV is devoted to the problem of extending the proposed algorithm to multi-output situations. Chapter V deals with the problem of jumps in noise statistics. Their effect on the behavior of the algorithm is analized. In chapter VI the application of this algorithm to the power system state estimation (for relaying purposes) problem is discussed, and results of some computer simulation are presented. Appendices A and B contain simulation results comparing the combined filter algorithm with other alternatives. They include the ordinary Kalman filter, finite memory Kalman filter, and random walk model Kalman filter. They are tested for different kinds of parameter variations including slow drifts, small and large jumps.

Chapter II

FAST TRACKING OF TIME-VARYING PARAMETERS

Summary: In this chapter the problem of on-line estimation of time varying parameters is discussed. An algorithm which is capable of very fast detection of abrupt jumps is introduced. The algorithm also has the capability of tracking some slower changes. This procedure combines the aspects of finite memory filtering, hypothesis testing, and constant parameter model filtering. The design minimizes the upper bounds of missed jump and false alarm probabilities. Most of the decisions are made on a single sample basis, although there is an alternative to include a desired number of consecutive samples. The proposed algorithm is called the 'Combined Filter Algorithm (CFA)'. In sections II.1, II.2, II.3 and II.4 the necessary background material for the description of the algorithm is developed, and in section II.5 the CFA is described in detail.

II.1 The Random Walk Model Filter: An Alternative to Finite Memory Filtering

The simplest example of Gauss-Markov stochastic difference equation is the following:

$$X_{k+1} = X_k + W_k, \qquad (2.1)$$

with

$$E(W_k) = 0$$

and

$$E(W_k W_j^t) = \begin{cases} 0 & k \neq j \\ S_k, & k = j. \end{cases}$$

The signal model is, as before,

$$Z_k = H_k^t X_k + V_k,$$

with

$$E(V_k) = 0$$

and

$$E(V_k V_j^t) = \begin{cases} 0, & k \neq j \\ R_k, & k = j. \end{cases}$$

The Kalman filter designed for the above model has the property that its gains do not tend to zero as k tends to infinity. The classical use of this filter has been in the estimation of parameters which are subject to small fluctuations at each time step k, the expected amount of the fluctuations being given by the matrix S_k. From this point of view, the designer has to know, or be able to assume, the elements of S_k. However, another point of view is possible: one can interpret S_k as a design parameter, and the filter can be interpreted as an alternative to the finite memory filter. It is contended that by manipulating S_k this filter can be made to behave much like a finite memory filter. Small S_k corresponds to a large window length, and large S_k has the effect of having a short window length. Moreover, there is one definite advantage over the finite memory filter: it has more flexibility of design. For example, if one needs to have the estimates of a particular parameter to be smoother than others, one can just make the corresponding element of S_k smaller. In other words, there are more design

parameters instead of just one, namely the window length.

The basic similarity between the two algorithms is that both are capable of tracking time varying parameters, with the same tradeoff between the alertness of the filter and the smoothness of the estimates. Shorter window (larger S_k) results in a large forgetting factor, which means the filter is capable of quickly estimating a parameter change, if one should occur. On the other hand, if a change does not occur, then one has to accept noisy estimates of the same old parameters. Thus, the behavior of the random walk filter is similar to that of the finite memory filter, while being much simpler to implement, and having more design flexibility. The tracking capability of this filter can be improved by using tests of hypothesis, which is the subject of the next section.

II.2 Tests of Hypothesis as a Tool in Detecting Sudden Jumps

Hypothesis test has long been used as a statistical tool. In this section, its application in conjunction with the Kalman filter algorithm is discussed. The residual of the filtering process is given by

$$\tilde{Z}_k = Z_k - \hat{Z}_k \qquad (2.2)$$

$$= H_k^t X_k + V_k - H_k^t \hat{X}_{k/k-1}$$

$$= H_k^t e_{k/k-1} + V_k. \qquad (2.3)$$

Under the assumption of parametric invariance, the residuals constitute a sequence of zero-mean, identically distributed independent random variables (iid). With Gaussian noise assumption, they are also Gaussian

[12]. It can be shown by straightforward computation that the covariance matrix of the residual is given by

$$Q_{\tilde{z}_k} = [\tilde{Z}_k \tilde{Z}_k^t] = H_k^t Q_{k-1} H_k + R_k \qquad (2.4)$$

for the constant parameter model, and

$$Q_{\tilde{z}_k} = H_k^t [Q_{k-1} + S_{k-1}] H_k + R_k \qquad (2.5)$$

for the random walk model.

For the single output case, the residual can be normalized and the resulting variable will be Gaussian with zero-mean and unit variance. The normalized residual is given by

$$\tilde{Z}_{nk} = \frac{\tilde{Z}_k}{\sqrt{Q_{\tilde{z}_k}}} . \qquad (2.6)$$

The multi-output case will be considered in a later chapter. For a large class of problems the system can be modeled as single output, and the properties and decision rules of the filter are more conveniently discussed under the single output assumption. Therefore, $\tilde{Z}_{nk}$ will be considered a scalar in the following sections.

The magnitude of the normalized residual is used as the test statistic, and the following is the null hypothesis:

H_0=jump has not occured.

The alternate hypothesis is

H_1=jump has occured.

The decision rule must be designed to reject(accept) $H_0(H_1)$ on the basis of the behavior of $|\tilde{Z}_{nk}|$. Since $\tilde{Z}_{nk}$ is N(0,1) it is possible to use the

normal distribution tables for determining the thresholds for rejecting and accepting the hypotheses.

The basic element of the decision rule is:

i) reject H_0 if $|\tilde{Z}_{nk}|>T_r$,

where T_r stands for rejection threshold. This threshold is chosen by the designer in accordance with the maximum allowable probability of false alarm, α. This is a conditional probability defined as follows.

Definition 2.1: The probability of false alarm, denoted by α, is the conditional probability of the null hypothesis being rejected even though there is no jump. Thus,

$$\alpha = \Pr(H_0 \text{ is rejected/no jump has occured}). \qquad (2.7)$$

In statistical literature α is also called the probability of type I error. The quantity $1-\alpha$ is called the confidence level of the hypothesis test.

The other aspect of the decision rule concerns the acceptance of the null hypothesis. The classical approach has been binary, i.e. if H_0 can not be rejected, then it is automatically accepted. In most of the literature where hypothesis testing has been used to detect parameter jumps ([3], [6], [8], [30], [34], [48]), this binary decision rule (BDR) is the norm. Thus,

ii) accept H_0(reject H_1) if $|\tilde{Z}_{nk}|<T_r$.

However, this kind of binary decision rule has certain disadvantages. The performance of the algorithm can be much improved by replacing it by a "combined decision rule" (CDR). This CDR is the subject of the next section. As a preliminary step the following definitions are stated and discussed.

Definition 2.2: The probability of missed jump, denoted by β, is the conditional probability of the null hypothesis being accepted even though there is a jump. Thus,

$$\beta = \Pr(H_0 \text{ is accepted/jump has occured}). \qquad (2.8)$$

It is also known as the probability of type II error.

Numerical Computation of β

Numerical computation this probability is possible only as a function of the amount of shift in the mean of the test statistic. In the present context, this shift (from zero) is given by

$$M = \frac{H_k^t \, \Delta X}{\sqrt{Q_{z_k}}},$$

where

ΔX - amount of jump in the parameter,

Q_{z_k} - variance of the normalized residual.

Only positive M is considered here because of the symmetry of the standard normal density curve. As a result of the jump, $\tilde{Z}_{nk} \sim N(M, 1)$. The probability of a type II error (β) is the following:

$$\beta = \Pr(-T_a < \tilde{Z}_{nk} < T_a).$$

Subtracting M throughout in order to standardize $\tilde{Z}_{nk}$, one gets

$$\beta = \Pr(-T_a - M < \eta < T_a - M) \qquad (2.8a)$$

where

$$\eta \sim N(0, 1).$$

Definition 2.3: The power of the test, denoted by P^t, subject to the constraint

$$\alpha \leq \alpha_f \text{ (fixed level)}, \tag{2.9}$$

is the conditional probability of the alternate hypothesis being accepted when there is a jump. Thus,

$$P^t = \Pr(H_1 \text{ is accepted/jump has occured}) = 1 - \beta. \tag{2.10}$$

The objective here is to maximize the power of the test (2.10) for a fixed level. However, in the on-line framework the criterion is to maximize the power of the test *and* to minimize the detection delay.

For example, suppose the problem is to decide which of the following hypotheses to accept:

H_0 = there is no jump, i.e. the system is

$$Z_k(0) = H_k X_k(0)$$

H_1 = there has been a jump, and the system is

$$Z_k(1) = H_k X_k(1)$$

where both the mean values of $Z_k(0)$ and $Z_k(1)$ are known. This is simple hypothesis testing, and the problem is to decide whether a switch between the two regimes has taken place. This problem of switching statistics has been widely investigated in mathematical literature, and good detection techniques exist ([37], [38], [39]). However, the problem becomes more complicated when the mean value of the jump is unknown [$X_k(1)$ can come from an infinite set of values]. Then it is impossible to compute the power of the test.

When the value of the parameter is unknown after the jump (as in the present context of Kalman filtering), it constitutes a "multiple hypotheses testing". In that case H_1 does not have a known distribution,

but comes from a set of unknown distributions. Then (2.10) is an incorrect statement, and the statistician's goal is to design *uniformly most powerful* (UMP) tests, i.e. tests which have the highest power for each distribution of the alternate hypotheses H_1. This problem is strongly related to the Neyman-Pearson lemma [35]. In [36] Albert Benveniste mentions the fact that change detection problems are multiple hypotheses testing problems for which the Neyman-Pearson lemma is not valid, so UMP tests do not exist. The problem would have been simpler to handle if the mean parameter values after the change takes place were known.

In this work the hypothesis test is designed to operate on a single sample. Since the test may fail, it becomes necessary to define a term "detection delay" which carries a slightly different meaning from the classical one.

Definition 2.4: Detection delay (D_d) is the number of time steps elapsed between the occurance of the jump and its detection. It is shown in chapter III that the algorithm actually operates in such a way that

$$Pr(D_d=1) > Pr(D_d=j) \text{ for all } j > 1.$$

II.3 Random Walk Model Filter with Hypothesis Testing

It has been mentioned earlier that the random walk filter performs similar to the finite memory filter in tracking time varying parameters. However, just like the finite memory filter, the larger the jump, the slower the tracking. To increase the speed of estimation in the case

of a large jump, tests of hypothesis can be incorporated into the algorithm. If a jump is detected, then the filter is re-initialized with a large covariance matrix, and the last (pre-fault) estimates used as the initial estimates.

With the proposed approach it is enough for the jump detector to detect the presence (absence) of a jump. Any detector which gives an estimate of the jump magnitude and direction can be regarded as doing redundant work. It may be argued that if the detector is able to provide a reasonable initial estimate of the parameter, then it will speed up the convergence of the Kalman filter. The counterargument is that no detector will give a good estimate of the jump magnitude on the basis of one sample. Hence it is necessary to wait for a few samples before starting the new Kalman filter. On the other hand, if the Kalman filter is started as soon as the jump is detected, it is known to converge within a few steps if the model and the measurement noise statistics are correct.

Therefore, it is the contention of this dissertation that it is better to rely on the very simple one step detector as proposed here, and leave the estimating to the Kalman filter. Computationally this is the simplest possible scheme; it is also one which requires the minimum apriori knowledge. Of course, it will miss some small jumps because the rejection threshold has to be kept high in order to avoid a large α value. But missing some jumps is not as dangerous in this case as in the case of using a constant parameter filter. The basic concept behind this algorithm is that the one-step detector will detect large jumps and random walk model will track small jumps. The filter will be re-initia-

lized only in the event of large jumps.

To summarize, the random walk filter is started with the following parameters chosen by the designer:

1. S_k – large values ensure more alertness, but also noisier estimates,
2. α – depending on how much false alarm probability can be accepted,
3. T_r – determined from the standard Gaussian distribution table, for the chosen α,
4. re-initialization parameters for the case when a jump is detected, and
5. initialization parameters for the Kalman filter.

The test statistic is the normalized residual, $\tilde{Z}_{nk}$, given by eq.(2.6). The covariance of the residual, $Q_{\sim}$,is given by eq.(2.4) or eq.(2.5).

The decision rule is:

reject (accept)H_0 if $|\tilde{Z}_{nk}| \gtrless T_r$.

If a jump is detected, the Kalman filter is re-initialized. If no jump is detected, no action is taken.

In section II.6 a new kind of algorithm is described, which is called the Combined filter algorithm (CFA). It uses the so-called combined decision rule (CDR), which is developed in section II.5. As a preliminary to the CDR, the disadvantage of the binary decision rule (BDR) is discussed in section II.4.

II.4. The Disadvantage of the Binary Decision Rule

The binary decision rule states that when H_0 can not be rejected, it must be accepted. In statistical literature it is usually mentioned that the alternative of rejecting a hypothesis is not necessarily to accept it. However, in most applications a binary decision rule is adopted. It is not usually taken into consideration that borderline cases may arise.

Suppose, for example, that $|\tilde{Z}_{nk}| = T_r - \epsilon$, with $\epsilon>0$. Then according to the binary decision rule (BDR) hypothesis H_0 is accepted. However, this action carries a large probability of missing a jump. This probability gets larger as ϵ gets smaller. In accepting H_0 the probability of being wrong may be greater than the probability of being right. There is no way to minimize both α and β at the same time, using a BDR. However, the upper bound of β can be reduced without increasing that of α by introducing the combined decision rule (CDR).

II.5 The Combined Decision Rule

The combined decision rule is capable of saying *yes*, *no* or *maybe* as opposed to the usual BDR, which says only *yes* or *no*. This flexibility is achieved by accepting the null hypothesis if $|\tilde{Z}_{nk}|$ is below a certain chosen acceptance threshold, T_a . It will be shown in chapter III that by doing this the upper bound of the probability of missed jump can be reduced. The choice of the acceptance threshold depends, in general, on the *minimum acceptable* power of the test which is decided upon by the designer. The rejection threshold T_r is chosen in the same way as in the

case of the BDR. Obviously $T_a < T_r$, the limiting case being $T_a = T_r$ when the CDR degenerates into the BDR.

The combined decision rule is summarized in the following:

1) reject H_0 if $|\tilde{Z}_{nk}| > T_r$,

2) accept H_0 if $|\tilde{Z}_{nk}| < T_a$,

3) neither accept nor reject H_0 if $T_a \leq |\tilde{Z}_{nk}| \leq T_r$.

Similar to the case of the BDR, when H_0 is rejected, the filter is re-initialized. No action is taken when H_0 is accepted. If H_0 can neither be accepted nor be rejected, then basically two options are available: *i)* wait for more data, and *ii)* activate a RWM filter and use it until H_0 is either rejected or accepted.

In the present work option *ii)* is adopted. It is contended that option *i)* may result in missing some jumps for a long time. If the jump is such that the test statistic falls in the region in between T_a and T_r for a number of consecutive steps, then the RWM filter is clearly the better option.

II.5.1 Choice of the Rejection Threshold

The rejection threshold is chosen in a straightforward manner. It is directly obtainable from the standard normal distribution table, depending only on the choice of the probability of false alarm, α. For example, if one chooses $\alpha = 0.01$, then the corresponding rejection threshold is $T_r = 2.58$. Another option available to the designer is to require that the null hypothesis be rejected n consecutive times before

re-initialization of the filter is done. Then, for example, if n = 2, one obtains $\alpha = 10^{-4}$ for the same previous threshold 2.58. The choice of a large n slows the response of the filter while reducing the false alarm probability. Therefore, in the interest of speed, n should not be too large.

II.5.2 Choice of the Acceptance Threshold

The acceptance threshold, T_a, is more difficult to choose than T_r. Basically it depends on the power of the test. However, this being a case of multiple hypothesis testing, the best one can do is to select a 'minimum allowable power of the test' ($P^t_{min} = 1 - \beta_{max}$) and compute T_a accordingly. Since β is a function of T_a and M, one has to choose a certain fixed M in order to obtain a fixed T_a for a given β_{max}. The problem is more compounded by the fact that a closed form expression does not exist for T_a in terms of β and M. Therefore an iterative method of computation has to be adopted.

It should also be mentioned here that although smaller T_a gives higher power of the test (which is desirable), one can not reduce T_a indefinitely. In order to retain a reasonable amount of probability of making a correct decision in the case of no jump, some $T_{a(min)}$ should be specified. Since

$$\Pr(H_0/\bar{J}) = \int_{-T_a}^{T_a} f(\eta)d\eta \ ,$$

a pre-specified $T_{a(min)}$ such that $T_a \geq T_{a(min)}$ will result in

$$\Pr(H_0/\bar{J}) \geq \Pr(H_0/\bar{J})_{min}.$$

Thus the choice of T_a can be viewed as an iterative process with a constraint. Any reasonable iteration method can be used, and the choice of the error criterion ϵ, such that $|\beta - \beta_{max}| \leq \epsilon$ should depend on the intended application.

According to equation (2.8a), the probability β can be computed as

$$\beta = \int_c^d f(\eta)d\eta,$$

where $f(\eta)$ is the Gaussian density function, and

$$c = -T_a - M,$$

$$d = T_a - M.$$

M is the amount of bias created in $\tilde{Z}_{nk}$ as the result of a parameter jump,

$$\text{i.e. } M = \frac{H_k^t \Delta X}{\sqrt{Q_{z_k}}}$$

where ΔX is the amount of jump.

The choice of T_a can be summarized in the following steps.

1. Select $T_{a(min)}$ according to the desired $\Pr(H_0/\bar{J})_{min}$ from the standard normal distribution table.
2. Choose β_{max} based on the minimum allowable power (P^t_{min}).
3. Choose M_{min}, which is interpreted as the minimum bias. For all $M \geq M_{min}$, one obtains $P^t \geq P^t_{min}$ and for all $M \leq M_{min}$, one obtains $P^t \leq P^t_{min}$ (for a fixed T_a). In this sense M_{min} is the cut-off point in the bias corresponding to the chosen P^t_{min}.

4. Select an initial value for $T_a = \hat{T}_a$. A good initial value is the pre-chosen $T_{a(min)}$.

5. Compute $\hat{\beta} = \int_c^d f(\eta)d\eta$,

where

$$c = -\hat{T}_a - M_{min},$$

$$d = \hat{T}_a - M_{min}.$$

6. Compare $\hat{\beta}$ to the chosen β_{max}, and adjust $\hat{T}_a$ accordingly. Since β is an increasing function of T_a for each fixed M, if $\beta_{max} - \hat{\beta} > 0$ then the value of T_a should be incremented by a suitably small amount. However, if $\beta_{max} - \hat{\beta} < 0$, and the initial choice of T_a was $T_{a(min)}$, then T_a can not be decremented because of the constraint $T_a \geq T_{a(min)}$. In that case one has to adjust the value of M_{min}, and start anew. If the initial choice of T_a was larger than $T_{a(min)}$, then T_a may be decremented.

7. When $|\hat{\beta} - \beta_{max}|$ is sufficiently small, the iteration stops.

It is concievable that in some applications the power of the test is not very important. However, it may be essential that $Pr(H_0/\bar{J})$ be high. In such a situation T_a may be selected directly from the required value of $Pr(H_0/\bar{J})$. Computationally this would be a much simpler situation.

II.6 The Combined Filter Algorithm

In this section an algorithm which is a combination of the constant parameter model filter, random walk model filter and hypothesis testing

is described. The random walk model filter used here is the one where S_k is viewed as a design parameter. The detection is one-step for presence or absence of jump, but there is also a region of non-detection. This is the result of using the CDR which was discussed in the previous section.

This algorithm is claimed to be capable of behaving as self-learning in the following sense.

i) If the parameter changes are such that the residuals are not shifted enough (from zero) to reject H_0 with the chosen level of confidence, then the RWM filter, capable of tracking small changes, is activated. This means that if there are small fluctuations, then the filter operates as a RWM filter.

ii) If the parameters remain constant for a long time, then it acts as a constant parameter filter.

iii) if the residual behaves in such a way that a jump is detected, then the filter is re-initialized and proceeds to estimate the new values of the parameters immediately.

Of course, as can be expected with any statistical method, there is some element of uncertainty which must be accepted. These are the pre-determined amounts of false alarm and missed jump probabilities.

The combined filter algorithm starts with the constant parameter model,

$$X_{k+1} = X_k.$$

The signal model is the following:

$$Z_k = H_k^t X_k + V_k,$$

where H_k is a known matrix of appropriate dimensions (as in the

classical case of Kalman filtering), or is given by the following equation:

$$H_k^t = [-Z_{k-1} \; -Z_{k-2} \; \bullet\bullet\bullet \; -Z_{k-\nu} \; U_{k-1} \bullet\bullet\bullet \; U_{k-\mu}].$$

The parameter vector is then given by

$$X^t = [\, A_1 \; \bullet\bullet\bullet \; A_\nu \; B_1 \; \bullet\bullet\bullet \; B_\mu \,],$$

where A_i, B_i are the coefficients of the ARMA model.

As mentioned earlier (in chapter I), the same algorithm is valid for both kinds of models, except that in the case of an ARMAX model the Kalman gains and covariances can not computed off-line because the H_k matrix depends on past outputs. Thus, H_k is unknown off-line, although known at every step on-line. For the case of the ARMAX model the whole algorithm must be implemented on-line.

In the following paragraphs, the CFA is described step by step.

1) Choose Q_0, $\hat{X}_0$, S_k; R_k is either known or assumed.
   Also choose α and P^t_{min}.
2) From the values of α and P^t_{min} determine T_r and T_a according to the method described in section II.3.
3) The estimation equation is:

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k \tilde{Z}_k,$$

   where $\tilde{Z}_k$ is the residual, defined as:

$$\tilde{Z}_k = Z_k - H_k^t \hat{X}_{k/k-1}.$$

   For both of the following models

$$X_k = X_{k-1} \quad \text{(CPM)}$$

and $X_k = X_{k-1} + W_{k-1}$ (RWM),

the prediction equation is given by

$$\hat{X}_{k/k-1} = \hat{X}_{k-1/k-1}.$$

The Kalman gain, K_k, is given by

$$K_k = Q_{k/k-1} H_k [\, R_k + H_k^t \, Q_{k/k-1} H_k]^{-1}.$$

The covariance of the estimates, Q_k, is updated for both the models by the recursive equation

$$Q_{k/k} = [\, I - K_k^t H_k] \, Q_{k/k-1}.$$

For the RWM, the projected Q_k is given by

$$Q_{k+1/k} = Q_{k/k} + S_k,$$

where

$$S_k = E[W_k W_k^t].$$

For the CPM,

$$Q_{k+1/k} = Q_{k/k}.$$

4) The residual is normalized by dividing it by its own standard deviation at each step. The variance of the residual is given by

$$Q_{\tilde{z}_k} = H_k \, Q_{k/k-1} H_k + R_{2k}$$

for both the models.

The normalized residual

$$\tilde{Z}_{nk} = \frac{\tilde{Z}_k}{\sqrt{Q_{\tilde{Z}_k}}}$$

is used as the test statistic.

5) The hypotheses are:

H_0 = the null hypothesis: no jump has occured,

H_1 = the alternate hypothesis: a jump occured.

6) The test is:

Check whether or not $\tilde{Z}_{nk}$ exceeds the chosen rejection threshold, T_r, and also whether or not $\tilde{Z}_{nk}$ falls inside the acceptance threshold, T_a.

7) The decision rule is:

Take no action if $|\tilde{Z}_{nk}| < T_a$, i.e. H_0 is accepted.

Activate RWF if $T_a < |\tilde{Z}_{nk}| < T_r$, i.e. neither H_0 nor H_1 is accepted.

Re-initialize constant parameter filter if $|\tilde{Z}_{nk}| > T_r$, i.e. H_0 is rejected (H_1 is accepted).

In the next chapter the properties of this filter are discussed. A number of computer simulations have been performed to test this algorithm. The results and plots are presented in appendix B.

Chapter III

PROPERTIES OF THE COMBINED FILTER

Summary: In this chapter the basic properties of the Combined Filter Algorithm are listed and discussed, with mathematical proofs where necessary. The general development is for the single output case, although most of the properties can be generalized or extended to the multiple output situation. Those properties are discussed in the next chapter. Some basic assumptions are required in order for the properties to hold. These assumptions are given in the next section, along with a list of notations.

III.1 Assumptions, Notations and Definitions

The following is a list of the required assumptions.

1) Any changes or jumps in the parameters are reflected in the output of the system, and consequently in the residual of the filter. Larger changes have larger impact on the residual.

2) Jumps are sufficiently far removed from one another so that each jump can be analyzed separately.

3) The system is stable for any change in the parameters.

The following is a list of notations and definitions.

T_r- rejection threshold.

T_a- acceptence threshold.

H_0 is the null hypothesis.

J is the event that there was a jump.

D is the event $|\tilde{Z}_{nk}| > T_r$.

F is the event $|\tilde{Z}_{nk}| < T_a$.

G is the event $T_a \leq |\tilde{Z}_{nk}| \leq T_r$.

$M = |E(\tilde{Z}_{nk})|$, amount of bias (in the normalized residual).

$\alpha = Pr(D/\bar{J})$, prob. of false alarm.

$\beta = Pr(F/J)$, prob. of missed jump.

$\omega = Pr(J/D)$, prob. of rightfully rejecting H_0.

$\gamma = Pr(D/J)$, prob. of detection.

$\tau = Pr(G/\bar{J})$, prob. of invoking random walk filter given no jump has occured.

$\psi = Pr(G/J)$, prob. of invoking random walk filter if a jump has occured.

The above probabilities are conditional, but it is the convention to use terms like 'probability of false alarm' instead of 'conditional probability of false alarm'.

III.2 Preliminary Considerations

When a jump occurs, it may or may not be detected by the hypothesis test. On the other hand, the test may fail even when there is no jump. The RWM filter may be started when there was no need to do so. As a result the algorithm may or may not work optimally. Therefore, it is necessary to divide the properties into two broad groups: first, when the algorithm makes the correct decision and works optimally, and second, when it makes a wrong decision. The second group of properties

essentially considers the robustness of the algorithm against outliers and missed jumps. The following block diagram summarizes the different situations that may arise while the combination filter is running. It should be noted that the algorithm starts with the CPM.

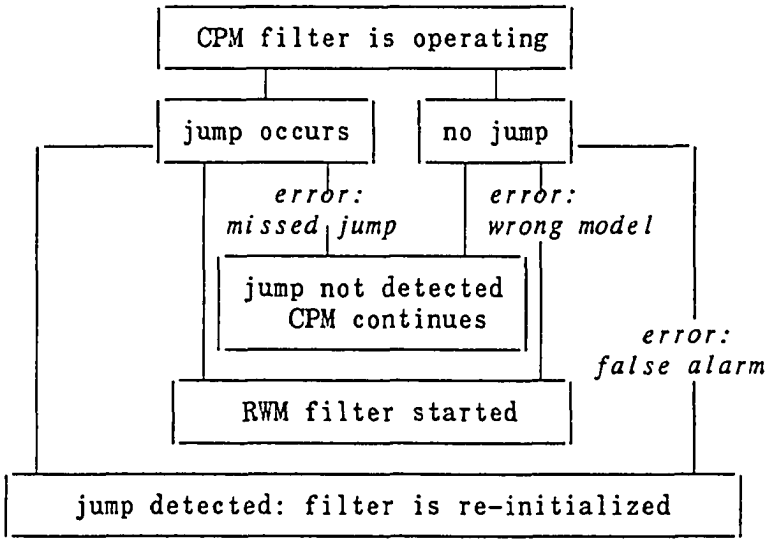

Fig. 3.1

Section III.3 contains the properties of the filter when a correct decision is made. It also discusses the probabilities of making correct decisions or various errors assuming there was no error upto the time step being considered.

The behavior of the algorithm when any one of the three kinds of error has been made is described in section III.4.

III.3 General Properties of the Filter

Lemma 1: If a jump occurs at step k and is detected at the following step or at a later time, then the reinitialized filter will give increasingly accurate estimates of the new parameter values until the next jump, or a false alarm, whichever occurs first.

Proof: The re-initialized filter is just an ordinary Kalman filter with constant parameter assumption. Therefore, all its convergence properties {[3],[4]} hold for each piecewise constant segment of the present algorithm.

Lemma 2: The probability of false alarm (α) can be made as small as desired at the expense of one of the following:

i) increase of estimation errors in some situations,

ii) decrease in the detection speed.

Remark 1: When using the binary decision rule, α can be reduced either at the cost of increasing β, or decreasing the speed of detection.

Proof: False alarm occurs when the null hypothesis is mistakenly rejected. The probability of this event is

$$\alpha = \Pr(D/\bar{J}) = 1 - \int_{-T_r}^{T_r} f(\eta)d\eta \qquad (3.1)$$

where $f(\eta)$ is the Gaussian density function with zero mean and unit variance, and T_r is the rejection threshold. Evidently, α can be made smaller by increasing T_r. However, this results in an outward expansion

of the region of RWF. As a result, the probability of invoking the RWF (even though there is no jump) increases.

Indeed,

$$\tau = \Pr(G/\bar{J}) = 2 \int_{T_a}^{T_r} f(\eta)d\eta \ .$$

Clearly if T_r is larger, then τ is larger. Thus the probability of noisier estimates is increased.

Another way to decrease α is to require k consecutive rejections before sounding an alarm, i.e., before re-initializing the filter. Since it is known that if no jump has taken place, then the normalized innovations constitute a sequence of mutually independent standard normal random variables, clearly it follows that the conditional probability of wrongful rejection remains the same (α) at each step. Therefore, the probability of k consecutive (wrongful) rejections is given by

$$\alpha(k) = \Pr(k \text{ cons.rej.}) = \alpha^k < \alpha.$$

However, in this case the detection is delayed by at least k steps.

Lemma 3: The upper bound of the probability of a missed ump can be reduced by reducing the acceptance threshold.

Proof: The probability of missed jump is given by:

$$\beta = \int_c^d f(\eta)d\eta \qquad (3.2)$$

where

$$c = -T_a - M,$$

$$d = T_a - M,$$

and

M = amount of bias (in the test statistic), created by the jump.

Let P^t_{min} be the minimum power of the test chosen by the designer. Then the corresponding probability of missed jump is given by

$$1 - P^t_{min} = \beta_{max} = \int_{c1}^{d1} f(\eta)d\eta \ ,$$

with

$$c1 = - T_a - M_{min}$$

and

$$d1 = T_a - M_{min},$$

where M_{min} is the cut-off point in the bias such that for all $M \geq M_{min}$ the power of the test $P^t \geq P^t_{min}$.

By the nature of the normal probability density function it is clear that for each fixed M_{min} one can reduce β_{max} by reducing T_a.

Remark 2: This method of reducing β_{max} results in increased noisiness of estimates because the region of the RWM filter is expanded inward. Also, whenever the RWM filter is in operation with the true parameters being constant, the filter becomes suboptimal. The behavior of the filter in this situation is considered in the next section.

Lemma 4: The probability of rightful rejection ω increases with the increase in the bias M.

Proof: Since

$$\Pr(\bar{J}/D) = \frac{\Pr(D/\bar{J})\Pr(J)}{\Pr(D)},$$

and

$$\Pr(D) = \Pr(D/J)\Pr(J) + \Pr(D/\bar{J})\Pr(\bar{J}),$$

it is possible to show that

$$\omega = \Pr(J/D) = \frac{\Pr(D/J)}{\Pr(D/\bar{J})\rho + \Pr(D/J)}, \qquad (3.3)$$

where

$$\rho = \frac{\Pr(\bar{J})}{\Pr(J)}$$

is a measure of the relative likelihood of jump. Note that $\Pr(D/\bar{J}) = \alpha$ is the probability of false alarm, and the probability of detection is

$$\gamma = \Pr(D/J) = 1 - \int_{c}^{d} f(\eta)d\eta, \qquad (3.4)$$

where the limits c and d are defined as

$$c = -T_r - M,$$

$$d = T_r - M.$$

Now, let

$$\lambda = \frac{\alpha\rho}{\gamma}; \qquad (3.4a)$$

then, from eq.(3.3) and (3.4a),

$$\omega = \frac{1}{1+\lambda} \quad . \qquad (3.5)$$

Eq.(3.4) shows that, for each fixed T_r, γ increases with an increase in M. According to eq.(3.4a), then, λ will decrease provided the probabilities ρ and α remain constant. Consequently, following eq.(3.5), the probability of rightful rejection ω will increase with the increase in M.

Lemma 5: Assuming that M is uniformly distributed between 0 and M_u, the (unconditional) probability of detection Pr(D) varies between α and 1. In particular,

a) $\lim_{M_u \to 0} \Pr(D) = \alpha$,

b) $\lim_{M_u \to \infty} \Pr(D) = 1$.

Proof: By definition,

$$\Pr(D) = 1 - \frac{1}{M_u} \frac{1}{\sqrt{2\pi}} \int_0^{M_u} \int_c^d e^{-\frac{\eta^2}{2}} d\eta dM.$$

Denote

$$A(M) = \frac{1}{\sqrt{2\pi}} \int_c^d e^{-\frac{\eta^2}{2}} d\eta.$$

Since

$$c = -T_r - M,$$

and

$$d = T_r - M,$$

from the nature of the standard normal probability density function,

it is evident that A(M) is a decreasing function of M. Then

$$\Pr(D) = 1 - \frac{1}{M_u}\int_0^{M_u} A(M)dM. \qquad (3.6)$$

Since M is distributed between 0 and M_u, it is seen that

$$-T_r - M_u < -T_r - M = c < -T_r$$

and

$$T_r - M_u < T_r - M = d < T_r,$$

which means that $M_u \to 0$ implies $c \to -T_r$ and $d \to T_r$. Thus, by definition of A(M),

$$\lim_{M_u \to 0} A(M) = 1 - \alpha.$$

Consequently

$$\lim_{M_u \to 0} \Pr(D) = \lim_{M_u \to 0} [1 - \frac{1}{M_u}\int_0^{M_u}(1 - \alpha)dM]$$

$$= 1 - (1 - \alpha)$$

$$= \alpha.$$

Part a) is thus proved. For part b), first it is shown that for large M, the function A(M) is bounded by an exponentially decreasing function. Denote

$$C(M) = \frac{1}{\sqrt{2\pi}}\int_c^d e^{\frac{x}{2}} dx.$$

From the nature of the standard normal probability density function, there exists $K > 0$ such that $A(M) < C(M)$ for all $M > K$. The function $C(M)$ can be written, after carrying out the integration and rearranging, as

$$C(M) = K_c \, e^{-\frac{M}{2}},$$

where

$$K_c = \frac{2}{\sqrt{2\pi}} \left(e^{\frac{T_r}{2}} - e^{\frac{-T_r}{2}} \right) = \text{constant for fixed } T_r.$$

It was shown in a) that $A(0) = 1 - \alpha$, and that $A(M)$ is a decreasing function of M. Therefore,

$$\int_0^L A(M)dM < L(1 - \alpha), \quad \text{for all } L > 0.$$

It follows that

$$\int_0^\infty A(M)dM = \int_0^K A(M) + \int_K^\infty A(M)$$

$$< K(1 - \alpha) + \int_K^\infty C(M)dM,$$

where K is such that $C(M) > A(M)$ for all $M > K$. The integral

$$\int_K^\infty C(M)dM = K_c \int_K^\infty e^{-\frac{M}{2}} dM$$

$$= -2K_c\left(e^{-\infty} - e^{-\frac{K}{2}}\right)$$

$$= 2K_c\, e^{-\frac{K}{2}} = K_a$$

is constant. Thus,

$$\int_0^{\infty} A(M)dM < K(1-\alpha) + K_a = K_b \qquad (3.6a)$$

From eq.(3.6) and inequality (3.6a), it follows that

$$\lim_{M_u \to \infty} \Pr(D) = 1 - 0 = 1.$$

This completes the proof of lemma 5.

Lemma 6: If there is a jump of amount ΔX_k at step $k = k_0$, then the mean-square value of the residual at step k is given by

$$E(\tilde{Z}_k \tilde{Z}_k^t) = \underset{z_k}{Q} \text{ (model)} + H_k^t \Delta X_k \Delta X_k^t H_k.$$

Proof: Assume there is a jump at step $k = k_0$, and that

$$X_k = X_{k-1} + \Delta X_k.$$

The residual is given by

$$\tilde{Z}_k = H_k^t\, e^a_{k/k-1} + V_k\ ,$$

where $e^a_{k/k-1}$ is the actual error, defined as

$$e^a_{k/k-1} = X_k - \hat{X}_{k/k-1}.$$

The actual error and the error computed in the model are related by

$$e^{a}_{k/k-1} = e^{m}_{k/k-1} + \Delta X_k.$$

Therefore,

$$E(e_{k/k-1} e^{t}_{k/k-1})(\text{actual}) = Q^{m}_{k/k-1} + \Delta X_k \Delta X_k^{t}.$$

Hence

$$E(\tilde{Z}_k \tilde{Z}_k) = H_k^{t} Q^{m}_{k/k-1} H_k + R_k + H_k^{t} \Delta X_k \Delta X_k^{t} H_k$$

$$= Q_{\underset{z_k}{\sim}} (\text{model}) + H_k^{t} \Delta X_k \Delta X_k^{t} H_k.$$

Theorem 3.1: If there is a parameter jump accompanied by an increase (decrease) in noise variance, with

$$R_k \text{ (actual)} = R_k \text{ (model)} + \Delta R_k,$$

and

$$X_k \text{ (actual)} = X_{k-1} + \Delta X_k,$$

then

i) the normalized residual has non-zero mean,

ii) its variance is greater (less) than unity.

Furthermore, if $\Delta R_k > 0$, then

iii) the mean square value of the normalized residual is greater than unity.

Proof: (For single output case) Part i) follows from the fact that a jump produces a bias in the residual. Observe that the mean value is given by

$$E(\tilde{Z}_{nk}) = \frac{H_k^t \Delta X_k}{\sqrt{Q_{z_k}}} .$$

To show part ii), observe that from lemma 6, the actual mean square value is

$$E(\tilde{Z}_k \tilde{Z}_k^t) = H_k^t Q^m_{k/k-1} H_k + R_k(\text{actual}) + H_k^t \Delta X_k \Delta X_k^t H_k. \quad (3.7)$$

Also,

$$E(\tilde{Z}_{nk} \tilde{Z}_{nk}^t) = \frac{E(\tilde{Z}_k \tilde{Z}_k^t)}{H_k^t Q^m_{k/k-1} H_k + R_k(\text{model})} . \quad (3.8)$$

The variance of the normalized residual is given by

$$\text{var}(\tilde{Z}_{nk}) = E(\tilde{Z}_{nk} \tilde{Z}_{nk}^t) - E(\tilde{Z}_{nk}) E(\tilde{Z}_{nk})^t.$$

Since

$$E(\tilde{Z}_{nk}) = \frac{H_k^t \Delta X_k}{\sqrt{H_k^t Q^m_{k/(k-1)} H_k + R_k(\text{model})}} ,$$

it is clear that

$$\text{Var}(\tilde{Z}_{nk}) = \frac{H_k^t Q^m_{k/(k-1)} H_k + R_k(\text{actual})}{H_k^t Q^m_{k/(k-1)} H_k + R_k(\text{model})} ,$$

$$= 1 + \frac{\Delta R_k}{Q_{z_k}} .$$

Thus

$$\mathrm{Var}(\tilde{Z}_{nk}) \gtrless 1 \text{ for } \Delta R_k \gtrless 0,$$

which proves part ii).

Next, observe that from eq. (3.7) and (3.8) it follows that

$$E(\tilde{Z}_{nk}\tilde{Z}_{nk}^t) = 1 + \frac{H_k^t \Delta X_k \Delta X_k^t H_k + \Delta R_k}{Q_{z_k}} \qquad (3.8a)$$

If $\Delta R_k > 0$, then eq. (3.8a) implies that

$$E(\tilde{Z}_{nk}\tilde{Z}_{nk}^t) > 1,$$

which proves part iii). Thus, the theorem is proved.

Theorem 3.2: The probability that the detection delay is equal to one is maximal, i.e.,

$$\Pr(D_d = 1) > \Pr(D_d = j) \text{ for all } j > 1$$

Proof: It is evident that

$$\Pr(D_d = j) = \beta_{k_0}\beta_{k_0+1}\cdots\beta_{k_0+j-1}\gamma_{k_0+j},$$

where k_0 is the step when the jump occured, and k_0+j is the step when it is detected. Since $\beta < 1$, clearly

$$\Pr(D_d = j) < \Pr(D_d = i) \text{ for all } i < j.$$

It follows that $\{\Pr(D_d = j)\}$ is a decreasing sequence in j, with the maximum occuring at j = 1.

III.4 Robustness Considerations: Behavior of the Algorithm if an Incorrect Decision Has Been Made

In this section the consequences of making incorrect decisions are analyzed. The goal is to show that even when the errors are never corrected, there is no danger of filter divergence.

In the case of an undetected jump, the estimates will be biased at the worst. However, the bias in the residual, being a function of the time-varying observation matrix, changes at each time step. The probability of detection varies accordingly, so that a missed jump may or may not be detected later. In order for the filter to diverge, however, the bias will have to keep increasing consistently. This will automatically lead to detection (since the conditional probability of detection is an increasing function of the bias). Hence the algorithm is robust against divergence.

If there has been no jump, but the RWM filter is activated, then a model mismatch is created. This situation does not pose any great danger as will be shown later in this section. In the case of a false alarm, the needless re-initialization creates noisy estimates for a short time. This condition corrects itself, provided there is not another false alarm too close in the future. In chapter IV it is shown that even under the very unlikely circumstance of a chain of false alarms, the filter can not diverge. The following properties are stated and proved for the single output case.

Theorem 3.3: If there is an undetected jump at step k_0, and the CPM filter is operating, then the mean estimation error vector is given by

$$E(e_k) = \Psi_k \bullet\bullet\bullet \Psi_{k_0} \Delta X,$$

where

$$\Psi_k = I - K_k H_k^t$$

and

$$k \geq k_0.$$

Proof: Let ΔX_k be the jump at step $k = k_0$, such that

$$X_k = X_{k-1} + \Delta X_k.$$

Define

$$e_k = X_k - \hat{X}_{k/k}. \qquad (3.9)$$

Then,

$$e_k = X_{k-1} - (\hat{X}_{k-1} + K_k \tilde{Z}_k) + \Delta X_k$$

$$= e_{k-1} - K_k[H_k^t(X_{k-1} + \Delta X_k - \hat{X}_{k-1}) + V_k] + \Delta X_k.$$

$$= (I - K_k H_k^t)(e_{k-1} + \Delta X_k) + K_k V_k. \qquad (3.9a)$$

The mean estimation error is given by

$$E(e_k) = (I - K_k H_k^t)[E(e_{k-1}) + \Delta X_k]. \qquad (3.10)$$

Assume that there was no jump before k_0 and there is no jump after k_0. Then,

$$\Delta X_k = 0 \quad \text{for all } k > k_0,$$

$$\Delta X_k = \Delta X \quad \text{for } k = k_0.$$

Consequently, for all $k > k_0$,

$$E(e_k) = (I - K_k H_k^t) E(e_{k-1}).$$

Denoting

$$(I - K_k H_k^t) = \Psi_k, \tag{3.11}$$

it follows that

$$E(e_k) = \begin{cases} \Psi_k E(e_{k-1}) & \text{for } k > k_0 \\ \Psi_k \Delta X & \text{for } k = k_0. \end{cases} \tag{3.12}$$

Therefore, for $k \geq k_0$ it can be written that:

$$E(e_k) = \Psi_k \Psi_{k-1} \cdots \Psi_{k_0} \Delta X. \tag{3.13}$$

Theorem 3.4: Assuming that the filter operated optimally up to step $k_0 - 1$, the covariance of the estimation error does not depend on the occurrence of a parameter jump at step k_0.

Proof: The covariance Q_k is defined as

$$Q_k = E\{[e_k - E(e_k)][e_k - E(e_k)]^t\} = E(\Delta e_k \Delta e_k^t).$$

If there is a jump of magnitude ΔX at step $k = k_0$, then by eq. (3.9a) and (3.12), for all $k \geq k_0$,

$$\Delta e_k = e_k - E(e_k) = \Psi_k [e_{k-1} - E(e_{k-1})] + K_k V_k,$$

$$= \Psi_k \Delta e_{k-1} + K_k V_k.$$

Thus, for all $k \geq k_0$, the actual Q_k is given by

$$Q_k^a = E(\Delta e_k \Delta e_k^t) = \Psi_k E(\Delta e_{k-1} \Delta e_{k-1}^t) \Psi_k^t + K_k R_k K_k^t$$

$$= \Psi_k Q_{k-1}^a \Psi_k^t + K_k R_k K_k^t.$$

Since it is assumed that the filter was optimal up to the step k_0-1, it is true that, for $k = k_0$,

$$Q_k^a = \Psi_k Q_{k-1}^m \Psi_k^t + K_k R_k K_k^t = Q_k^m,$$

where Q_k^m denotes the covariance computed in the filter, and the matrix Ψ_k involves Kalman gains K_k which are optimal at this step. Next it is shown that if at any step $k \geq k_0$, $Q_k^a = Q_k^m$, then it follows that $Q_{k+1}^a = Q_{k+1}^m$.

At step k, $Q_k^a = Q_k^m$ means that the gain at the next step is optimal, because

$$K_{k+1} = Q_k H_{k+1}(H_{k+1}^t Q_k H_{k+1} + R_k).$$

The Ψ_{k+1} matrix is then optimal, since by definition

$$\Psi_{k+1} = (I - K_{k+1} H_{k+1}^t),$$

and H is an actual matrix consisting of past inputs and outputs. Therefore,

$$Q_{k+1}^a = \Psi_{k+1} Q_k^m \Psi_{k+1}^t + K_{k+1} R_{k+1} K_{k+1}^t = Q_{k+1}^m.$$

Thus, it has been shown that if there is a jump (ΔX) at step k_0, then although the mean of the estimation error is shifted from zero, the covariance of the estimation error remains unchanged.

Corollary to theorem 3.4:

If a jump is not accompanied by a change in the noise variance, then the covariance of the normalized residual, Q_{z_k}, is not

affected by it.

Proof: Since

$$\underset{z_k}{Q} = H_k^t \, Q_{k/k-1} H_k^t + R_k,$$

and $Q_{k/k-1}$ remains unaffected, it follows that $\underset{z_k}{Q}$ will remain unaffected as long as R_k does not change.

Remark 3: In some particular situations it may so happen that the bias in the normalized residual, which is a function of the mean error, keeps decreasing consistently. This means the conditional probability of detection also decreases. This case is discussed in theorem 3.5.

Theorem 3.5: (A sufficient condition for the bias to be a decreasing sequence.) If at each step, for all vectors b which are non-orthogonal to H_k , the signs of $b^t H_k$ and K_k^t $(I + \Psi_k^t)$ b are the same, then the bias is bounded above by a decreasing sequence.

Proof: In order to prove this theorem, the following two lemmas are required.

Lemma 7: (Since the expressions are valid for each step k, the subscript k is dropped temporarily.) If for all vectors b which are non-orthogonal to H, the signs of $b^t H$ and $K^t(I + \Psi^t)b$ are the same, then $\|\Psi b\| < \|b\|$.

Proof: Assume that the signs of $b^t H$ and $K^t(I + \Psi^t)b$ are the same. Then,

$$b^t H K^t(2I - K H^t)b > 0$$

$$\Rightarrow \quad -b^t HK^t b + b^t HK^t KH^t b - b^t KH^t b + b^t b < b^t b$$

$$\Rightarrow \quad -b^t HK^t (I - KH^t) b + b^t (I - KH^t) b < b^t b$$

$$\Rightarrow \quad b^t (I - HK^t)(I - KH^t) b < b^t b$$

$$\Rightarrow \quad (\Psi b)^t \Psi b < b^t b$$

$$\Rightarrow \quad \|\Psi b\| < \|b\|.$$

Lemma 8: The sequence of bias in the normalized residual is bounded above by the norm of the mean estimation error times a constant.

Proof: The bias in the normalized residual is given by

$$M_k = \frac{1}{\sigma_k} |E(\tilde{Z}_k)|$$

$$= \frac{1}{\sigma_k} |H_k^t E(e_{k-1}) + H_k^t \Delta X_k|,$$

where σ_k is the standard deviation of the residual, with

$$\sigma_k^2 = Q_{\tilde{z}_k}.$$

Assuming that the jump occured at step k_0, the bias can be written as:

$$M_k = \begin{cases} \frac{1}{\sigma_k} |H_k^t \ \Delta X_k| & \text{for } k = k_0 \\ \frac{1}{\sigma_k} |H_k^t E(e_{k-1})| & \text{for } k > k_0. \end{cases} \qquad (3.14)$$

By theorem 3.4, the covariance Q_k does not depend on the occurance of jumps in parameters. So σ_k also does not depend on the jump. WLOG assuming that $\|H_k\|$ is bounded, and noting that $\frac{1}{\sigma_k}$ is bounded by $\frac{1}{\sqrt{R_k}}$,

the following is true:

$$M_k = \frac{1}{\sigma_k} |H_k^t E(e_{k-1})| \text{ for } k > k_0$$

$$\leq \frac{1}{\sigma_k} ||H_k|| \, ||E(e_{k-1})|| \qquad (3.15)$$

Thus the bias is dominated by a bounding sequence.

Proof of theorem 3.5: It has been shown earlier (in theorem 3.3) that

$$E(e_k) = \Psi_k \cdots \Psi_{k_0} \Delta X.$$

Assuming the non-orthogonality condition to hold at each step for ΔX, it follows from lemma 7 that

$$||\Delta X|| > ||\Psi_{k_0} \Delta X|| > \cdots > ||\Psi_k \cdots \Psi_{k_0} \Delta X|| = ||E(e_k)||.$$

Thus, $||E(e_k)||$ is a decreasing sequence. Then it follows from lemma 8 that since the bias M_k is bounded above by $||E(e_k)||$, and since $||E(e_k)||$ is decreasing, then M_k is dominatated by a decreasing sequence.

Theorem 3.6: If at some step k_0 the RWM filter is started although the parameters are constant, and if the CPM filter has been operating optimally until that time, then the computed error covariance is the sum of the actual error covariance and a positive semidefinite matrix. i.e.,

$$Q_k^m - Q_k^a = B_k, \qquad (3.16)$$

where

Q_k^m = covariance computed in the model,

Q_k^a = actual covariance,

and

$$B_k = \begin{cases} 0, & \text{if } k = k_0 \\ \Psi_k(B_{k-1} + S_{k-1})\Psi_k^t, & \text{if } k > k_0. \end{cases} \tag{3.17}$$

Proof: The proof is by induction. First it is proved for $k_0 + 1$. By definition,

$$Q_k = \Psi_k Q_{k/(k-1)} \Psi_k^t + K_k R_k K_k^t.$$

This equation is true for both actual and computed covariances, for optimal or suboptimal gains. It is assumed that

$$Q_k^a = Q_k^m \quad \text{for } k = k_0.$$

Then the projected covariance is

$$Q^m_{(k0+1)/k0} = Q^m_{k_0} + S_{k_0},$$

$$= Q^a_{k_0} + S_{k_0}.$$

But the actual covariance is

$$Q^a_{(k_0+1)/k_0} = Q^a_{k_0}.$$

Then, for $k = k_0$,

$$Q^a_{k+1} = \Psi_{k+1}[\, Q^m_{(k+1)/k} - S_k]\Psi^t_{k+1} + K_{k+1} R_{k+1} K^t_{k+1}$$

$$= Q^m_{k+1} - \Psi_{k+1}(B_k + S_k)\Psi^t_{k+1},$$

with $B_k = 0$. Hence,

$$Q^a_{k-1} = Q^m_{k+1} - B_{k+1}.$$

Assume the theorem to be true for step i - 1. Hence,

$$Q^a_{i-1} = Q^m_{i-1} - B_{i-1}.$$

Showing that the relationship holds for step i will complete the proof. At step i,

$$Q^a_i = \Psi_i Q^a_{i/(i-1)} \Psi^t_i + K_i R_i K^t_i.$$

By hypothesis,

$$Q^a_{i/(i-1)} = Q^a_{i-1} = Q^m_{i-1} - B_{i-1}$$

$$= Q^m_{i/(i-1)} - (B_{i-1} + S_{i-1}).$$

Hence,

$$Q^a_i = \Psi_i [Q^m_{i/(i-1)} - (B_{i-1} + S_{i-1}) \Psi^t_i + K_i R_i K^t_i$$

$$= Q^m_i - \Psi_i (B_{i-1} + S_{i-1}) \Psi^t_i$$

$$= Q^m_i - B_i,$$

and

$$Q^m_k - Q^a_k = B_k \text{ , for } k > k_0, \qquad (3.18)$$

where B_k is positive semidefinite. This completes the proof.

Remark 4: The matrix Q^m_k remains bounded because $Q^m_k = \Psi_k (Q^m_{k-1} + S_{k-1})$, and S_k is chosen by the designer.

Theorem 3.7: If the CPM filter has been operating optimally up to step k_0 and the RWM filter is started at that time (although the

parameters are still constant), then the normalized residual at $k > k_0$ has i) zero mean, and ii) variance less than one.

Proof: The proof of the first is by induction. Since the filter is assumed optimal up to step k_0,

$$E(\tilde{Z}_k) = 0 \text{ and } E(e_{k-1}) = 0 \qquad \text{for } k = k_0.$$

The residual at the next step is given by

$$\tilde{Z}_{k+1} = H_{k+1}^t e_{k-1} - H_{k+1}^t K_k \tilde{Z}_k + V_{k+1}.$$

Taking expected values of both sides,

$$E(\tilde{Z}_{k+1}) = 0.$$

Also,

$$E(e_k) = E(e_{k-1}) - K_k E(\tilde{Z}_k) = 0,$$

so that

$$E(e_{k+1}) = E(e_k) - K_{k+1} E(\tilde{Z}_{k+1}) = 0.$$

Assume for some $i > k_0$, $E(\tilde{Z}_i) = 0$ and $E(e_i) = 0$.

Then

$$E(\tilde{Z}_{i+1}) = H_{i+1}^t E(e_i) = 0.$$

This completes the proof of part i).

For part ii), note that the variance of the normalized residual is given by

$$\text{var}(\tilde{Z}_{nk}) = \frac{H_k^t Q_{k/(k-1)}^a H_k + R_k}{H_k^t Q_{k/(k-1)}^m H_k + R_k}. \qquad (3.19)$$

But, from previous result,

$$Q^a_{k/(k-1)} = Q^a_{k-1} = Q^m_{k-1} - B_{k-1}.$$

Therefore,

$$Q^a_{k/(k-1)} = Q^m_{k/(k-1)} - S_{k-1} - B_{k-1}. \qquad (3.20)$$

Eq. (3.19) and (3.20) show that $\text{var}(\tilde{Z}_{nk}) < 1$.

Corollary to theorem 3.7:

Under the conditions of theorem 3.7: i) the false alarm probability is lower, and ii) the probability of correctly accepting H_0 is higher than the designed value.

Proof: Both of these results are based on the observation that the normal density function, when var < 1, encompasses more area inside interval $[-T_a, T_a]$ and less area outside $[-T_r, T_r]$, compared to the case when variance = 1.

Remark 5: The above properties show that when the RWM filter is invoked although there is no parameter jump, the filter will not diverge.

In the next chapter, the proposed algorithm is extended to the multiple output case. The properties of the algorithm are modified, and necessary proofs are presented.

Chapter IV

EXTENSION TO MULTI-OUTPUT SYSTEMS

Summary: In this chapter the hypothesis testing and decision rule is extended to the multioutput case. Methods for the normalization of the residual are discussed. Since the hypothesis test is based on only the current sample, it is possible to use the norm squared of the normalized residual as the test statistic. This variable is a scalar with Chi-square distribution. Moreover, since this quantity can be obtained directly from the residual vector and the inverse of its covariance matrix, the actual normalization is avoided. This technique results in significant computational reduction and simplification of the decision rule. Properties of the CFA for the multi-output case are also given in this chapter.

IV.1 Normalization of the Residual

The observation equation is

$$Z_k = H_k X_k + V_k, \qquad (4.1)$$

where

Z_k is mx1

H_k is mxn

X_k is nx1

V_k is mx1.

The residual vector is given by

$$\tilde{Z}_k = H_k e_{k-1} + V_k, \qquad (4.2)$$

with

$$E(\tilde{Z}_k) = 0,$$

and covariance matrix

$$Q_{\tilde{z}_k} = E(\tilde{Z}_k \tilde{Z}_k^t) = H_k Q_{k/k-1} H_k^t + R_k.$$

The matrix $Q_{\tilde{z}_k}$ is mxm, symmetric, positive definite.

To normalize $\tilde{Z}_k$, denote $Q_{\tilde{z}_k} = L\, L^t$. Then

$$\tilde{Z}_{nk} = L^{-1}\, \tilde{Z}_k.$$

It is easy to show that $\tilde{Z}_{nk}$ is a vector consisting of zero-mean, unit variance, independently distributed Gaussian random variables. Indeed,

$$E(\tilde{Z}_{nk}) = L^{-1} E(\tilde{Z}_k) = 0,$$

and

$$\begin{aligned} E(\tilde{Z}_{nk} \tilde{Z}_{nk}^t) &= L^{-1} E(\tilde{Z}_k \tilde{Z}_k^t)(L^{-1})^t \\ &= L^{-1} L\, L^t (L^{-1})^t \\ &= I. \end{aligned}$$

The mean of the normalized residual vector, assuming a jump of magnitude ΔX_k in the parameter vector, is given by

$$E(\tilde{Z}_{nk}) = H_k^t\, \Delta X_k + H_k^t\, E(e_{k-1}).$$

Clearly, similar to the single output case, if there is no jump at the

present step, $k = k_0$, and there was no jump previously, $k < k_0$, then $E(\tilde{Z}_{nk}) = 0$. A jump in the parameter vector creates a bias (in the normalized residual), defined below.

Definition 4.1: The bias in the normalized residual is defined, in analogy with the single output case, as the norm of the mean of the normalized residual vector. This bias is denoted by M_n. i.e.,

$$M_n = \|E(\tilde{Z}_{nk})\|.$$

IV.2 Decision Theory for the Multi-Output Case

The decision for the multi-output case can be based on either of the following approaches:

i) test each component of the vector $\tilde{Z}_{nk}$

or *ii)* test the vector $\tilde{Z}_{nk}$ as a whole.

These two approaches are described in the next two subsections.

IV.2.1 Testing Each Component of the Normalized Residual Vector

If the components are to be tested seperately, the hypothesis test must be used m times. The decision can be based on the number of failures. In this case, each component of $\tilde{Z}_{nk}$ will have its own bias, defined as

$$M_i = |E(\tilde{Z}_{nki})|.$$

Each $\tilde{Z}_{nki}$ is distributed $N \sim (0,1)$ under the null hypothesis. Thus, for each component, the testing procedure and all the associated probabi-

lities are the same as in the single output case. In other words, the multi-output case is treated as a collection of single output cases. The i^{th} test fails if

$$|\tilde{Z}_{nki}| > T_r.$$

The i^{th} test passes if

$$|\tilde{Z}_{nki}| < T_a.$$

If all measurements are equally important, a possible decision rule is the following:

a. H_0 is accepted if at least j of the m tests pass,

b. H_0 is rejected if at least k of the m tests fail,

c. Otherwise, H_0 is neither accepted nor rejected.

In such decision theory, the tests are Bernoulli trials, and the various probabilities can be computed accordingly.

Let α be the probability of false alarm for each test. The actual probability of false alarm is

$$\Pr(k \text{ or more tests fail}/\bar{J}) = \sum_{i=k}^{m} \binom{m}{i} \alpha^i (1 - \alpha)^{m-i}.$$

The probability of rightfully accepting H_0 is

$$\Pr(j \text{ or more tests pass}/\bar{J}) = \sum_{i=j}^{m} \binom{m}{i} \delta^i (1 - \delta)^{m-i},$$

where $\delta = \Pr(F/\bar{J})$, is the probability of passing for each test. The other probabilities can also be computed in a straight forward, though lengthy, manner.

IV.2.2 Testing the Norm-Squared of the Normalized Residual Vector

If the vector $\tilde{Z}_{nk}$ is to be tested as a whole, then only one hypothesis test is necessary. For the present purpose this is a better approach because:

a. It is generally true that jumps in any parameter can result in deflecting any one, or more of the components of the measurement vector. Therefore, each component of the residual is not of separate interest.

b. If $\tilde{Z}_{nk}$ is tested as a whole, it results in significant computational simplification. The following theorem illustrates the point.

Theorem 4.1: If $\tilde{Z}_{nk}^t \tilde{Z}_{nk}$ is used as the test statistic, then the test statistic can be computed without actually computing the matrices L, L^{-1} and vector $\tilde{Z}_{nk}$.

Remark 1: This theorem implies that one matrix square-root extraction and one matrix inversion can be avoided at each time step.

Proof: The test statistic is given by

$$R_t = \tilde{Z}_{nk}^t \tilde{Z}_{nk} = (L^{-1}\tilde{Z}_k)^t L^{-1}\tilde{Z}_k$$

$$= \tilde{Z}_k^t L^{-1}L^{-1} \tilde{Z}_k$$

$$= \tilde{Z}_k^t (L^2)^{-1} \tilde{Z}_k$$

$$= \tilde{Z}_k^t (Q_{\tilde{Z}_k})^{-1} \tilde{Z}_k.$$

Since the matrix $(Q_{\tilde{Z}_k})^{-1}$ is already computed in the Kalman filter

algorithm, this approach requires only one additional multiplication compared with the single output case.

Lemma 4.1: The test statistic R_t has Chi-square distribution with m degrees of freedom.

Proof: $R_t = \sum_{i}^{m} Z_{nki}$ has Chi-square distribution with m degrees of freedom because each $\tilde{Z}_{nki}$ is $N \sim (0,1)$.

Decision rule for this approach:

a. H_0 is accepted if $R_t < T_a$,

b. H_0 is rejected if $R_t > T_r$,

c. Otherwise, H_0 is neither accepted nor rejected.

If there has been a significant change in the parameter values, then it can be expected that R_t will not be Chi-square with m degrees of freedom any more. In particular, if any one or more of the components of the $\tilde{Z}_{nk}$ vector ceases to be zero-mean, then R_t will have non-central Chi-square distribution. This non-centrality can be tested for. The thresholds T_r and T_a have to be chosen from the Chi-square distribution table. The hypothesis test in this case is one tailed, as opposed to the usual two-tailed hypothesis test for Gaussian distribution. It should be noted that the bias in the test statistic is not the same as the bias in the normalized residual. The relationship between the two is given in theorem 4.2. The following definition is required before the theorem is stated.

Definition 4.2: The bias in the test statistic, desoted by M_t, is defined as the shift in its mean value. i.e.,

$$M_t = E(R_t) - m.$$

Theorem 4.2: If the filter is optimal up to step $k_0 - 1$, and a jump ΔX occurs at step k_0, and the noise statistics do not change, then the bias in the test statistic is given by the square of the bias in the normalized residual. i.e.,

$$M_t = M_n^2.$$

Proof: The residual vector is given by

$$\tilde{Z}_k = H_k^t(\Delta X_k + e_{k-1}) + V_k. \qquad (4.3)$$

Assuming that a jump of magnitude ΔX occurs at step $k = k_0$, and is not detected,

$$E(\tilde{Z}_k) = \begin{cases} H_k^t \Delta X, & \text{for } k = k_0 \\ H_k^t E(e_{k-1}), & \text{for } k > k_0. \end{cases}$$

The mean value of the test statistic is the following:

$$E(R_t) = E(\tilde{Z}_{nk}^t \tilde{Z}_{nk})$$

$$= \text{Trace } E\{\tilde{Z}_{nk}\tilde{Z}_{nk}^t\}$$

$$= \text{Trace } E(L^{-1}\tilde{Z}_k\tilde{Z}_k^t {L^{-1}}^t).$$

Since $E(e_{k-1}) = 0$ at step k_0, and $\Delta X_k = 0$ at steps $k > k_0$, the cross products in the expression for $E(\tilde{Z}_k\tilde{Z}_k^t)$ vanish for all steps $k \geq 0$. Then

$$E(R_t) = \text{Trace } L^{-1}\{H_k^t[E(e_{k-1}e_{k-1}^t) + \Delta X \Delta X^t]H_k + R_k\}{L^{-1}}^t.$$

Also, for the step $k = k_0$,

$$E(e_{k-1}e_{k-1}^t) = Q_{k-1}^m, \tag{4.4}$$

and

$$H_k^t Q_{k-1}^m H_k + R_k = L\,L^t. \tag{4.4a}$$

Thus

$$E(R_t) = \text{Trace } \{I + L^{-1}H_k^t\, \Delta X \Delta X^t H_k L^{-1^t}\}.$$

The bias in the test statistic, M_t, for the step $k = k_0$ is then given by

$$E(R_t) - m = \text{Trace } \{L^{-1}H_k^t\, \Delta X \Delta X^t H_k L^{-1^t}\}. \tag{4.5}$$

For the steps $k > k_0$, $\Delta X_k = 0$; but $E(e_{k-1}) \neq 0$. Therefore, eq. (4.4) does not hold. Then, however,

$$E(R_t) = \text{Trace } L^{-1}\{H_k^t\, E(e_{k-1}e_{k-1}^t)H_k + R_k\}L^{-1^t} \tag{4.6}$$

But since

$$E(e_{k-1}e_{k-1}^t) = Q_{k-1} + E(e_{k-1})E(e_{k-1}^t),$$

and by theorem 3.4,

$$Q_k^a = Q_k^m \qquad \text{for all } k \geq k_0,$$

eq. (4.6) can be written as

$$E(R_t) = \text{Trace } L^{-1}[Q_{z_k}(\text{model}) + H_k^t\, E(e_{k-1})E(e_{k-1}^t)H_k]L^{-1^t}$$

$$= \text{Trace } \{I + L^{-1}H_k^t\, E(e_{k-1})E(e_{k-1}^t)H_k L^{-1^t}\}$$

Therefore, for steps $k > k_0$,

$$E(R_t) - m = \text{Trace } \{L^{-1}H_k^t \, E(e_{k-1})E(e_{k-1}^t)H_k L^{-1^t}\}. \quad (4.7)$$

The square of the bias in the normalized residual is given by

$$M_n^2 = \|E(\tilde{Z}_{nk})\|^2 = E(\tilde{Z}_{nk})^t E(\tilde{Z}_{nk})$$

$$= \text{Trace } E(\tilde{Z}_{nk})E(\tilde{Z}_{nk})^t$$

$$= \text{Trace } L^{-1}E(\tilde{Z}_k)E(\tilde{Z}_k)^t L^{-1^t}.$$

Thus, for step $k = k_0$,

$$M_n^2 = \text{Trace } \{L^{-1}H_k^t \, \Delta X \Delta X^t H_k L^{-1^t}\}. \quad (4.8)$$

For steps $k > k_0$,

$$M_n^2 = \text{Trace } L^{-1}\{H_k^t \, E(e_{k-1})E(e_{k-1}^t)H_k L^{-1^t}. \quad (4.9)$$

Comparing eq. (4.5) and (4.8), and (4.7) and (4.9),

$$M_t = M_n^2 \qquad \text{for } k \geq k_0,$$

which is the desired result.

IV.3 Properties of the CFA for the Multi-Output Case

Some properties of the CFA presented in chapter III are valid for the multi-output case without any change. These are lemmas 1 and 6, theorems 3.2, 3.3, 3.4, and 3.6. If the normalized residual vector is treated component by component, than lemmas 1 through 6 are valid

without any change. The only difference is that 'bias' means each component of the bias vector, and 'normalized residual' means each component of the $\tilde{Z}_{nk}$ vector. However, if the test statistic is $R_t = \tilde{Z}_{nk}^t \tilde{Z}_{nk}$, then the statements of lemmas 2, 3, and 4 are valid, while their proofs require minor adjustments. In these proofs, the lower limits of integration $-T_a$, $-T_a-M$, $-T_r$ and $-T_r-M$ must be replaced by 0, and the Gaussian density by the Chi-square density function. Lemma 5 can not be stated because uniform distribution for the bias can not be assumed. Regardless of the approach taken for the test statistic, theorems 3.1, 3.5 and 3.7 require modifications, which are presented below.

Theorem 4.3: (Modification of theorem 3.1 for multi-output case.) If there is a parameter jump accompanied by a change in the noise variance such that $\Delta R_k = R_k(\text{actual}) - R_k(\text{model})$ is positive (negative) definite, then

i) the normalized residual has non-zero mean,

ii) its variance is the sum of the identity and a positive (negative) definite matrix.

Furthermore, if ΔR_k is positive, then

iii) the mean square value of the normalized residual is the sum of the identity and a positive definite matrix.

Proof: For part i), observe that

$$\tilde{Z}_{nk} = L^{-1} \tilde{Z}_k,$$

and following a jump of magnitude ΔX_k,

$$E(\tilde{Z}_k) = H_k^t \Delta X_k.$$

Therefore,

$$E(\tilde{Z}_{nk}) = L^{-1} H_k^t \Delta X_k \neq 0.$$

For part ii), note that according to lemma 6 of chapter III, the following expression holds:

$$E(\tilde{Z}_k \tilde{Z}_k^t) = H_k^t Q^m_{k/k-1} H_k + R_k + H_k^t \Delta X_k \Delta X_k^t H_k.$$

Also,

$$E(\tilde{Z}_{nk} \tilde{Z}_{nk}^t) = L^{-1} E(\tilde{Z}_k \tilde{Z}_k^t) L^{-1^t}.$$

However,

$$R_k(\text{actual}) = R_k(\text{model}) + \Delta R_k,$$

so that

$$E(\tilde{Z}_{nk} \tilde{Z}_{nk}^t) = L^{-1} Q_{\tilde{Z}_k}(\text{model}) L^{-1^t} + L^{-1}(\Delta R_k + H_k^t \Delta X_k \Delta X_k^t H_k) L^{-1^t}.$$

Denote

$$P = \Delta R_k + H_k^t \Delta X_k \Delta X_k^t H_k. \qquad (4.10)$$

Then

$$E(\tilde{Z}_{nk} \tilde{Z}_{nk}^t) = I + L^{-1} P L^{-1^t}, \qquad (4.11)$$

and hence

$$\text{Var}(\tilde{Z}_{nk}) = E(\tilde{Z}_{nk} \tilde{Z}_{nk}^t) - E(\tilde{Z}_{nk}) E(\tilde{Z}_{nk})^t,$$

$$= I + L^{-1}(P - H_k^t \Delta X_k \Delta X_k^t H_k) L^{-1^t}$$

$$= I + L^{-1} \Delta R_k L^{-1^t}.$$

Thus

$$\text{Var}\,(\tilde{Z}_{nk}) = I + L^{-1}\Delta R_k L^{-1^t}. \qquad (4.12)$$

Eq. (4.12) proves part ii) of the theorem. For the proof of part iii), note that P is positive if in eq. (4.10), ΔR_k is positive. Then eq. (4.11) proves part iii) of the theorem.

Theorem 4.4: (Modification of theorem 3.5 for multi-output case.) If at each step the norm of the Ψ_k matrix is less than one, then the bias in the normalized residual is bounded above by a decreasing sequence.

Proof: The bias in the normalized residual is given by

$$M_n = \begin{cases} \|L^{-1}H_k^t \Delta X\| & \text{for } k = k_0 \\ \|L^{-1}H_k^t E(e_{k-1})\| & \text{for } k > k_0. \end{cases}$$

Assuming that the norms of L^{-1} and H_k are bounded by some constants, it is clear that M_n is bounded by a constant times the norm of $E(e_{k-1})$. Then it has to be proved that $\|E(e_k)\|$ is a decreasing sequence. By theorem 3.3 and the fact that $\|\Psi_k\| < 1$, the following is true:

$$\|\Delta X\| > \|\Psi_{k_0} \Delta X\| > \cdots > \|\Psi_k \cdots \Psi_{k_0} \Delta X\| = \|E(e_k)\|.$$

Thus, $\|E(e_k)\|$ is a decreasing sequence. This completes the proof of the theorem.

Theorem 4.5: (Modification of theorem 3.7) If the CPM filter has been operating optimally up to step k_0 and the RWM filter is started at that time (although the parameters are still constant), then the normalized residual

i) is zero mean, and

ii)the norm of its covariance matrix is less than one

for all steps $k > k_0$.

Proof: The proof is the same as for the single output case, except for one change in the eq. (3.19). That equation has to re-written for the multi-output case in the following way:

$$\mathrm{Var}(\tilde{Z}_{nk}) = L^{-1}(H_k^t\, Q^a_{k/k-1} H_k + R_k)L^{-1^t}. \qquad (4.13)$$

But eq. (3.20) is valid for the multi-output case,

$$Q^a_{k/k-1} = Q^m_{k/k-1} - S_{k-1} - B_{k-1},$$

where the matrix $(S_k + B_k)$ is positive definite. Therefore,

$$\mathrm{Var}(\tilde{Z}_{nk}) = L^{-1}[Q_{\underset{z_k}{\sim}}(\text{model}) - (S_{k-1} + B_{k-1})]L^{-1^t}$$

$$= I - L^{-1}(S_{k-1} + B_{k-1})L^{-1^t},$$

which proves the theorem. The corollary to theorem 3.7 is also valid for the multi-output case. Simply replace the lower limits $-T_a$ amd $-T_r$ by 0, and normal density by Chi-square, and observe that Var $(\tilde{Z}_{nk}) < I$ implies that $E(R_t) < m$.

Remark 2: The single output case can also be analyzed as a special case of the multi-output situation. Hypothesis tests for the single output case can be designed based on the Chi-squared distribution. However, there is no computational or conceptual advantage over the technique developed in the previous chapters.

Chapter V

CHANGES IN THE NOISE STATISTICS

Summary: In the previous chapters the Combined Filter Algorithm has been developed, and its properties have been studied. The underlying assumption in all the previous discussion has been that the measurement noise statistics do not change from the initially assumed values. In this chapter this assumption is removed, and the effects of a jump in the *i)* mean of the noise process and *ii)* variance of the noise process are studied.

First, the case of a change in the noise mean is discussed. Next, the case of a jump in the noise variance is studied. It is found that although the algorithm can not distinguish between jumps in the system parameters and jumps in the noise statistics, the estimates remain bounded even under the worst possible circumstances, namely, a chain of wrong detections.

V.1 The Noise Becomes Non-Zero Mean

Assumption 5.1: There has been no jump in the parameter vector at step k or at any previous step k-j, for j = 1,2, •••, k, and the filter has been operating optimally upto the step k.

Theorem 5.1: If at some time instant k the measurement noise becomes a non-zero mean process satisfying the following equation:

$$V_k = \mu_v + \xi_k, \tag{5.1}$$

with

$$E(\xi_k) = 0, \tag{5.2}$$

$$\mu_v \neq 0,$$

and

$$E(\xi_k \xi_k^t) = R_k, \tag{5.3}$$

(in other words, the variance of the noise process does not change), then:

i) the estimates of the parameters become biased,

ii) the variance of the estimation error does not change,

iii) the probability of false alarm increases.

Proof: i) The actual observations are given by

$$Z_k = H_k^t X_k + \mu_v + \xi_k. \tag{5.4}$$

Replacing this expression in the estimation equation, one gets

$$\hat{X}_k = \hat{X}_{k-1} + K_k [H_k^t X_k + \mu_v + \xi_k - H_k^t \hat{X}_{k/k-1}]. \tag{5.5}$$

Taking expected values and considering $X_k = X_{k-1}$ (according to assumption 5.1), one can easily show that

$$E(\hat{X}_k) = E(\hat{X}_{k-1}) + K_k \mu_v$$

$$= X_k + K_k \mu_v, \tag{5.6}$$

where $E(\hat{X}_{k-1}) = X_{k-1}$ because the estimates are unbiased at each step before k. Thus,

$$E(e_k) = X_k - \hat{X}_k$$
$$= K_k \mu_v$$
$$\neq 0,$$

i.e., the estimate ceases to be unbiased.

Therefore, a shift in the noise mean is treated by the filter just like a jump in the parameter vector, with the consequence that the normalized residual $\tilde{Z}_{nk}$, $j \geq k$, is not independent, standard normal. In chapter III it has been shown that the probability of detection is an increasing function of the bias in the normalized residual. In the present context, the probability of wrong detection or false alarm increases.

ii) The mean of the residual is shifted from zero, and is given by

$$E(\tilde{Z}_k) = E(Z_k - H_k^t \hat{X}_{k/(k-1)})$$
$$= E(H_k^t X_{k-1} + \mu_v + \xi_k - H_k^t \hat{X}_{k-1})$$
$$= \mu_v.$$

Also, it can be shown by straightforward computation that

$$E(\tilde{Z}_k \tilde{Z}_k^t) = H_k^t Q_{k/(k-1)} H_k + R_k + \mu_v \mu_v^t \quad (5.7)$$

and

$$Q_{\tilde{z}_k} = E(\tilde{Z}_k \tilde{Z}_k^t) - \mu_v \mu_v^t$$
$$= H_k^t Q_{k/(k-1)} H_k + R_k.$$

This expression is the same as eq.(2.4), showing that the variance of the residual does not change. Hence,

$$\tilde{Z}_k \sim N(\mu_v, Q_{\tilde{z}_k}) .$$

Thus the normalized residual, $\tilde{Z}_{nk}$ will not be zero-mean, but will be shifted with respect to the origin. The shift is

$$E(\tilde{Z}_{nk}) = L^{-1} \mu_v, \text{ where } Q_{\tilde{z}_k} = L L^t.$$

All the properties described in chapter III and IV which concern the bias (in the normalized residual) are applicable here. Some expressions have to be re-written, replacing $H_k^t \Delta X$ by μ_v. It must be mentioned that the algorithm has no way of distinguishing between a bias created by a parameter jump and one created by a shift in the noise mean. The variance of the normalized residual is not affected by the change in the noise mean.

iii) In order to prove the properties concerning the probability of false alarm, define

$$\alpha^0 = Pr(|\eta| > T_r),$$

where η is N (0,1). The actual probability of false alarm will be

$$\alpha = Pr(|\tilde{Z}_{nk}| > T_r).$$

To show that $\alpha > \alpha^0$ it is enough to show that

$$Pr(|\tilde{Z}_{nk}| > T_r) < Pr(|\eta| < T_r).$$

Let

$$\bar{\eta} = \tilde{Z}_{nk} - E(\tilde{Z}_{nk}).$$

Clearly $\bar{\eta}$ is N(0,1), and

$$\Pr(|\tilde{Z}_{nk}|<T_r) = \Pr[-T_r - E(\tilde{Z}_{nk}) < \bar{\eta} < T_r - E(\tilde{Z}_{nk})].$$

By the nature of the normal probability density curve, it is true that

$$\Pr(-T_r - M < \eta < T_r - M) < \Pr(-T_r < \eta < T_r) \text{ for any } M.$$

Hence,

$$1 - \alpha < 1 - \alpha^0.$$

Thus, $\alpha > \alpha^0$, and the theorem is established.

V.2 The Noise Variance Changes

Theorem 5.2: An increase(decrease) in the noise variance implies an increase(decrease) in the probability of false alarm.

Proof: Let

$$E(V_k V_k^t) = R_k + \Delta R_k.$$

Then, it is easy to show that

$$\begin{aligned} Q_{\tilde{z}_k} &= H_k^t Q_{k-1} H_k + R_k + \Delta R_k \\ &= Q^0_{\tilde{z}_k} + \Delta R_k. \end{aligned}$$

The normalized residual is not standard normal any more, and

$$1 - \alpha = \Pr(-T_r < \tilde{Z}_{nk} < T_r)$$

$$= \Pr\left[\frac{-T_r \sqrt{Q^0_{\tilde{z}_k}}}{\sqrt{Q^0_{\tilde{z}_k} + \Delta R_{2k}}} < \eta < \frac{T_r \sqrt{Q^0_{\tilde{z}_k}}}{\sqrt{Q^0_{\tilde{z}_k} + \Delta R_{2k}}} \right].$$

Let $F = \dfrac{\sqrt{Q^0_{\tilde{z}_k}}}{\sqrt{Q^0_{\tilde{z}_k} + \Delta R_{2k}}}$; clearly $F < 1$ for $\Delta R_{2k} > 0$, i.e.

if noise variance is increased. Consequently

$$[-T_r.F,\ T_r.F] \subseteq [-T_r,\ T_r],$$

and

$$1 - \alpha = \Pr(-T_r.F < \eta < T_r.F)$$

$$< \Pr(-T_r < \eta < T_r)$$

$$= 1 - \alpha^0,$$

or, $\alpha > \alpha^0$.

On the other hand, if $\Delta R_k < 0$, i.e. noise variance is decreased, then $F > 1$. In that case

$$[-T_r,\ T_r] \subseteq [-T_r\ F,\ T_r\ F]$$

and

$$1 - \alpha = \Pr(-T_r\ F < \eta < T_r)$$

$$> \Pr(-T_r < \eta < T_r) = 1 - \alpha^0.$$

Hence,

$$\alpha < \alpha^0.$$

Thus, the theorem is established.

V.3 Analysis of Incorrect Detection

In the past sections it has been shown that the probability of false alarm increases with an increase in the measurement noise variance or a shift in its mean. This can pose a potential danger to the estimation process. The worst possible case is when a sequence of re-initializations occur, without any updating of the estimates. The following is an analysis of the behavior of the algorithm in such a situation.

Let the first re-initialization occur at step k. This means the estimation error covariance Q_k is returned to a fixed large value, Q_0. Then,

$$K_k = Q_0 H_k[R_{2k} + H_k^t Q_{k-1} H_k]^{-1}$$

is also large, and

$$\hat{X}_k = \hat{X}_{k-1} + K_k \tilde{Z}_k$$

is far removed from the true value of the parameters, and also from the latest estimates just before the re-initialization took place. However, a large gain is not desirable in this case, because this was a false alarm (probably due to an outlier, or due to a change in the noise mean or variance).

Theorem 5.3: Even if there is a sequence of re-initializations, the Kalman gain remains bounded as long as the observation matrix is bounded. Furthermore, Q_0 can be chosen such that

$$0 \leq KH^t \leq I.$$

Proof: It will be shown that for a particular choice of Q_0, the norm of the matrix KH is bounded by one. (Dropping the subscripts for K, R and H temporarily),

$$K = Q_0 H[R + H^t Q_0 H]^{-1}$$

$$K(R + H^t Q_0 H) = Q_0 H$$

$$K R = (I - K H^t) Q_0 H$$

$$K = \Psi Q_0 H R^{-1}$$

$$\Psi = I - K H^t = I - \Psi Q_0 H R^{-1} H^t$$

Therefore,

$$\Psi^{-1} = I + Q_0 H R^{-1} H^t.$$

Since Q_0 is a design parameter, it can be chosen to be a scalar matrix, namely $Q_0 = q I$. Then,

$$\Psi = (I + q H R^{-1} H^t)^{-1},$$

where $H R^{-1} H^t$ is positive semidefinite. Hence

$$0 \leq \Psi \leq I.$$

Since $K = \Psi Q_0 H R^{-1}$, the boundedness of K follows. Replacing Ψ by its equivalent expression one obtains

$$0 \leq I - K H^t \leq I,$$

which is the same as

$$0 \leq K H^t \leq I.$$

Thus, the theorem is proved.

Corollary to theorem 5.3:

If the chain of wrong detections occur as a result of outliers,

then the estimates remain bounded and they return assymptotically to the true values.

Proof: The wrong detections occuring as a result of outliers means that there was neither any parameter jump nor any jump in the noise statistics. Therefore, no model mismatch exists, the observation matrix is bounded, and all the previous assumptions about the system are valid. Thus, the Kalman gains are bounded (by theorem 5.3) and as a result the estimates are bounded also. After a number of steps the effect of the large Q_k will vanish and the estimates will return to the previous (true) values.

The above discussion shows that there is no danger of filter divergence if the measurement noise statistics change. The worst possible result of such changes is a chain of false alarms. It has been shown that a chain of false alarms does not result in filter divergence. The main drawback of the algorithm is that it can not differentiate between a failure of the hypothesis test due to an actual parameter jump and one due to some shift in the noise mean or variance, or due to an outlier. This problem is regarded as an interesting topic for further research.

In the next chapter, the application of the CFA for power system state estimation is discussed. It is shown that power system faults, being accompanied by large amounts of noise (transients), are ideally suited for single-sample hypothesis tests.

Chapter VI

APPLICATION OF THE COMBINED FILTER TO POWER SYSTEM STATE ESTIMATION

Summary: In this chapter the application of the Combined Filter Algorithm is demonstrated for power system fault detection. It is also used to estimate pre- and post-fault voltages and currents. Kalman filter has been applied for power system state estimation by some other researchers: however, the fault is detected by a separate mechanism, and then the Kalman filter is used to get estimates of the steady state values which are then utilized in order to determine the fault location. In this work it is shown that a power system fault is ideally suited for single sample hypothesis testing, making the CFA a good candidate to be used for the detection and estimation purposes. Computer simulation results are included.

VI.1 General Background and a Brief Literature Survey

When a fault (short circuit) occurs in a power transmission line, the distance to the fault location needs to be determined as quickly as possible so that proper decisions can be made as to which relays should trip, and where. Isolating the fault helps protect the remainder of the system, limit damage to major equipment and preserve system stability [24]. To locate the fault position the steady state values of the post-fault currents and voltages must be estimated.

During the last decade many methods have been developed to predict the fundamental frequency components of the current and voltage. The following brief survey of the different methods that have been suggested is based mainly on a 1981 paper by Girgis and Brown [28] in which they proposed the application of Kalman filtering for this problem. They mentioned most of the previous work done in this area before 1981. Some more papers have appeared on his topic since then, which have been added to the present survey. Mann and Morrison [37] used the fact that the amplitude and phase angle of a pure sinusoid can be determined from its value and rate of change at any instant using difference equations, and suggested an algorithm based on this. Gilcrest, Rockefeller and Udren [38] developed an algorithm based on first and second derivatives in order to reduce sensitivity to dc offset and other components below power system frequency. Makino and Miki [39] used a two-sample sinusoidal curve fit to obtain the peak squared values. Digital filters were used to attenuate the dc offset and high frequency components.

In a different approach, Ramamoorty [40] suggested the use of a full cycle Fourier transform. Phadke, Hiibka and Ibrahim [41] used a half cycle Fourier transform in order to reach a compromise between speed of response and sharpness of filtering. Within a very short time researchers started to investigate whether least squares methods could be effectively employed in this situation. Throp, Phadke, Horowitz and Beehler [42] showed that the half cycle Fourier transform is a least squares estimate, provided the noise signal is assumed to be stationary and white. In [43] Luckett, Munday and Murray developed an algorithm for a least squares fit to a general fault waveform. Sachdev and Baribeau in

[44] suggested a least squares fitting algorithm which solves simultaneous equations to obtain the real and imaginary parts of the fundamental componets. This algorithm neglects any frequency higher than the third order, but its accuracy is improved by using data-redundancy and pseudo-inverse techniques. Brooks [45] used still another least squares approach in which it is assumed that the waveform can be approximated by constant and fundamental frequency components. All this work paved the way for Kalman filtering to be applied for this estimation problem. Girgis and Brown [28] developed the necessary models and studied the transient components in order to tailor the problem to fit the Kalman filtering framework.

The use of Kalman filter for power system state estimation is a relatively new idea. A few papers have appeared in the last six years concerning this application. Girgis and Brown have proposed using a bank of Kalman filters for sensing the fault ([26], [27], [28]). After a fault has been identified, the ordinary Kalman filter with a time-varying observation matrix is used. They have developed appropriate models for voltages and currents, and also determined the noise statistics with reasonable accuracy. It must be mentioned here that immediately after a fault, transient components are superimposed on the steady state values. Since the steady state values are to be estimated, the transients can be modeled as random noise. They have presented simulation results to show that the Kalman filter is capable of giving very good estimates of the post-fault states in a very short time. A different type of system model has been proposed by Sachdev [25] in which the observation matrix is time-invariant, but the state model is

based on rotating reference frames. Sachdev also reports some simulation results showing the efficiency of the Kalman filter. In fact, both models work well, although the one developed by Girgis and Brown seems to follow the usual control literature more closely. Hence their model has been used in this work.

In a different application, Nishiwaki, Yokokawa and Ohtsuka proposed the use of Kalman filters for parameter identification for power system stability analysis [46]. It seems reasonable to say that in the near future Kalman filtering or its modifications will find many more applications in the general area of power system analysis. In particular, this chapter is concerned with the use of the combination filter for estimating pre- and post- fault states as well as fault detection. In the next section the typical noise situation during and after a fault is discussed. Then the system models are described, and results of simulation study are given.

In concluding this section the basic novelty of the CFA must be pointed out. It is designed as a tool capable of estimating the pre-fault states, then detecting any anomally in the system's behavior, and subsequently estimating the steady state post-fault values. In Girgis and Brown's approach one needs an extended Kalman filter for sensing the fault and an ordinary Kalman filter for estimating the states. The CFA is a single algorithm capable of handling the whole process, with considerably less complexity of on-line computation. The single-sample hypothesis testing has been shown to be very suitable for power system fault detection.

VI.2 Noise Process Associated with Power System Fault

Girgis and Brown have done extensive simulation study of the transients accompanying various types of faults for various lengths of time. In [27] and [28] they report that the transients can be approximated as nonstationary random processes. The voltage transients have high frequency components that decay exponentially with time. Therefore they are modeled as a white noise sequence with exponentially decaying variance. The current transients have high frequency components *and* a large time-constant exponential component. This is called a dc offset. This component is augmented to the state vector in the model. While it is true that the transient phenomena is affected by many factors such as line loading. structure of power systems and untransposition of transmission lines, it has been shown that [47] these effects are relatively small. Girgis and Brown simulated a simple power system which is shown in figure 6.1. It represents a 345 KV, 160 mile transmission line connected to a 400 MVA generating station at the sending end and a large interconnected network at the receiving end. The generator is represented by a constant voltage behind a subtransient reactance.

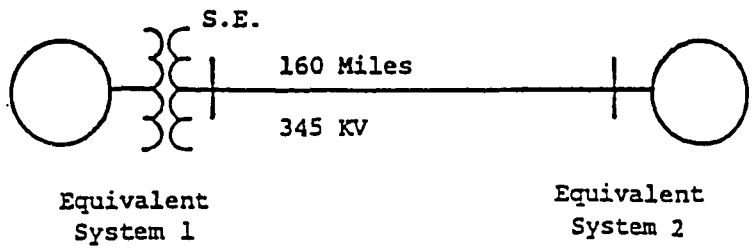

Fig. 6.1: Power system model.

The instrumentation errors were considered small compared to the high frequency transient noise. The noise in the voltage and current waveforms was determined by subtracting the steady state time solution from the transient solution in the first cycle after the fault occured. They simulated four different kinds of faults at four locations equally spaced along the transmission line. The worst condition case was studied, which is the maximum voltage at the fault location. Although the nature of the noise is slightly dependent on the type of fault, the dependence can be safely ignored. The noise was not found to change with different load conditions either. The auto-correlation function and the ensemble variance were also determined based on the frequency of occurance of different types of fault and equal probability of fault locations. Figure 6.2a shows the nature of the voltage noise variance in the faulted phase and the exponential approximation used in the Kalman filter. The variance of the voltage noise in the unfaulted phase, shown in fig. 6.2b, was found to be

reasonably within expected limits.

The following empirical formula for the variance of the voltage noise signal is proposed by Girgis and Brown:

$$R_k = k_v \, e^{-k\Delta t/T1} . \qquad (6.1)$$

The constant k_v is very closely equal to the mean square of the change in the sending end voltage in the faulted phase. T1 is roughly equal to half of the expected time constant of the protected line. Girgis and Brown suggest that there is no need to add a small constant to the R_k formula to keep it from approaching zero with time because the estimation time is small. However, it is the opinion of the present author that a small constant must be added to it since the combination filter is intended to stay in operation for a long time after the fault has been cleared and the system has reached steady state again.

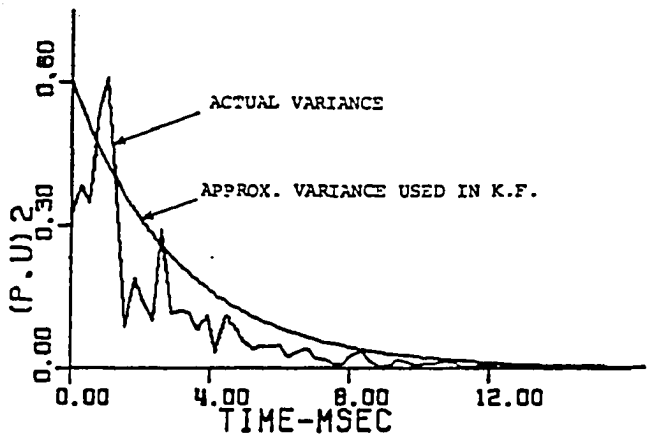

Fig. 6.2a: Variance of the voltage noise signal in the faulted phase.

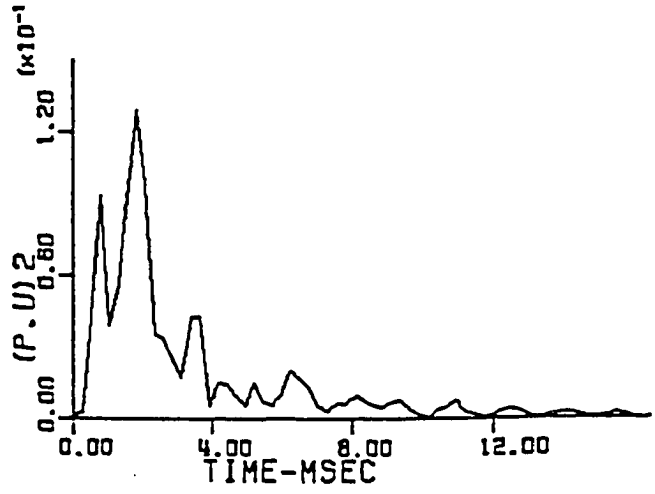

Fig. 6.2b: Variance of the voltage noise signal in the unfaulted phase.

For the current noise also, an empirical formula is given for the variance of the white noise process:

$$R_k = k_i \, e^{-k\Delta t/T1}. \qquad (6.2)$$

The constant k_i is found to be one-fourth of the mean square of the sending end current. Fig. 6.2c shows the variance of the current noise signal in the faulted phase. As mentioned earlier, the current noise consists of a white noise *and* an exponential component.

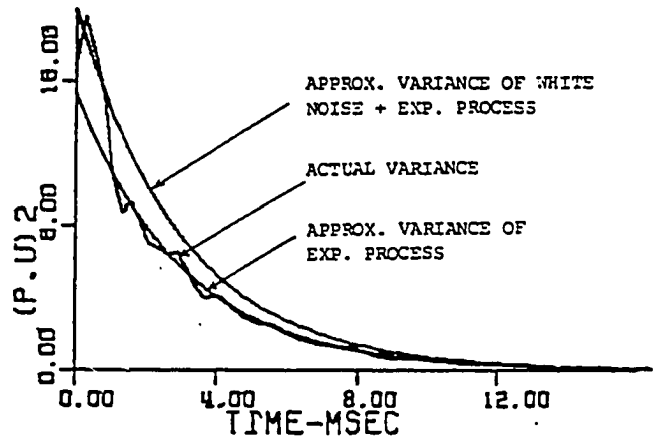

Fig. 6.2c: Variance of the current noise signal in the faulted phase.

The discrete state model for the exponential process is given by:

$$X_3(k+1) = e^{-\beta\Delta t} X_3(k) + w_k, \qquad (6.3)$$

where the initial variance is assumed to be large compared to the steady state variance. The constant β is obtained from a least-square exponential curve fit to the actual variance of the noise in the faulted phase. The initial variance σ_i^2 is approximately equal to the mean square of the sending end current in the faulted phase, and the constant β is given by 1/T1. Thus, the exponential part of the noise is incorporated in the state equatiotion. This component includes a small residual noise w_k. The variance of this w_k is obtained from experimental results. The third state of the current model is a nonstationary process with a large initial value, and it gradually relaxes to a Markov process with a relatively small rms value in the steady state condition.

VI.3 Significance of the Exponentially Decaying Noise Variance for the Jump Detector

It is interesting to note that the special kind of noise statistics in the case of power system fault actually is benificial to the jump detector. This is true because the parameter jump in this case is accompanied by a sudden large noise which decays gradually. The initial large value of the noise assists in the detection of the jump. According to theorem 3.1 of chapterIII, if a jump is accompanied by an increase in the noise variance, then the normalized residual has variance greater than unity. In chapter IV it was shown that an increase in the noise variance results in a larger probability of detection: wrong detection if there was no parameter jump, correct detection if there was. Because of this particular property, power system fault detection is an ideal case for applying hypothesis test based on a single sample. A numerical example will make this clear. Let at some time step k a fault occur, with

$$H_k^t = [\ 0.9 \quad -0.1]$$

$$\Delta X_k^t = [-0.3 \quad 0.7]$$

$$R_k(\text{actual}) = 0.6$$

$$R_k(\text{model}) = 0.03$$

$$Q_{k/(k-1)} = \begin{bmatrix} 0.02 & 0.01 \\ 0.01 & 0.02 \end{bmatrix}$$

These numerical values are not chosen in random: they comply with the fact that in power system models all the variables are in per unit (so

they are less than one), and the initial value of R_k(actual) is taken from Girghis and Brown's simulation study [28]. The value of R_k(model) is chosen somewhat larger than in [28] in order to obtain a more conservative result. Then,

$$\underset{\sim}{Q}_{z_k}(\text{model}) = H_k^t Q^m_{k/(k-1)} H_k + R_k(\text{model}) = 0.0446$$

$$\underset{\sim}{Q}_{z_k}(\text{actual}) = H_k^t Q^m_{k/(k-1)} H_k + R_k(\text{actual}) = 0.6146$$

thus,

$$\underset{\sim}{Q}_{z_{nk}} = \frac{0.6146}{0.0446} \approx 13.78$$

and

$$E(\tilde{Z}_{nk}) = \frac{H_k^t \Delta X_k}{\sqrt{\underset{\sim}{Q}_{z_k}(\text{model})}} = \frac{-0.34}{\sqrt{0.0446}} \approx -1.61.$$

Assuming that the rejection threshold is $T_r = 2.58$, chosen on the basis of 1% false alarm rate for standard normal r.v., the conditional detection probability is

$$\Pr(|\tilde{Z}_{nk}| > T_r) \approx 52.7\%.$$

To compare this result with the case when there is no increase in R_k with the jump, note that the conditional detection probability for the same amount of jump would be 16.6%. This shows the increase in R_k, in this example, results in a 217% increase in the conditional detection probability.

VI.4 Voltage and current models

(same for each phase)

The signal process for both current and voltage is a sine wave with random amplitude and phase. This is modeled as a two-element vector, where the state variables are the coefficients of the sine and cosine components of the wave. The noise free current or voltage signal can be expressed by the real or imaginary part of

$$S(t) = A \exp (j\omega_0 t + \varphi), \qquad (6.4)$$

where A is the amplitude and φ is the phase angle at t = 0. The same quantity can be expressed as

$$S(t) = X_1 \cos \omega_0 t - X_2 \sin \omega_0 t, \qquad (6.5)$$

where X_1 and X_2 are the in-phase and quadrature-phase components of the voltage or current phasor, and they are independent Gaussian random variables.

The state (voltage) model for both faulted and unfaulted phases is given by

$$\begin{bmatrix} X_1(k+1) \\ X_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1(k) \\ X_2(k) \end{bmatrix} \qquad (6.6)$$

The measurement equation for both faulted and unfaulted phases is the following:

$$Z_k = \begin{bmatrix} \cos(\omega_0 k\Delta t) & -\sin(\omega_0 k\Delta t) \end{bmatrix} \begin{bmatrix} X_1(k) \\ X_2(k) \end{bmatrix} + V_k \qquad (6.7)$$

where Δt is the sampling interval, and $\omega_0 = 2\pi f = 377\ \text{sec}^{-1}$ is the fundamental angular frequency. The current model has a three-state vector:

$$\begin{bmatrix} X_1(k+1) \\ X_2(k+1) \\ X_3(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-\beta\Delta t} \end{bmatrix} \begin{bmatrix} X_1(k) \\ X_2(k) \\ X_3(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ w_k \end{bmatrix} \qquad (6.8)$$

The measurement equation is similar to that of the voltage model,

$$Z_k = \begin{bmatrix} \cos(\omega_0 k\Delta t) & -\sin(\omega_0 k\Delta t) & 1 \end{bmatrix} \begin{bmatrix} X_1(k) \\ X_2(k) \\ X_3(k) \end{bmatrix} + V_k \qquad (6.9)$$

VI.5 Simulation Study of the Application of the Combined Filter Algorithm to the Power System State Estimation Problem

The purpose of this computer simulation work is to show the effectiveness of the proposed algorithm (with the single sample detector and combined decision rule) in power system state estimation. The voltage model is used to estimate pre- and post-fault values of the two states. The initial values of Q_k and $\hat{X}_k$, the sampling rate and the values of R_k are taken from the example studied by Girghis and Brown in [28].

The acceptance and rejection thresholds are left as variables in order to find out their effect on the performance of the algorithm. The S_k matrix is also left as a variable so that its effect on the performance of the filter can be seen. However, since in Girghis and Brown's study the Kalman filter is started only after a fault has been detected (by some other detection mechanism), it is not possible to get comparative results on the effect of the combined decision rule in this particular example. Therefore another example has been constructed where a small jump occurs, and the noise variance R_k after the fault is much lower than in the previous example. This can happen in practice if the time constant of the protected line is small.

VI.5.1 Example 1

(voltage model)

Initial values:

$\hat{X}_1 = 0.0$

$\hat{X}_2 = 0.0$

$Q_0 = I$

Pre-fault values ($k < k_0$, fault occurs at step k_0):

$X_1 = 1.0$

$X_2 = 0.0$

$R_k = 0.001.$

Post fault values ($k \geq k_0$):

$X_1 = 0.5$

$X_2 = 0.33$

$R_k = 0.6\ e^{-0.2(k-ko)} + 0.001.$

The sampling rate is 1920 samples per second, which translates into 32 samples per cycle. Girghis and Brown used 64 samples per cycle in their simulation, but later showed that the performance of the filter is practically the same for 32 or 64 samples per cycle.

The hypothesis test is based on the following.

Set i) of simulations: $T_r = 2.58$, $T_a = 1.0$.

This corresponds to $\alpha = 0.01$ and CDR.

Set ii) of simulations: $T_r = 2.58$, $T_a = 2.58$.

This corresponds to $\alpha = 0.01$ and BDR.

In 50 simulations with different sequences of random noise, the jump was detected without any delay each time. Fig. 6.3 through 6.6 illustrate the performance of the algorithm.

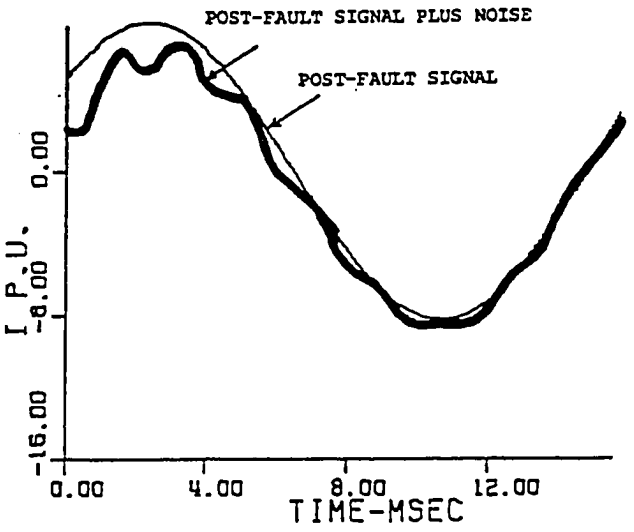

Fig. 6.3a: Post-fault signal, and signal plus noise (After Girghis and Brown [28]).

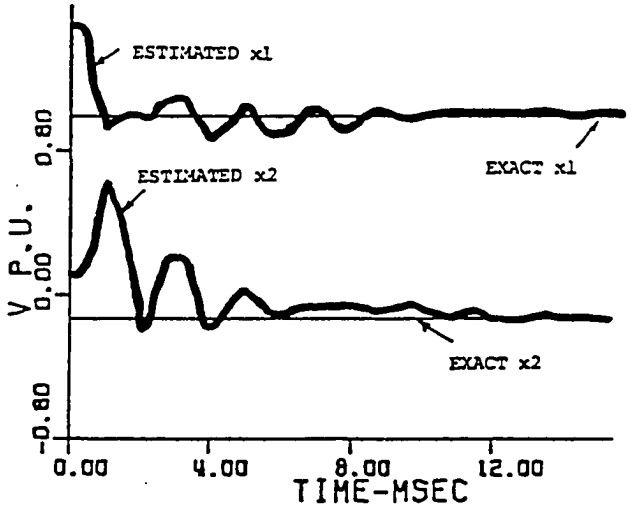

Fig. 6.3b: True values and estimates of the post-fault states (After Girghis and Brown [28]).

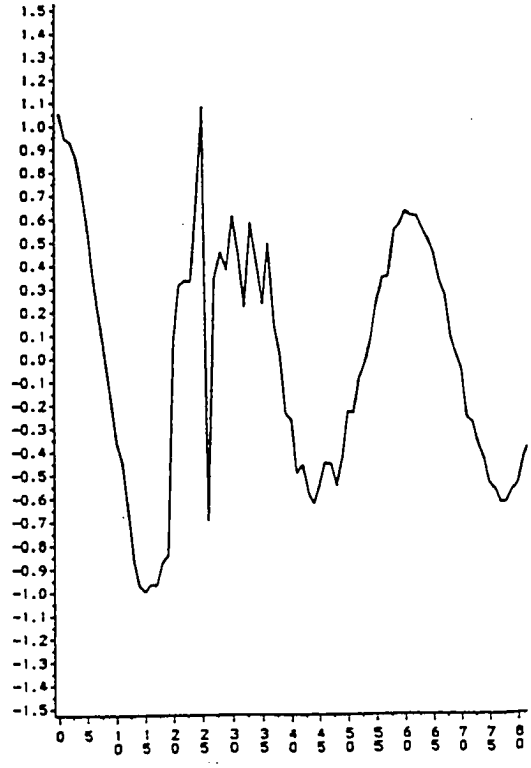

Fig. 6.4: Measured signal plus noise. At step 20, a fault occurs.

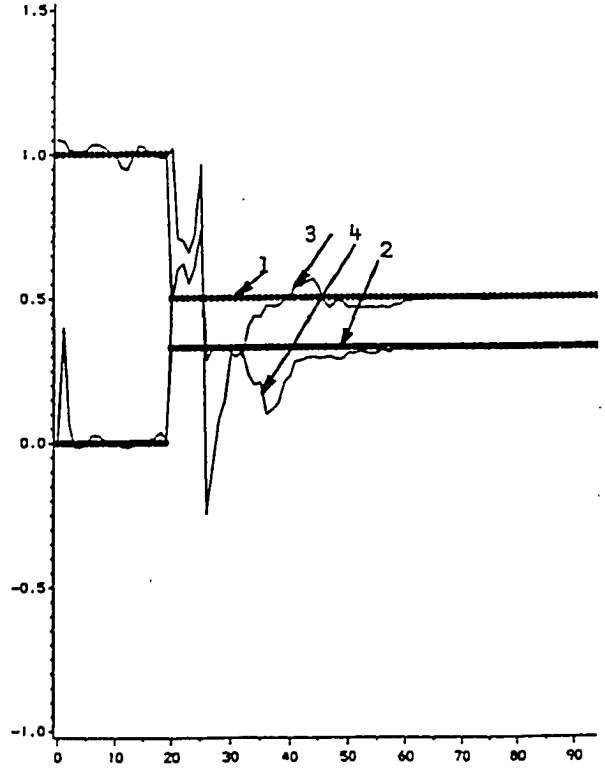

Fig. 6.5: Estimates and true values of the states.
1 - true X_1,
2 - true X_2,
3 - estimated X_1,
4 - estimated X_2.

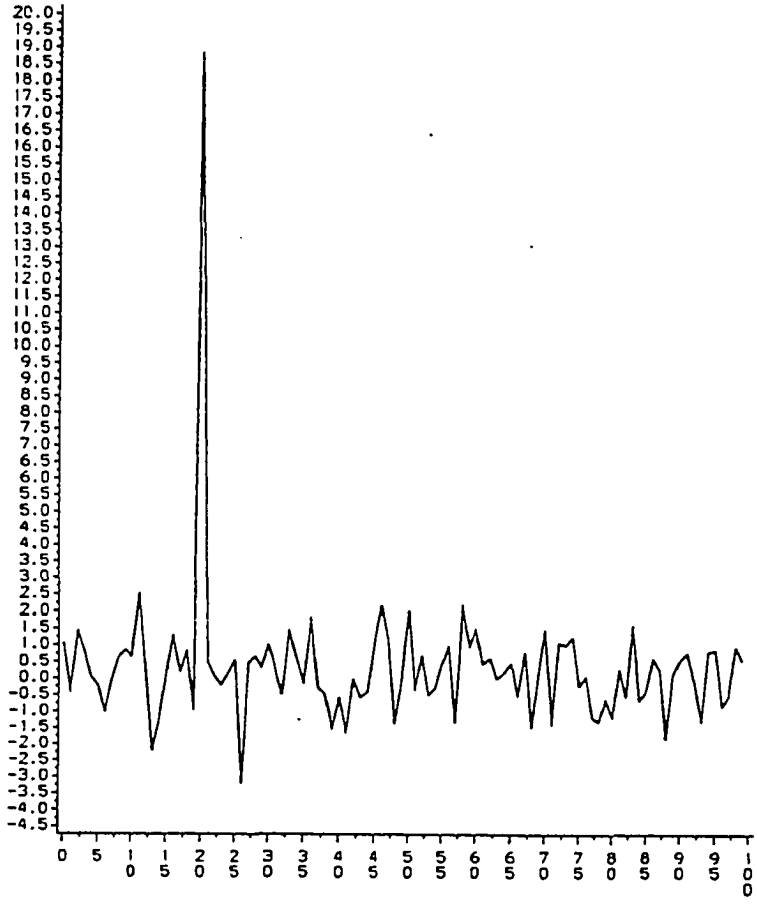

Fig. 6.6: Behavior of the normalized residual.

At step 20, it has a sharp increase.

It is observed that CDR and BDR yield the same performance. This is explained by the fact that the jump is large and the rejection threshold is low. Also, the post-fault noise variance is high which assists in correct detection. It should be noted again that the RWM (in using the CDR) is useful only if the jump is not detected by the hypothesis test. Under the circumstances of this example the one-step detector performed

almost as an ideal detector.

The post-fault estimates converged to the true values within 25 to 30 samples after the fault (12 to 14 miliseconds at 32 samples per cycle), which is comparable to Girghis and Brown's results, and very good for relaying purposes. During the course of this simulation study, some cases of false alarms were encountered. However, as mentioned earlier in chapters III and V, false alarms may disrupt the estimation process temporarily, but the estimates eventually converge to the true values. Fig. 6.7 shows an example of false alarm.

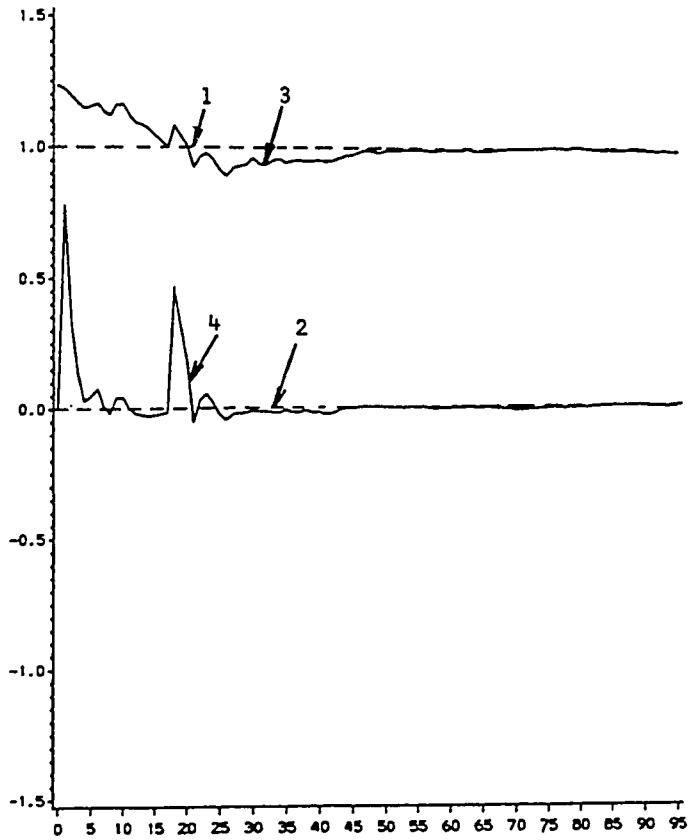

Fig. 6.7: Estimates and true values of the states in the presence of a false alarm.

VI.5.2. Example 2

a) To find out if the detection would be good for smaller jumps, a limiting case of zero jump was simulated, with the same post-fault noise variance as in Example 1), introduced after 60 steps ($k_0 = 60$). This program was written to terminate as soon as a detection occured, be it a false alarm or correct detection. The same thresholds as in Example 1 were used. In 1000 runs correct detection with the minimum possible delay (one step) occured in 697 cases; there was a two-step delay in 162 cases. The maximum detection delay was found to be 5 samples after the jump. In 54 cases the program was terminated by due to a false alarm.

b) For the case of a low time constant of the protected line, the program of part a) was repeated with a lower post- fault noise variance

$$R_k = 0.01\, e^{-0.2(k-k_0)}, \quad k \geq k_0, \quad k_0 = 60,$$

and the following thresholds:

$$T_r = 3.3$$

$$T = 1.0.$$

The rejection threshold corresponds to $\alpha = 0.001$. The simulations showed that the false alarm rate decreased compared to the previous thresholds. However, the detection rate also dropped. In 1000 runs immediate correct detection occured in only 360 cases. In 294 cases the detection was delayed by two steps, and in 103 cases by three steps. In 48 cases there was no detection even after 10 steps.

VI.6. Discussion of the Simulation Results

In the previous chapters it was shown that in the case of a missed jump it is useful to have the RWM filter available. If there is an undetected jump, then the BDR makes the algorithm perform similar to the ordinary Kalman filter, whereas the CDR makes it perform similar to a random walk model or finite memory filter. However, for power system state estimation it is observed that since the fault is accompanied by increased noise, the probability of missing a jump is very low. Thus, in most cases it is adequate to have the binary decision rule. It was observed in numerous simulations that the algorithms with BDR and CDR performed the same way. The novelty of the present algorithm lies in the fact that it is much simpler and faster than using an elaborate detection scheme. Since the problem of power system fault detection is really an ideal situation for using single-step hypothesis tests, it follows that the proposed algorithm has good potential for future use.

Example 2 indicates that detection is faster when the time constant of the protected line is high. In one thousand runs with a limiting case of zero jump, the detection rate is found to be very good. The detection rate decreases for lower time constants. Overall, because of its fast detection and accurate tracking capabilities, and because of the particular nature of the noise statistics during a power system fault, the CFA shows substantial promise for use in power system state estimation problems.

Chapter VII

APPLICATION OF THE COMBINED FILTER TO A CONTINUOUS SYSTEM

Summary: In this chapter a new kind of experiment is described. Motivated by the fact that in real life most systems are continuous, the proposed algorithm is tested on the sampled version of a continuous system. In this dissertation, the CFA has been tested for various discrete systems, the results of which are given in appendices A and B However, it must be admitted that in reality, abrupt changes may occur in a continuous system, and a discrete model of the changes may not be available. The goal in this chapter is to investigate the effectiveness of the Kalman filter in general, and hypothesis testing in particular, in such a situation. The Kalman filter with single-sample hypothesis testing is applied to the sampled input-output sequence of a simulated continuous-time system. The result of this study shows that parameter jumps occuring in the continuous system can be detected by the proposed method. It also indicates areas for further research.

VII.1 Problem Statement

The chosen continuous-time system is the well-known inverted pendulum problem. Figure 7.1 shows the pendulum attached to a nearly vertical rod, the motions of which are described by the equations

$$\left.\begin{aligned} &\ddot{\psi}(t) - \sigma^2 \psi(t) - \omega^2[\psi(t) + \theta(t)] = u(t) \\ &\ddot{\theta}(t) + \omega_0^2 \theta(t) + \epsilon \ddot{\psi}(t) = 0, \end{aligned}\right\} \qquad (7.1)$$

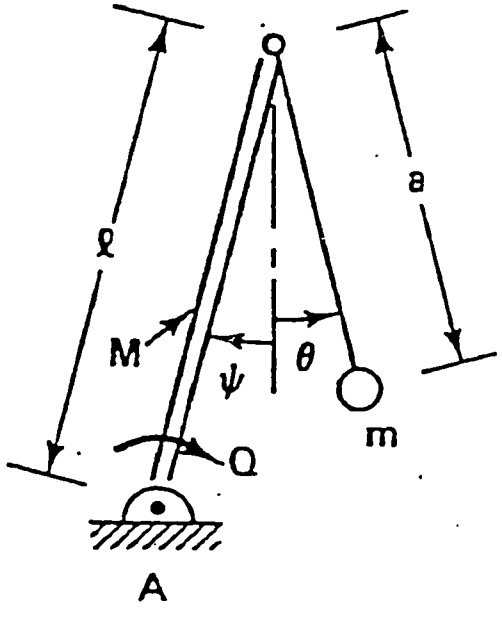

Fig. 7.1: The inverted pendulum.

where

$$\sigma^2 = \frac{3g}{2l}, \quad \omega^2 = \frac{3mg}{Ml}, \quad \epsilon = \frac{l}{a}, \quad \omega_0^2 = \frac{g}{a}, \quad u = \frac{Q}{Ml},$$

Q = torque applied at A,

g = gravitational force per unit mass.

Assuming that a = 1 and $\frac{\omega^2}{\omega_0^2} = \frac{3m}{M} = r$, and after some algebraic manipulations, the following state-space model can be derived:

$$\dot{\xi} = A\xi + bu, \qquad (7.2)$$

with

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1.5+r & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1.5-r & 0 & -2 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix},$$

where

$$\xi_1 = \psi, \quad \xi_2 = \dot{\psi}, \quad \xi_3 = \theta, \quad \xi_4 = \dot{\theta}.$$

The parameter r has been left as a variable in order to introduce changes in the hanging mass m in the continuous system. The nominal value of r is assumed to be 1.

After stabilizing the system with a state feedback, the model is simulated using the Continuous System Modeling Program (CSMP). The system is excited by a random input sequence. The angle ψ (state variable X_1) is considered as the output. The input and output are sampled, and the CFA is used to estimate the parameters of an assumed fourth-order ARMA model, which is given by the following:

$$Z_k = \sum_{i=1}^{4} a_i Z_{k-i} + \sum_{i=1}^{4} b_i w_{k-i}.$$

The constant parameter model is

$$X_{k+1} = X_k,$$

where

$$X = [\, a_1 \cdots a_4 \; b_1 \cdots b_4 \,]^t .$$

In order to investigate whether the test of hypothesis can detect a change in the parameters of the original continuous system, the parameter r in the A matrix of eq. (7.2) was changed from its nominal value of 1. Increase in its value is equivalent to the addition of extra mass, whereas decrease in its value means loss of some hanging mass. The goal is to find out whether the CFA can detect the change, and estimate the parameters.

VII.2 Results of the Simulation Study

The system of eq.(7.2) was simulated for 120 seconds. The input and output functions were sampled at the rate of 20/sec. At step 1000 (50 sec.), the value of r was changed from 1 to 4.5. Estimates of all the eight parameters were examined for signs of jumps. When an ordinary Kalman filter was used for the estimation, none of the parameters showed any jump. In fig. 7.2, the estimates of the first four parameters are shown when the jump exists but an ordinary Kalman filter is used.

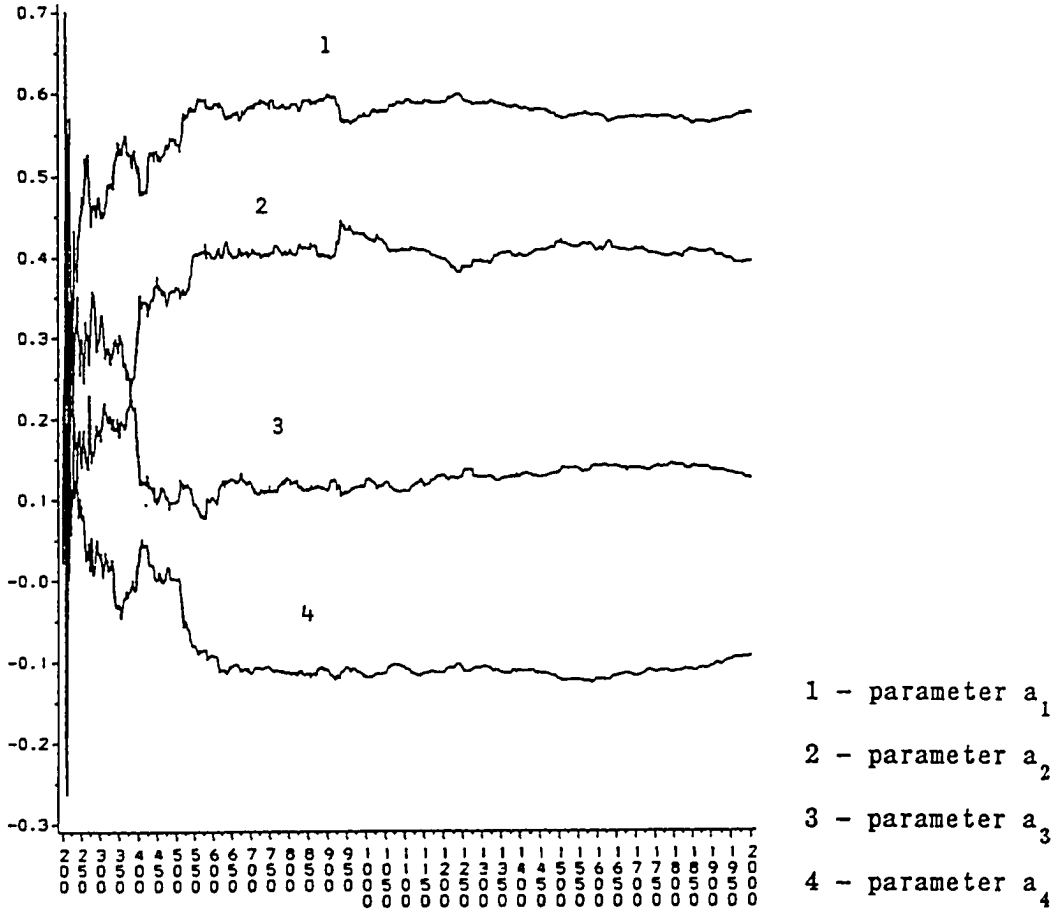

Fig. 7.2: Estimates of parameters a_1 through a_4, given by the ordinary Kalman filter. At step 1000 (50 sec.), parameter r in the continuous system changes from 1 to 4.5.

When the CFA was used to estimate the eight parameters, one parameter, a_4, was found to show a substantial jump. All the other parameter estimates showed the same average values before and after the jump. Fig. 7.3 shows the estimate of the parameter a_4. The pre-jump value is approximately -0.13, whereas the post-jump value is about -0.06. Other parameters, not showing any significant jumps, are not plotted.

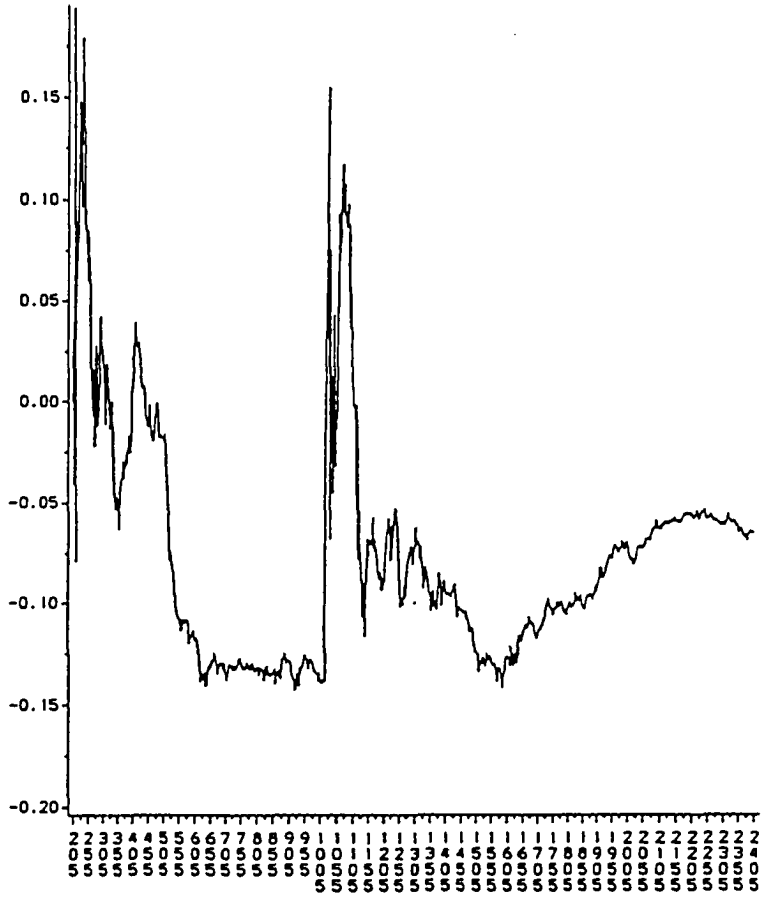

Fig. 7.3: Estimate of parameter a_3, given by the CFA. At step 1000 (50 sec.), parameter r in the continuous system changes from 1 to 4.5.

Two major problems are encountered during this study. These may be potential difficulties in a real life situation also. The next two sections are devoted to the discussion and possible remedy of these problems.

VII.3 Sensitivity Problem

The most serious problem encountered is that of the sensitivity of the parameters of the discrete model to the changes in the continuous system. Since an exact discrete model is assumed to be unknown, it is also unknown how sensitive its parameters will be to the changes in the original system. In some cases they may have very low sensitivity to the changes in a particular parameter of the continuous system.

In the present simulation study each of the four state variables, one by one, was considered as the output of the continuous system. It was found that X_1 and X_3 are more sensitive than the rest. It was also found that the parameters of the discrete model are insensitive to any decrease in the hanging mass. Lack of sensitivity is attributed to the state feedback used to stabilize the pendulum.

This problem of low observability may result in inadequate detection not only by the CFA, but by any other technique as well. It is important to remember that the algorithm is designed to detect jumps that take place in the parameters of the discrete system. It is assumed that any such jump produces sufficient deflection in the residual of the filter. If, for any reason, the changes in the parameters are such that they do not cause this deflection, then no test of hypothesis will detect the change. Making the abrupt change in the continuous system,

while imitating a real life situation, does not guarantee an adequate change in the discrete system parameters.

Another problem is that of false alarms. Although related to the problem of sensitivity, it deserves separate treatment. The next section is devoted to this problem.

VII.4 The Problem of False Alarms

A real system may or may not be susceptible to a large amount of false alarms. In the case of the inverted pendulum, the system takes a long time to reach steady state. Therefore, the Kalman filter converges very slowly. Operating the hypothesis test in that situation results in a larger than designed probability of false alarms. This fact is demonstrated by the simulation study. However, it is reasonable to assume that, in a practical situation, the conventional Kalman filter can be operated long enough to get "good" initial estimates before the CFA is started. Then the filter is optimum, and the probabilities associated with the hypothesis tests are valid.

Another source of false alarms lies in the fact that just after a jump, there may be strong transient behavior in the system. The detection mechanism does not take that into account, thus resulting in a large probability of false alarms. Again, the hypothesis test may be switched off for some time just after a detection. It can even be switched off altogether, if the assumption of "at most one jump" is valid. In the present simulation the hypothesis test was turned off for long periods of time in the beginning of the run, and also just after the jump.

There is one other reason for false alarms, namely, a sudden very bad measurement. Some systems may be more prone to this kind of behavior than others. An effective way to deal with this problem is to require two or more consecutive rejections of the null hypothesis before sounding an alarm. This has been mentioned earlier in chapter III. This is a trade-off between the number of false alarms and speed of detection. With a high rejection threshold, one decreases the probability of false detections, but also reduces the probability of detecting an actual jump. However, due to the CDR (combined decision rule), an 'undetected' jump may still be traced by the RWM filter.

The choice of thresholds also depends upon the sensitivity of the system. For example, the inverted pendulum system turned out to have very low sensitivity to the change in the ratio of masses r. This meant that the jump in the value of r in the continuous system did not produce a large deviation in the normalized residual of the Kalman filter. Therefore a low rejection threshold was necessary for detecting the jump, which resulted in a large probability of false alarms.

Besides the aforementioned two major problems, one has to be careful about choosing the proper sampling rate. Inadequate sampling may result in loss of information regarding the continuous system. This well known fact is not unique to the present situation. Therefore, a detailed discussion is not necessary. It is enough to mention that improper or inadequate sampling may result in biased estimates or detection delays.

Despite problems, the result of this study is promising. It is felt that more research is needed in this particular area of implementing discrete filters for continuous systems.

Chapter VIII

CONCLUSIONS

The CFA has been designed to detect large jumps with the help of single-sample hypothesis tests, and trace small jumps with the help of the RWM filter. It degenerates into an ordinary Kalman filter if the parameters remain constant for a segment of time. The algorithm is simple and efficient. Its superiority over other available methods has been established in a number of simulation studies.

The primary development in this dissertation is for the single output case. However, the multi-output case has also been studied, and the method has been extended to it. The approach proposed here results in a Chi-square distributed test statistic, which implies that a matrix square-root extraction and inversion is avoided at each step. This is a significant computational simplification. However, it is felt that the multi-output case needs to be studied in more detail for the case of jumps in the noise statistics. The test of hypothesis is based on the Chi-square distribution, thus it is one-tailed. The asymmetry of the Chi-square distribution has to be taken into consideration while modifying some properties and their proofs for the multi-output case.

The main application of the CFA in this work is that of power system fault detection and state estimation. Detailed simulation studies have been carried out. They show that single sample hypthesis test is ideal for detecting short circuit faults. The efficiency of the Kalman filter in estimating the steady state voltages and currents has been

shown earlier by other researchers. More work is needed in the area of applicability of hypothesis tests in order to distinguish between different types of power system faults. If that can be achieved, then good initial post fault estimates will be available which will make the estimates converge much faster than is possible at present. Moreover, accurate modeling of the power system during and immediately after different kinds of faults is very important. At the present time the same model is assumed for all kinds of faults. It is felt that an open ground for more research exists in this area.

Another area that needs to be investigated in great detail is that of implementation of Kalman-type discrete filters to continuous systems. Jumps in the continuous system may produce several changes in its discrete model, which have to be identified quickly. In this dissertation one such example has been studied, and computer simulations have been performed. It is found that sensitivity problems play a great role in the success or failure of the method. An area of potential difficulty is that of diminished sensitivity due to state feedbacks. However, state feedbacks are sometimes necessary for the stabilization of the model. Methods of avoiding high incidents of false alarms are also important. Apriori test runs of the conventional Kalman filter may provide required initial estimates. It is contended that the performance of the CFA for the continuous system, while promising, has room for improvement. The improvement has to come from more detailed studies of the sensitivity problem, and also from better understanding of the transition from continuous to discrete models. It is important to remember that in real life most systems are continuous, and one has to

rely on sampled versions of those systems for applying various digital algorithms. Thus the problems of efficient sampling need also to be studied.

In conclusion it is reasonable to state that the CFA shows great promise for application to discrete systems. With future research for implementation on continuous systems, the proposed method or its modifications may be beneficial in solving some important identification problems.

BIBLIOGRAPHY

[1] Sage, A. P. and J. L. Melsa, *Estimation Theory with Application to Communications and Control*, Academic Press, 1971.

[2] Sage, A. P. and C. C. White III., *Optimal Systems Control*, Prentice-Hall, 1977.

[3] Perriot-Mathonna, D. M. "Improvements in the application of stochastic estimation algorithms- parameter jump detection". *IEEE Trans. Auto. Control*, Vol.AC-29, pp. 962-969, Nov. 1984.

[4] Basseville, M. "Detection and diagnosis of changes in the eigenstructure of nonstationary multivariable systems". *Automatica*, Vol.23, No.4, pp. 479-489,1987.

[5] Basseville, M. "The two models approach for the on-line detection of changes in AR process". In *Detection of abrupt changes in signals and dynamical systems*, LNCIS No.77, Springer-Verlag, 1986.

[6] Basseville, M and A. Benveniste, "Sequential detection of abrupt changes in spectral characteristics of digital signals". *IEEE Trans. Information Theory*, Vol.20, No.5, pp. 709-723, Sep. 1983.

[7] Nikiforov, I. V. "Sequential detection of changes in stochastic systems". In *Detection of abrupt changes in signals and dynamical systems*, LNCIS No.77,Springer-Verlag, 1986.

[8] Willsky, A. S. "Detection of abrupt changes in dynamic systems". In *Detection of abrupt changes in signals and dynamical systems*, LNCIS No.77,Springer-Verlag, 1986.

[9] Alexander, S. T. *Adaptive Signal Processing Theory and applications*. Springer-Verlag, 1986.

[10] Mohanty, N. *Random Signals Estimation and Identification — Analysis and Applications*. VNR, 1986.

[11] Brown, R. G. *Introduction to Random Signal Processing and Kalman Filtering*. John Wiley and sons, 1983.

[12] Anderson, B. D. O. and J. B. Moore, *Optimal Filtering*. Prentice Hall, 1979.

[13] Young, P. *Recursive Estimation and Time Series Analysis*. Springer-Verlag, 1984.

[14] Aravena, J. L. "Finite memory estimation: moving window Kalman filter". 2^{nd} *symposium of ACCA, Santiago, Chile,* 1984.

[15] Bohlin, T. "Four cases of identification of changing systems". In *Mathematics in science and engineering,* Vol. 126, pp. 441-518, 1976.

[16] Basseville, M. and A. Benveniste, "Design and comparative study of some sequential jump detection algorithms for digital signals". *IEEE Trans. ASSP,* June 1983.

[17] Basseville, M. and A. Benveniste (Editors), *Detection of Abrupt Changes in Signals and Dynamical Systems*. In LNCIS No. 77, Springer- Verlag, 1986.

[18] Caulcott, E. *Significance Tests*. Unwin Brothers Ltd., Great Britain, 1973.

[19] Aplevich, J. D. and Y. L. Shi, "Simultaneous state and parameter estimation of linear multivariable systems with constrained parameters". *Proc.* 2^{nd} *IASTED International Conf., Applied Control and Identification,* Los Angeles, Dec. 1986.

[20] Cioffi, J. M. and T. Kailath, "Windowed fast transversal filters: adaptive algorithms with normalization". *IEEE Trans. ASSP*, Vol. ASSP-33, No.3, pp. 607-624, June 1985.

[21] Cioffi, J. M. and T. Kailath, "Fast, recursive least-squares transversal filters". *IEEE Trans. ASSP*, April 1984.

[22] Chen, H. F. "Recursive system identification and adaptive control by use of modified least squares algorithm". *Trans. Soc. for Industrial and Applied Mathematics*, Sep. 1984.

[23] Chow, "Tests of equality between two sets of coefficients in two linear regressions". *Econometrica*, Vol. 28, pp. 561-605, 1970.

[24] Wood, A. J. and B. F. Wollenberg, *Power Generation and and Control*. Power Technologies inc., 1984.

[25] Sachdev, M. S., Wood and Johnson, "Kalman filtering applied to power system measurements for relaying". *IEEE Trans. PAS*, Dec. 1985.

[26] Girgis, A and R.G. Brown, "Adaptive Kalman filtering in computer relaying". *IEEE Trans. PAS*, May 1985.

[27] Girgis, A. "A new Kalman filtering based digital distance relay". *IEEE Trans. PAS*, Sep. 1982.

[28] Girgis, A and R. G. Brown, "Application of Kalman filtering in computer relaying". *IEEE Trans. PAS*, July 1981.

[29] Sorenson, H. W. *Parameter Estimation: Principles and Problems*. Marcel Dekker, New York,1980.

[30] Mili, L., T. Van Cutsem and M. Ribbens-Pavella, "Decision theory applied to bad data identification in power system state

estimation". *Proc. IFAC Symp. Identification and System Parameter Estimation*, pp. 945-950. Univ. of York, U.K., July 1985.

[31] Watanabe, K., T. Yoshimura and T. Soeda, "A diagnosis method for linear stochastic systems with parametric failures". *ASME Trans.* March 1981.

[32] Shiryev, A. N. "On optimum methods in quickest detection problems". *Theory of Prob. and its Applic.* Vol. 8, No. 1, pp. 22-46, 1963.

[33] Lorden, G. "Procedures for reacting to a change in distribution". *Annals of Mathematical Statistics*, Vol. 42, No. 6, pp. 1897-1908, 1971.

[34] Soeda, T. and T. Yoshimura, "A practical filter for systems with unknown parameters", *Trans. ASME, Journal of Dynamic Systems, Measurement, and Control*, pp. 396-401, Dec. 1973.

[35] Lehmann, E. L. *Testing Statistical Hypotheses*. N. Y. Wiley, 1959.

[36] Benveniste, A. "Advanced methods of change detection: an overview". In *Detection of Abrupt Changes in Signals and Dynamical Systems*, LNICS No. 77, pp. 77-96, 1986.

[37] Mann, B. J. and I.F. Morrison, "Digital calculation of impedence for transimission line protection". *IEEE Trans. PAS*, Vol. 90, No. 1, pp. 270-279, Jan/Feb. 1971.

[38] Gilchrist, G. B., G.D. Rockefeller and E.A. Urden, "High speed distance relaying using a digital computer: part I - system description", *IEEE Trans. PAS*, Vol. 91, No. 3, pp. 1235-1243, May/June, 1972.

[39] Makino, J. and Y. Miki, "Study of operating principles and digital filters for protective relays with digital computers", *IEEE Publication No. 75CH0990-2 PWR, Paper No. C-75 197-9, IEEE PES Winter Meeting*, New York, pp. 1-8, Jan. 1975.

[40] Ramamoorty, M. "Application of digital computers to power system protection", *Journal of Inst. Engr. (India)*, Vol. 52, No. 10, pp. 235-238, June 1972.

[41] Phadke, A. G., T. Hiibka and M. Ibrahim, "Fundamental basis for distance relaying with symmetrical components", *IEEE Trans. PAS*, pp. 635- 646, March/April 1976.

[42] Throp, J. S., A. G. Phadke, S. H. Horowitz and J. E. Beehler, "Limits to impedance relaying", *IEEE Trans. PAS*, Vol. PAS-98, No. 1, pp. 246-260, Jan./Feb. 1979.

[43] Luckett, R. G., P. J. Munday and B. E. Murray, "A substation based computer for control and protection", *Developments in power system protection, IEEE Conference Publication* No. 125, pp. 291-296, London, March 1975.

[44] Sachdev, M. S. and Baribeau, "A new algorithm for digital impedance relays", *IEEE Trans. PAS*, Vol. PAS-98, No.6, pp. 2232-2240, Nov./Dec. 1979.

[45] Brooks, A. V. Jr. "Distance relaying using least-squares estimates of voltage, current and impedance", *Proceedings of IEEE PICA Conference* No. 77CH 1131-2-PWR, pp. 394-402, May 1977.

[46] Namba, M., T. Nishiwaki, S. Yokokawa, and K. Ohtsuka, "Identification of parameters for power system stability

analysis using Kalman filter", *IEEE Trans. PAS*, Vol. PAS-100, No. 7, pp. 3304-3311, July, 1981.

[47] Swift, G. W. "The spectra of fault induced transients", No. F-78 717-1, pp. 1-7, *IEEE PES Summer Meeting*, Los Angeles, July 1978.

[48] Mili, L., T. Van Custem and M. Ribbens-Pavella, "Decision theory for fault diagnosis in electric power systems". *Automatica*, Vol.23, No.3, pp. 335-353, 1987.

[49] Chowdhury, F. N. and J. L. Aravena, "Fast detection of time-varying parameters". *Proc. 2nd IASTED conf.*, Los Angeles, Dec. 1986.

Appendix A

RESULTS OF COMPUTER SIMULATION STUDIES: CONSTANT PARAMETER MODEL, RANDOM WALK MODEL AND MOVING WINDOW KALMAN FILTERS

In this appendix the results of estimating the parameters of a first order ARMA model using an ordinary, constant parameter model (CPM) Kalman filter, a random walk model (RWM) Kalman filter and a moving window Kalman filter (MWKF) are presented. The actual parameters are subjected to jumps and linear variations. It was found that the performance of the moving window Kalman filter and that of the random walk model Kalman filter may be very similar depending on the choice of the design parameters. It was also found that all of these methods are inadequate or slow in case the jumps are substantial.

A.1 Parameters with Jumps

A first order ARMA model is simulated in order to study the behavior of the CPM, RWM, and MWKF. The system has two unknown parameters which are subject to a sudden jump. The system model is

$$Z_k = aZ_{k-1} + bU_{k-1} + V_k.$$

Hence the observation matrix is

$$H_k^t = [Z_{k-1} \quad U_{k-1}],$$

and the state vector, consisting of the two parameters, is given by

$$X_k^t = [a \quad b].$$

The observation equation is

$$Z_k = H_k^t X_k + V_k.$$

The actual values of the parameters are

$$a = \begin{cases} 0.7 \ , \ k < 40 \\ 0.5 \ , \ k \geq 40, \end{cases}$$

$$b = \begin{cases} 2.0 \ , \ k < 40 \\ 2.6 \ , \ k \geq 40. \end{cases}$$

The estimates obtained by the ordinary Kalman filter are shown in Figure A.1.1. Here, the parameter model is $X_{k+1} = X_k$, i.e., the parameters are assumed to be constants in the model. However, they undergo jumps at step 40. The conventional Kalman filter is seen to be quite inadequate in estimating the new values of the parameters.

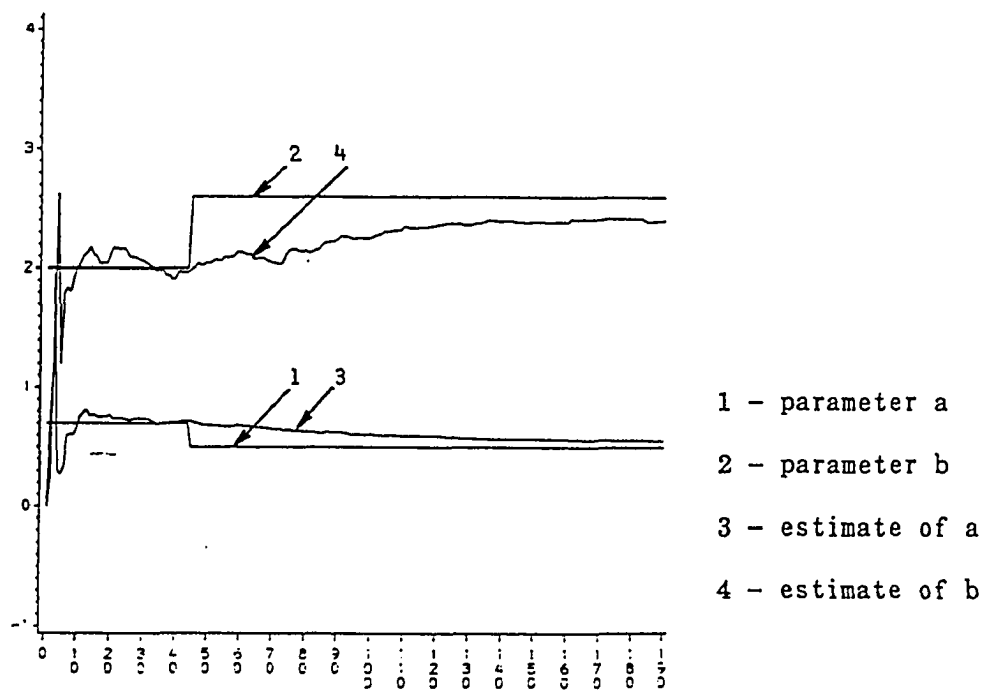

Fig. A.1.1: Behavior of the ordinary Kalman filter when parameters are subjected to jumps.

For the same kind of jumps in parameters, the random walk model filter is found to perform much better. The parameter model in this case is

$$X_{k+1} = X_k + W_k \ ,$$

with

$$E(W_k W_k^t) = S_k.$$

Figure A.1.2 shows the true values of the parameters and the estimates.

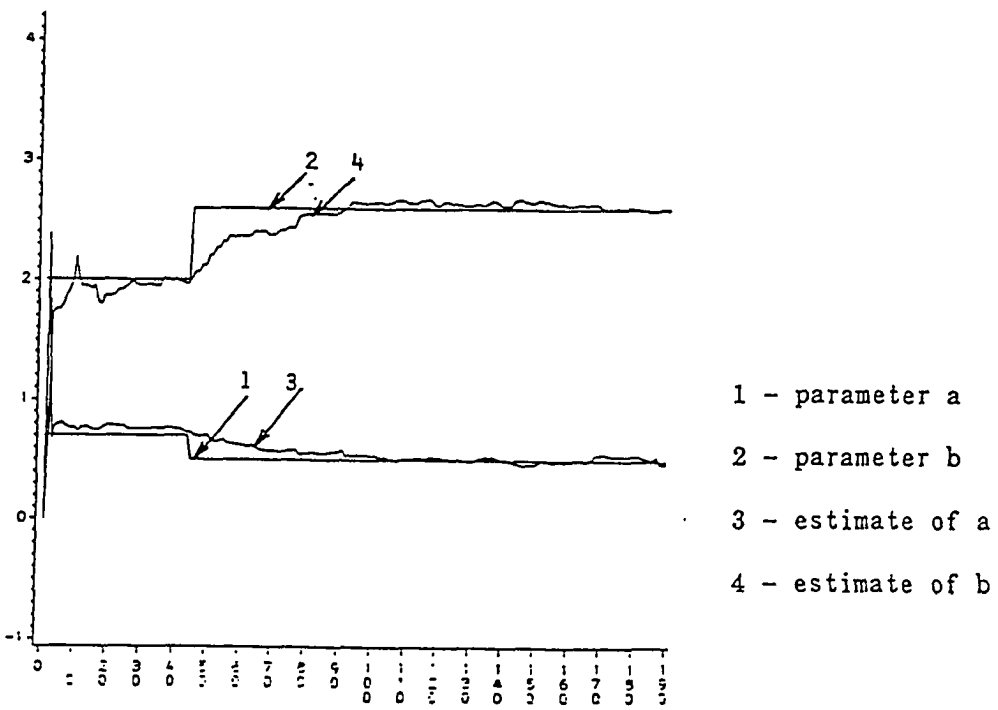

Fig. A.1.2: Performance of the RWM filter when parameters are subjected to jumps.

It must be mentioned here that the estimates of the RWM filter may not always be so good. In this particular run, the combined effects of a low measurement noise variance, the small size of the jumps, and the choice of the S_k matrix resulted in very good performance. To illustrate the point, a different set of values for noise variance, jump and S_k matrix were used, which produced the estimates shown in fig. A.1.3.

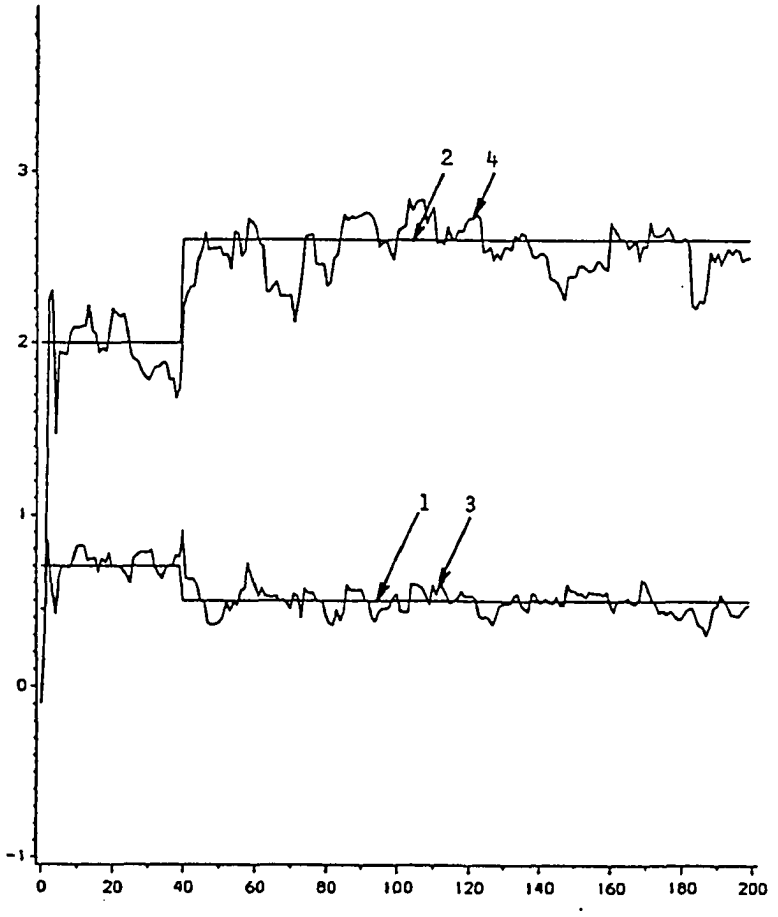

Fig. A.1.3: Another example of estimates from a RWM filter.

1 - parameter a, 2 - parameter b,

3 - estimate of a, 4 - estimate of b.

The next plot, fig. A.1.4, shows the result of using the moving window Kalman filter (MWKF). In this run the 'window' was equal to 30 steps, which implied that at each step, the effect of the most recent 30 data points was retained. Older data was forgotten. The estimates can be adjusted by changing the window length. The larger the window, the closer is the behavior to the ordinary Kalman filter (also to the RWM filter with a very small S_k matrix). Smaller window length makes the filter behave like the RWM filter with large values of S_k matrix. The similarity between the MWKF and the RWM filter is even more evident in a set of simulations described later in this section.

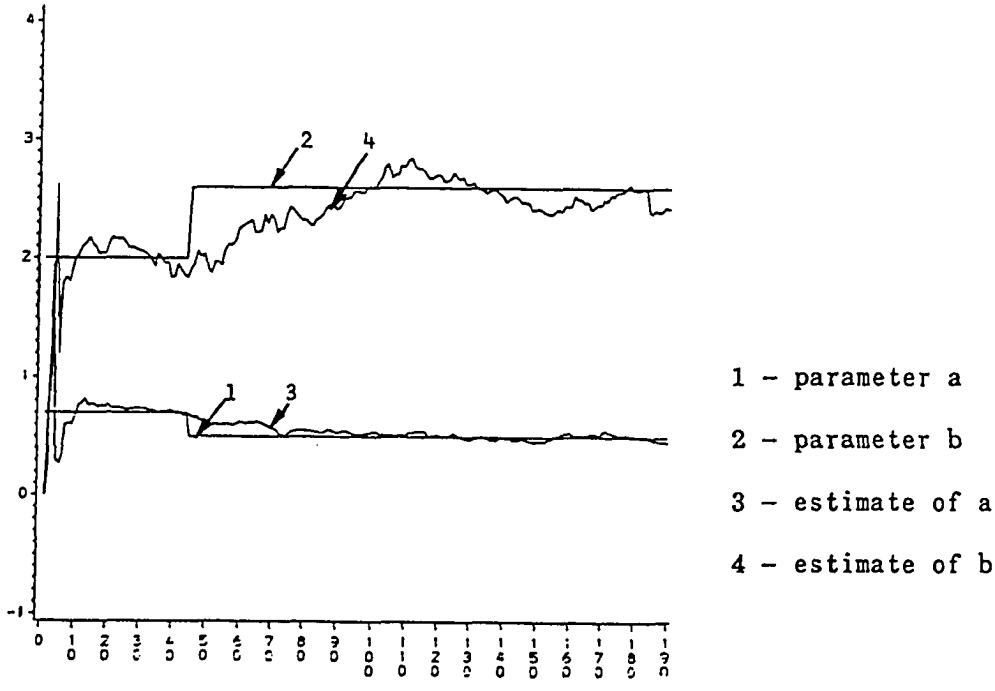

Fig. A.1.4: Performance of the moving window Kalman filter.

A.2. Linearly Varying Parameters

In this group of examples the parameters were varied linearly. Fig. A.1.5, A.1.6 and A.1.7 show the performances of the CPM, RWM and MWKF respectively. With a reduced measurement noise variance and reduced window length (20 steps), the RWM and MWKF were found to produce almost same estimates. Fig. A.1.8 and A.1.9 show these results.

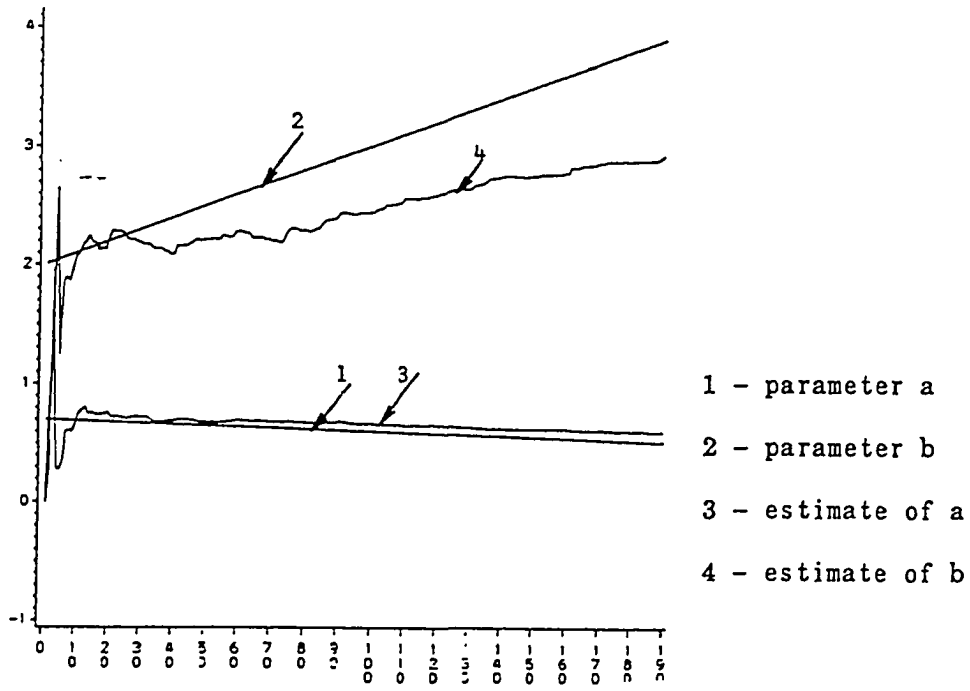

Fig. A.1.5: Estimates produced by the CPM Kalman filter when parameters are linearly varying.

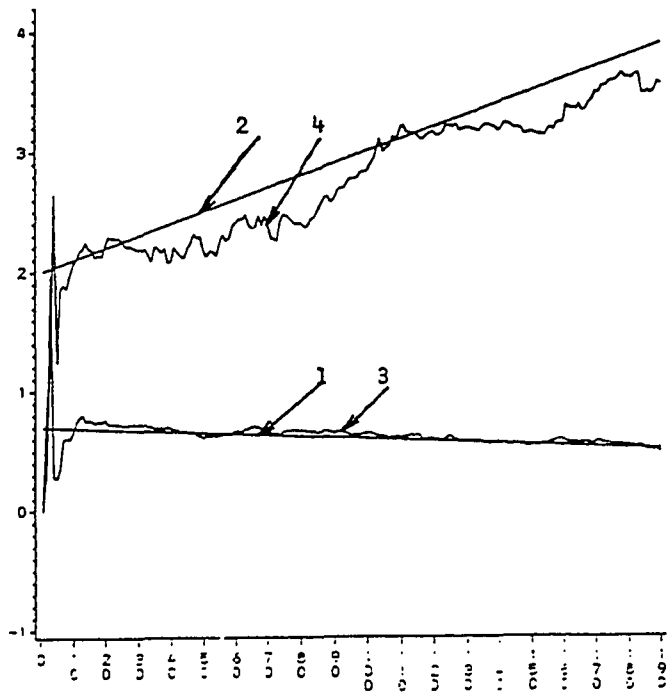

Fig. A.1.6: Estimates of linearly varying parameters produced by the RWM filter.

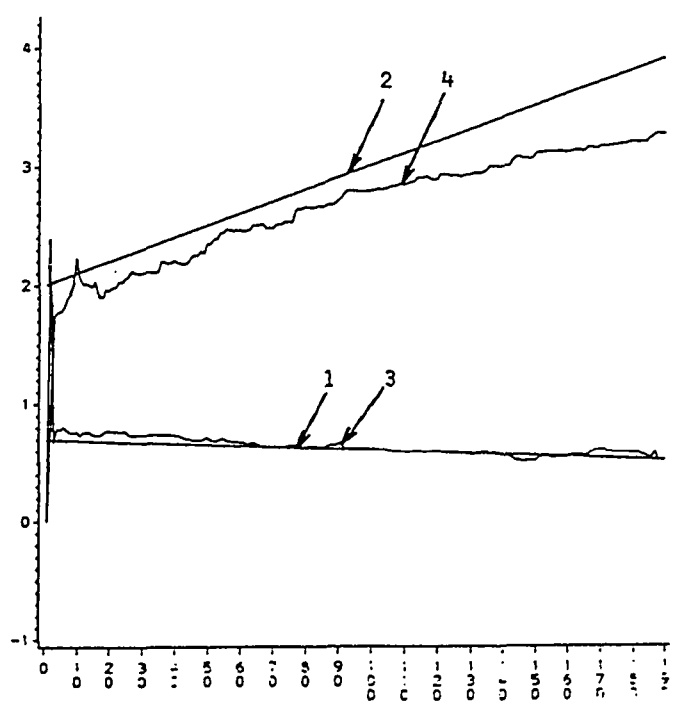

Fig. A.1.7: Estimates of linearly varying parameters produced by the MWKF.
Window length = 30.
1 - parameter a
2 - parameter b
3 - estimate of a
4 - estimate of b

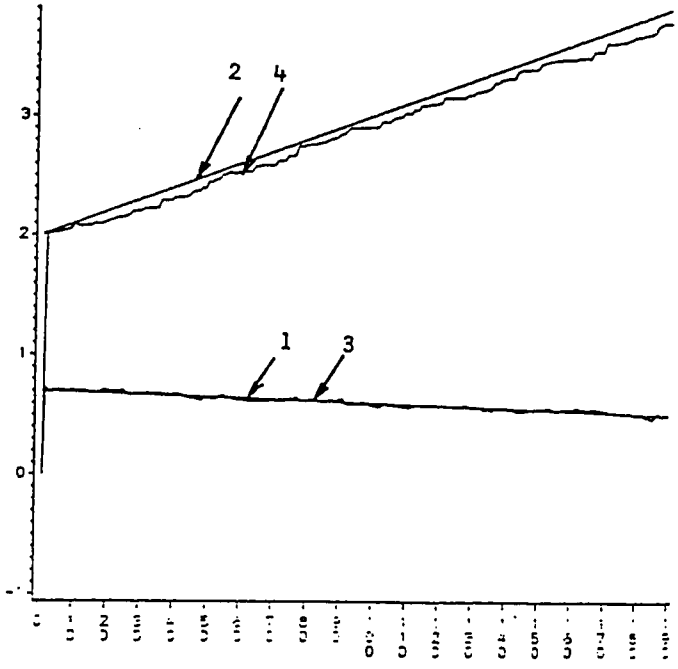

Fig. A.1.8: Estimates produced by the MWKF. Window length = 20.

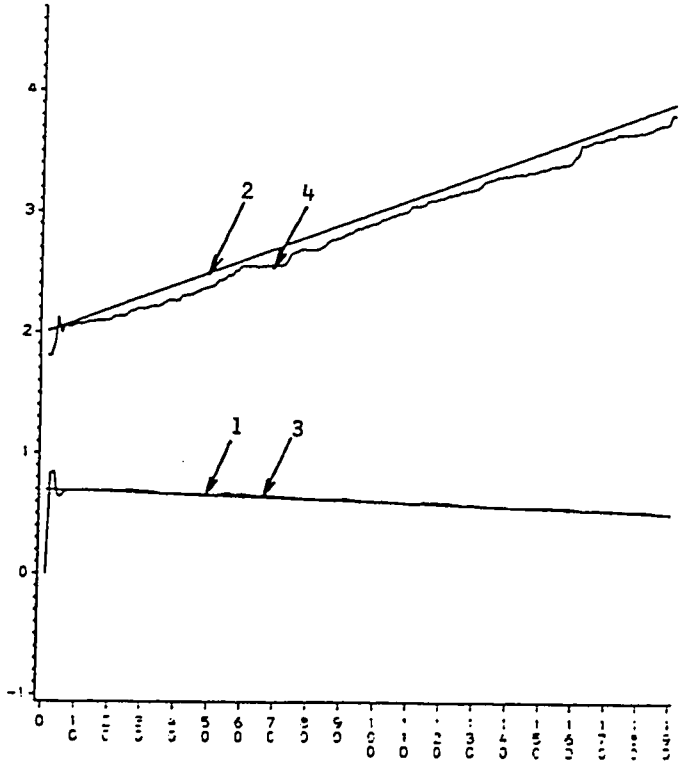

Fig. A.1.9: Estimates produced by the RWM filter.

1 - parameter a

2 - parameter b

3 - estimate of a

4 - estimate of b

This comparative simulation study shows that the conventional Kalman filter is ill-equipped to handle the case of parameter jumps. The random walk model (RWM) and moving window Kalman filters (MWKF) perform similarly. However, the MWKF is computationally more complex than the RWM filter. For small jumps, both of these filters may be adequate. They are also able to trace large jumps, but then the quality of the estimates is lower. In the present dissertation, tests of hypothesis are used to detect large jumps, and RWM filter is used to trace smaller jumps. The algorithm is called the "Combined Filter Algorithm (CFA)".

In appendix B the results of using the combined filter algorithm on first and higher order ARMA models are presented.

Appendix B

RESULTS OF COMPUTER SIMULATION STUDIES: COMBINED FILTER ALGORITHM

This appendix contains the results of testing the Combined Filter Algorithm (CFA) on various ARMA models. The CFA was applied to estimate the parameters of first and fourth order ARMA models, and the resulting estimates were plotted for different amounts of jumps and threshold values. It was found that the CFA is very efficient in detecting the jumps. After the re-initialization, the new values of the parameters were estimated accurately. The behavior of the normalized residuals was plotted for the fourth order model and its magnitude was found to have a sharp increase shortly after the parameter jump. The effect of high and low threshold-values, and that of large and small S_k matrix were also studied.

B.1 First Order ARMA Examples

The first group of these examples has two unknown parameters, one of which is subjected to a sudden jump. The system model is the same as the one described in appendix A:

$$Z_k = aZ_{k-1} + bU_{k-1} + V_k.$$

The observation matrix is

$$H_k^t = [Z_{k-1} \quad U_{k-1}],$$

and the state vector, consisting of the two parameters, is given by

$$X_k^t = [a \quad b].$$

The observation equation is

$$Z_k = H_k^t X_k + V_k.$$

In the first example, the threshold values are:

$T_r = 2.58$, which corresponds to $\alpha = 0.01$,

and $T_a = 1.0$.

Actual parameter values are:

$$a = \begin{cases} 0.7 \ , & k < 45 \\ 0.0 \ , & k \geq 45 \end{cases}$$

and $b = 2.0$ for all k.

The initial estimate is: $[\hat{a} \quad \hat{b}]^t = [0.0 \quad 0.0]^t$.

Neither the time nor the magnitude of the jump are assumed to be known. The Combined Filter Algorithm is found to detect the jump almost immediately. After the re-initialization, the new values of the parameters are estimated with reasonable accuracy within a short time. Figure B.1.1 shows the estimates and the true values. It is observed that the estimate of the parameter which did not jump undergoes some disturbance at the moment of detection, but it settles down to the original value shortly after the re-initialization.

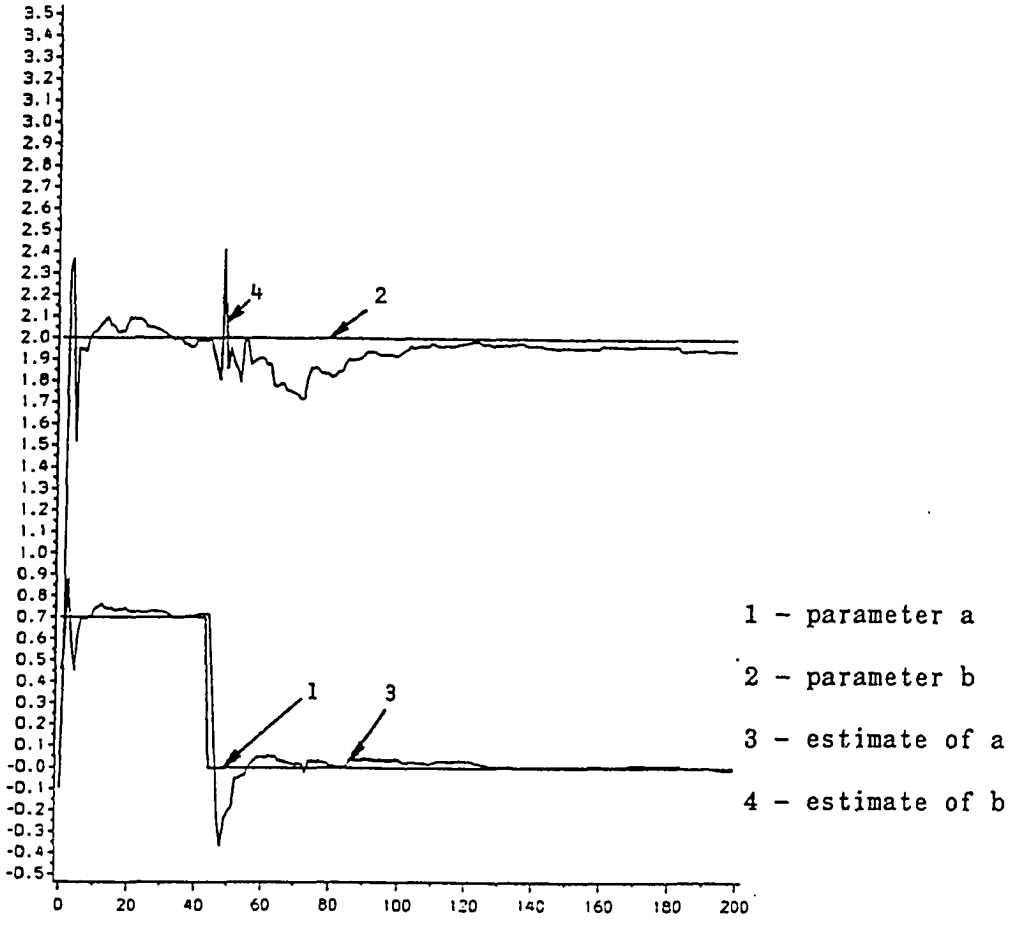

Fig. B.1.1: Estimates and true values when one parameter is subjected to jump.

If the rejection threshold is increased, then the false alarm probability decreases. This automatically increases the probability of missed jump. In such a situation the benefit of using the CDR is seen clearly. If the BDR is used, then the algorithm operates like the ordinary Kalman filter and is inefficient in estimating the post-jump parameter. Fig. B.1.2 illustrates this situation.

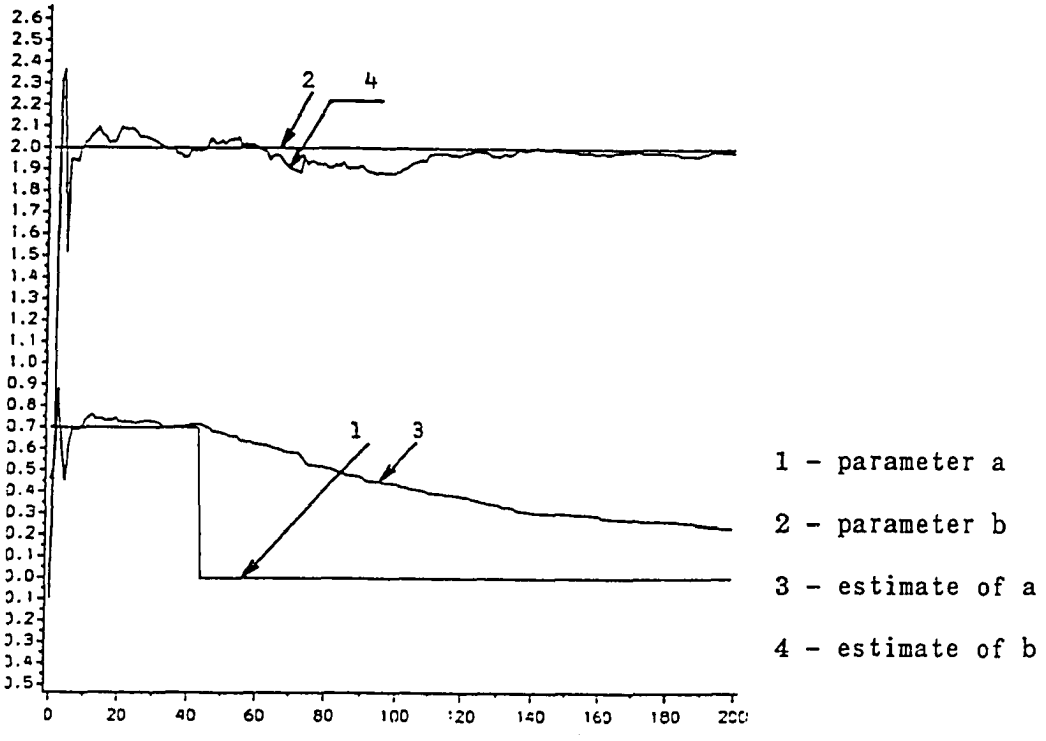

Fig. B.1.2: Estimates and true values when the jump is not detected. Due to BDR, the algorithm behaves like a CPM Kalman filter.

On the other hand, the CDR makes the algorithm behave as a random walk model filter, which results in much better estimates. The same effect is obtained by lowering the magnitude of the jump, while keeping the rejection threshold unchanged. A smaller jump goes undetected by the test of hypothesis; but the normalized residual, although not large enough to cause failure of the hypothesis test, is large enough to cause

the RWM filter to be started. Figure B.1.3 shows the estimates when the jump is smaller than before, and the CDR is used. The a parameter changes from 0.7 to 0.2 at step 45. The b parameter remains the same as before. If the BDR were used, the result would have been similar to that shown in fig. B.1.2.

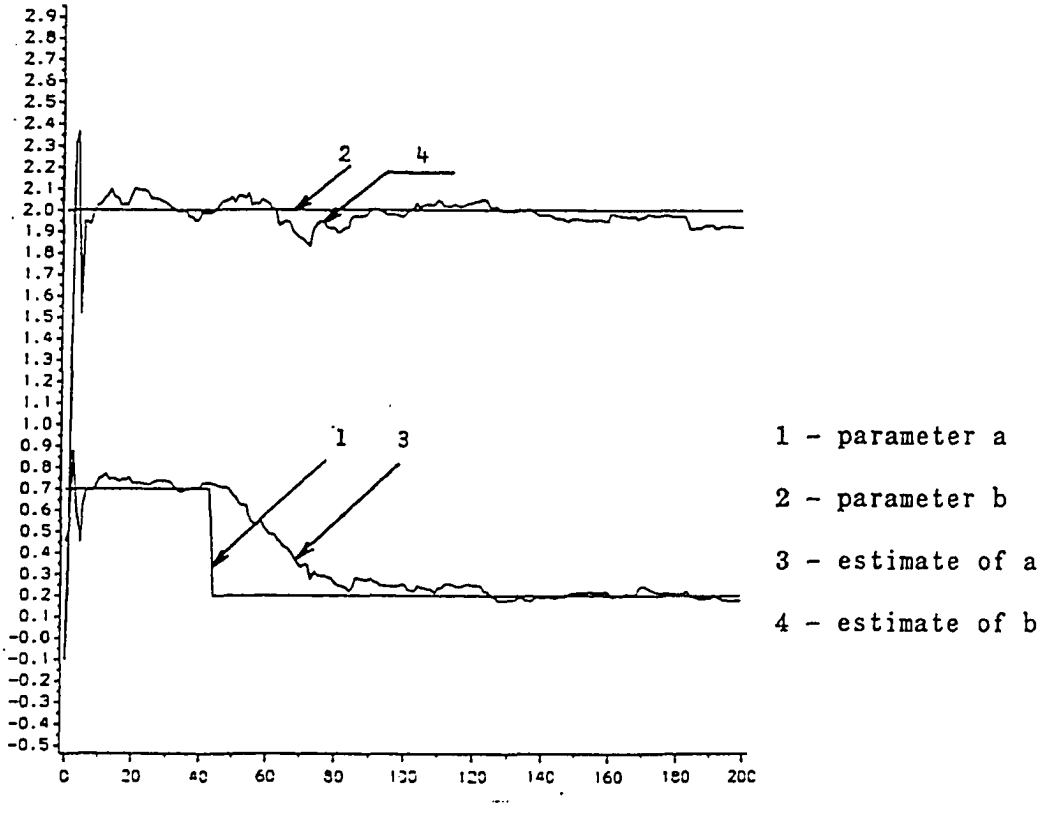

Fig. B.1.3: Estimates and true values of the parameters when the jump is too small to be detected, but too large to be ignored, and the RWM filter is triggered.

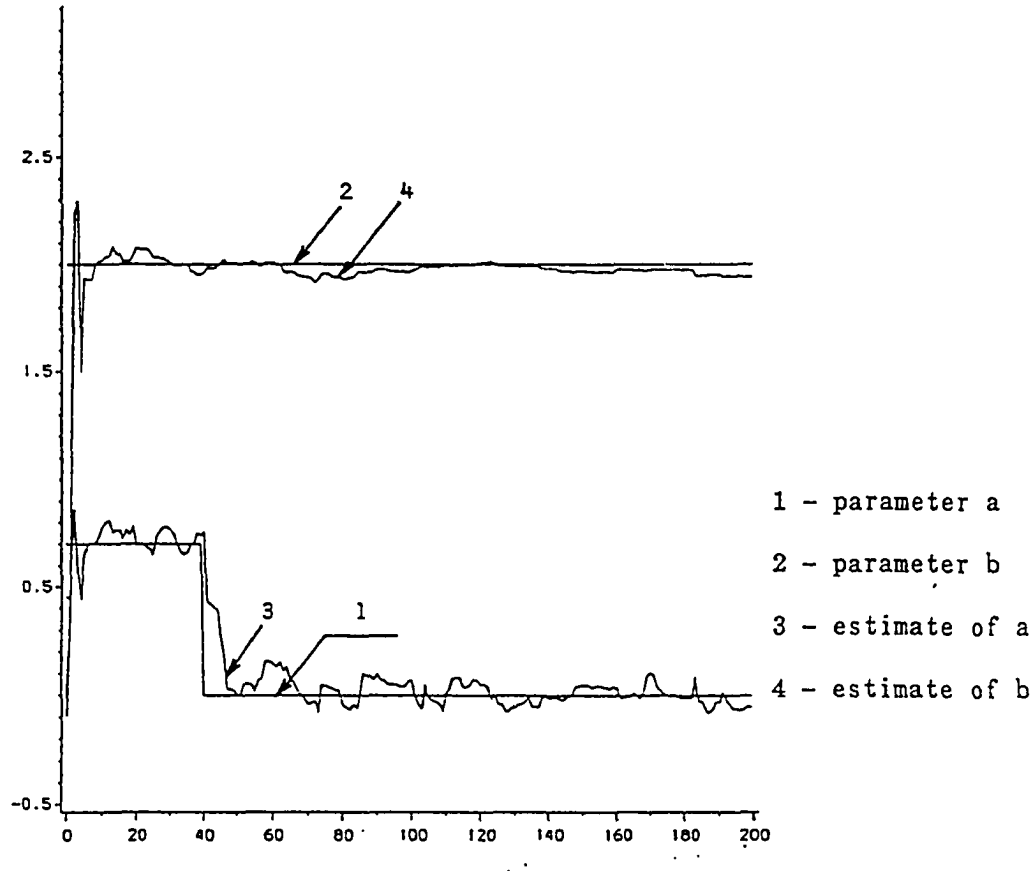

Fig. B.1.4: Estimates and true values when T_r is so high that the hypothesis test does not fail. RWM filter is activated.

This simulation study shows that for piecewise constant parameters, the CPM filter with hypothesis test yields very good performance. When the jumps are substantial, the BDR and CDR give essentially the same results. However, if the jumps are smaller, or if the rejection threshold has to be high, then the CDR is the better alternative.

In the second group of these examples both parameters are subjected to jump. In the first example of this group, parameter a changes from 0.7 to 0.0 and parameter b changes from 2.0 to 2.5 at step 40. Fig. B.1.5 shows the estimates and the true values.

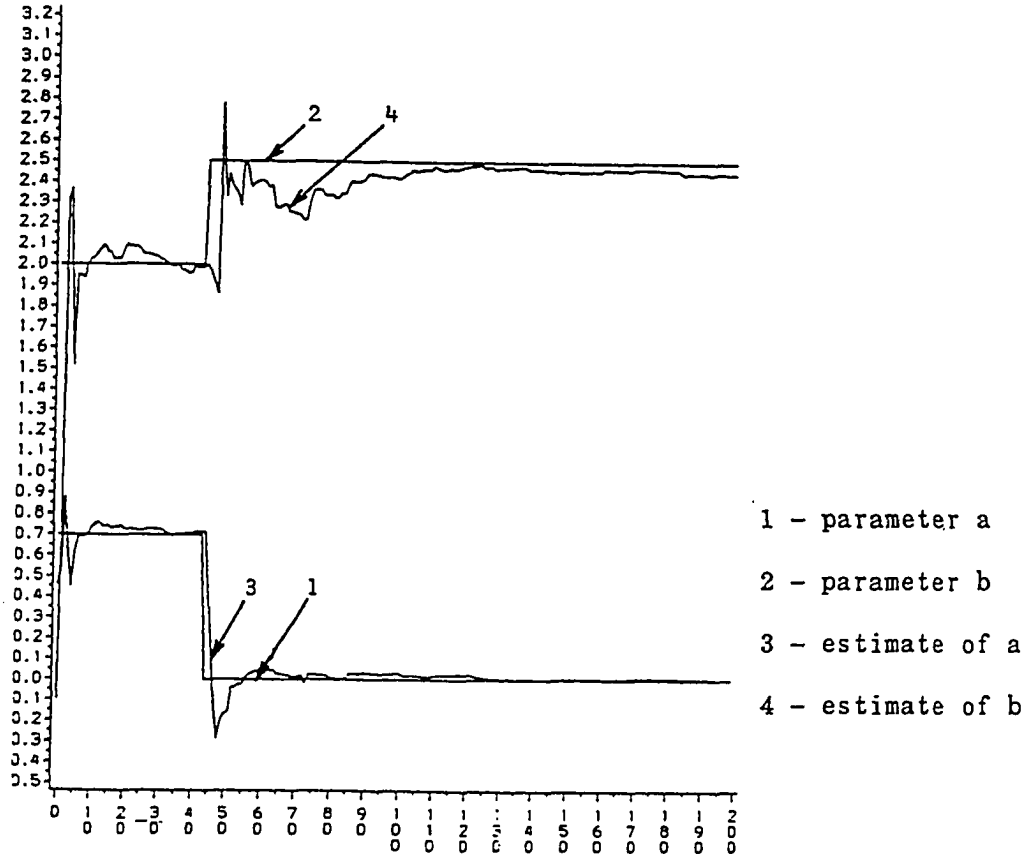

Fig. B.1.5: Performance of the CFA when both parameters are subjected to jumps.

In the second example of this group, the b parameter changes from 2.0 to 1.5, the change in the a parameter remaining the same as before. It is found that the hypothesis test does not fail, but due to the effect of the CDR the estimates are quite satisfactory. Fig. B.1.6 shows the result of this simulation.

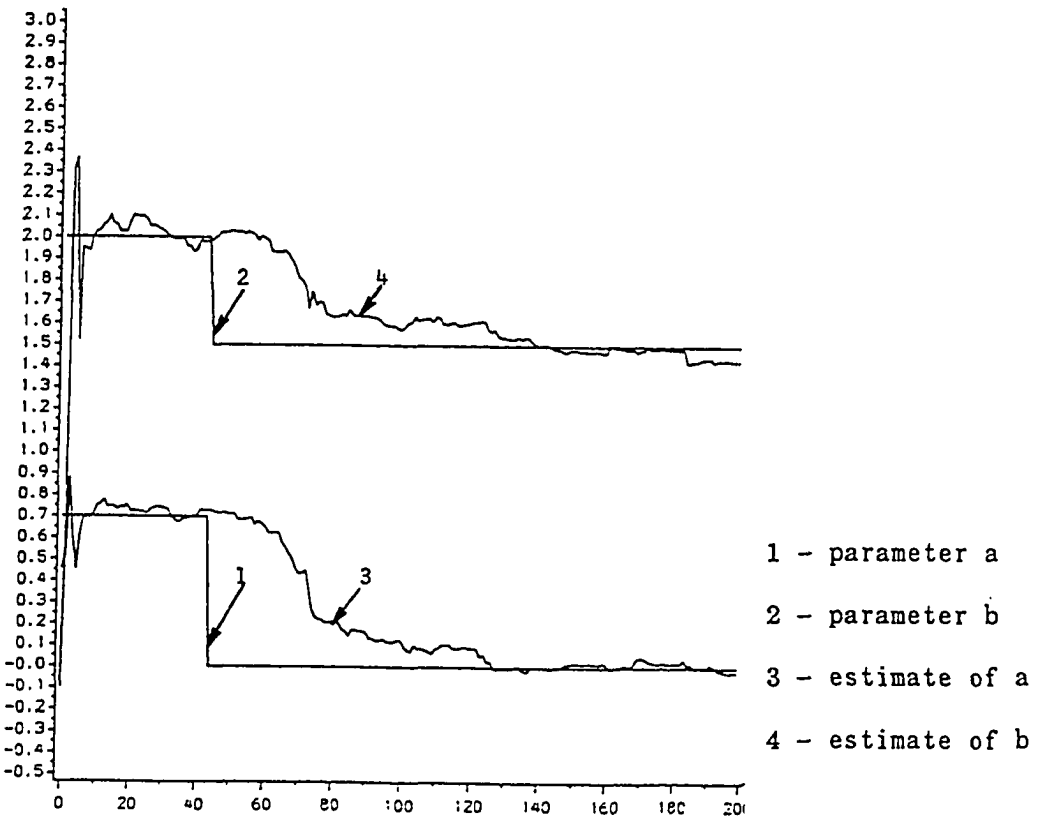

Fig. B.1.6: Performance of the CFA when the jumps are such that instead of re-initialization, the RWM filter starts operating.

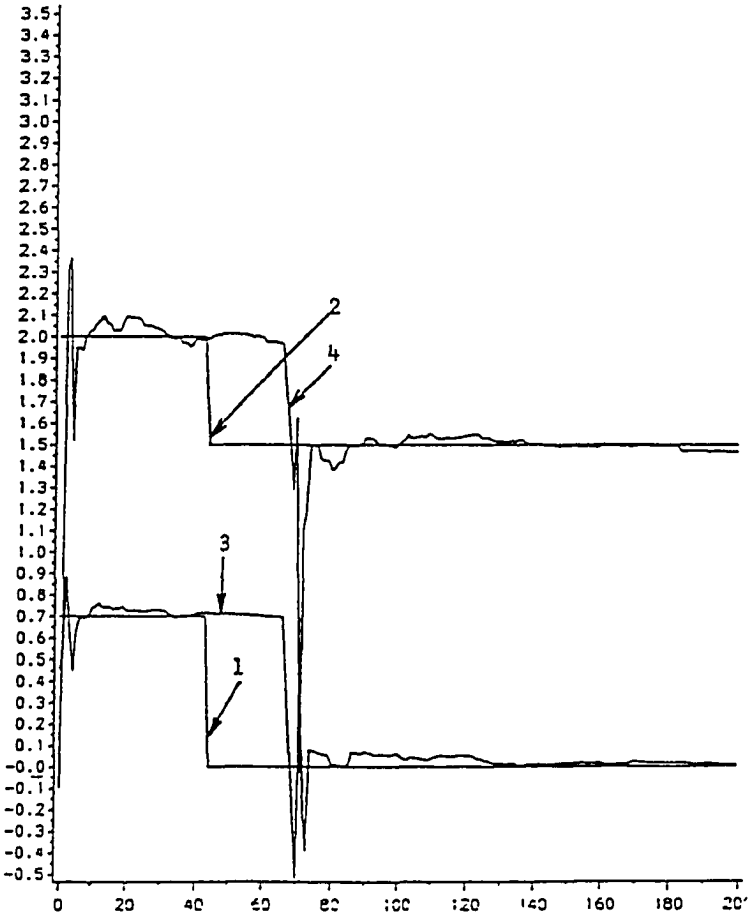

Fig. B.1.7: Example of detection delay.

1 - parameter a, 2 - parameter b,

3 - estimate of a, 4 - estimate of b.

The performance of the filter, as expected, is found to be independent of how many parameters have jumped. Rather, it depends on whether or not the shift created in the normalized residual is large enough to cause a failure of the hypothesis test. The performance also depends on the choice of the design parameters of the algorithm, namely

Q_0, S_k, T_r and T_a. The nature of dependence of the estimates on Q_0 has been well documented in the Kalman filtering literature, and is not studied here. The choice of S_k for the RWM segment of the algorithm is very important. Low S_k makes the performance close to the ordinary Kalman filter. High S_k makes it noisy. Higher T_r and lower T_a make the algorithm approach the RWM Kalman filter by diminishing the segments of CPM operation. Higher T_a makes the algorithm behave close to an ordinary Kalman filter. Lower T_r results in better detection, but also in more false alarms. Fig. B.1.8 through B.1.11 show these effects.

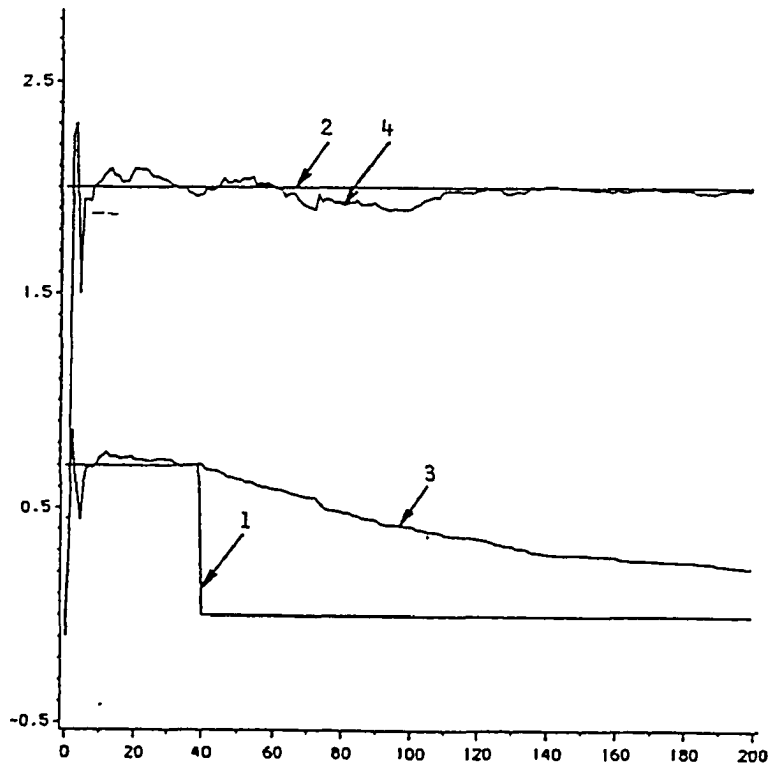

Fig. B.1.8: Elements of the S_k matrix are very small. The operation of the RWM filter is sluggish.

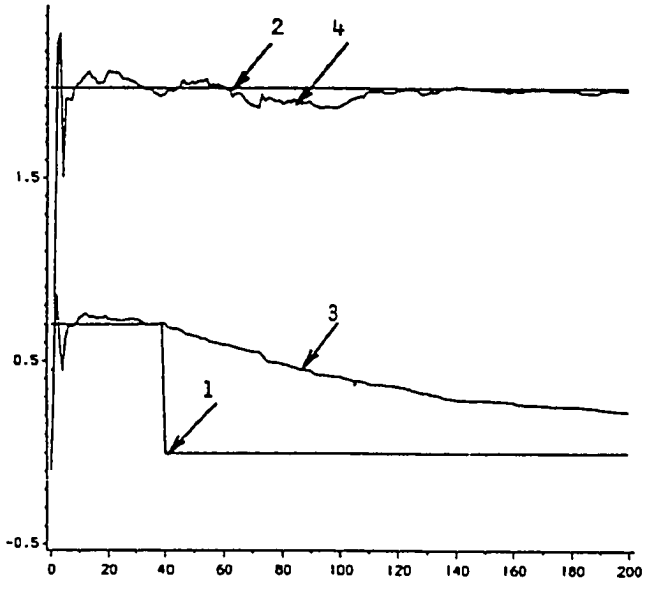

Fig. B.1.9: Result of using high T_a and high T_r. Behavior is similar to CPM filter.

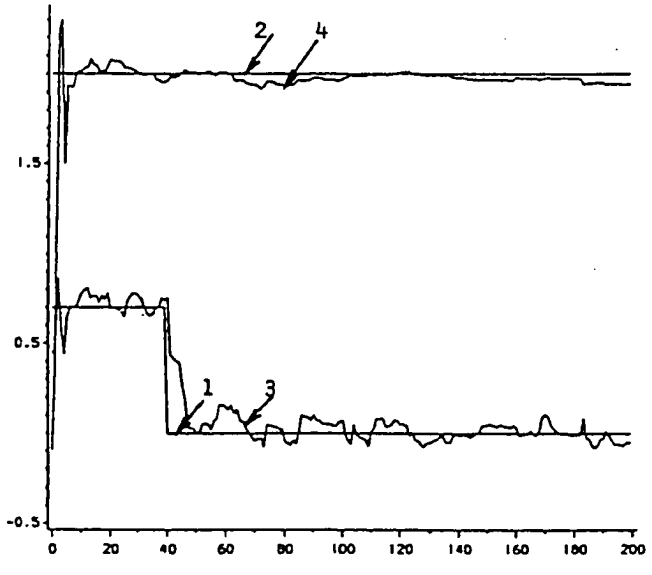

Fig. B.1.10: Result of using high T_r and low T_a. The filter behaves similar to the RWM filter.

1 - parameter a

2 - parameter b

3 - estimate of a

4 - estimate of b

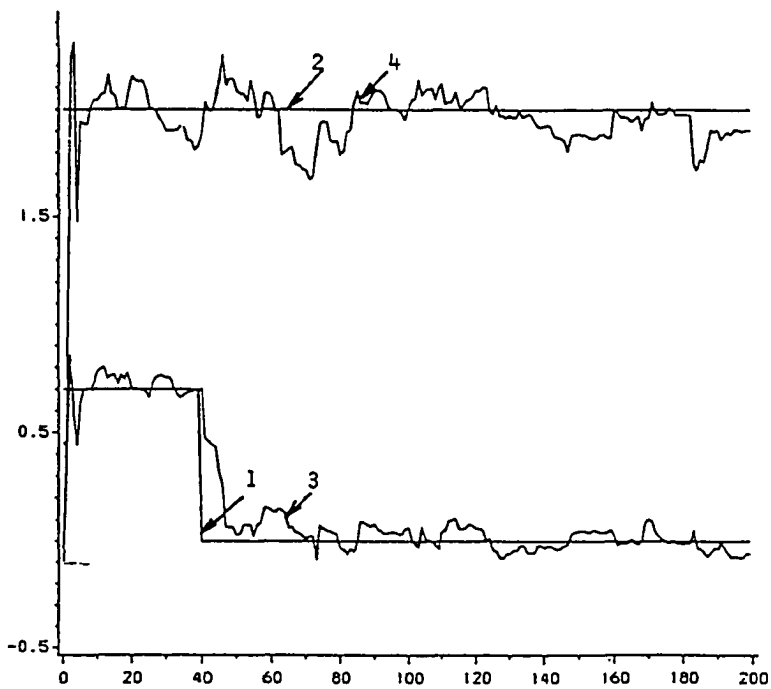

Fig. B.1.11: Result of having large S_k matrix.
The jump is traced, but the estimates are noisy.
1 - parameter a, 2 - parameter b,
3 - estimate of a, 4 - estimate of b.

B.2 Higher Order ARMA Example

A fourth order model is simulated. Here, eight unknown parameters must be estimated. The system is described by the following equation:

$$Z_k = \sum_{i=1}^{4} a_i Z_{k-i} + \sum_{j=1}^{4} b_j U_{k-j} + V_k.$$

In the Kalman filter framework,

$$H_k^t = [\, Z_{k-1} \cdots Z_{k-4} \quad U_{k-1} \cdots U_{k-4}\,],$$

$$X_k^t = [\, a_1 \cdots a_4 \quad b_1 \cdots b_4\,].$$

$\{U_k\}$ and $\{V_k\}$ are timewise uncorrelated Gaussian sequences with standard deviations 2.0 and 0.1 respectively. The true values of the parameters are assumed to be the following:

$$X_k^t = [-0.8 \quad -0.6 \quad -0.4 \quad -0.2 \quad 0.0 \quad 0.05 \quad 0.1 \quad 0.7].$$

In the first experiment the parameters are kept constant. In the second experiment, parameter a_2 changes from -0.6 to 0.0, and b_3 changes from 0.1 to 0.5 at step 90. The resulting estimates are shown in fig. B.2.1 through B.2.4. The performance of the algorithm is found to be satisfactiry. The estimates converge very well.

The behavior of the test statistic is depicted in fig. B.2.5. It is seen that the normalized residual exceeds the rejection threshold (3.3 in this example), which causes the hypothesis test to fail at that point.

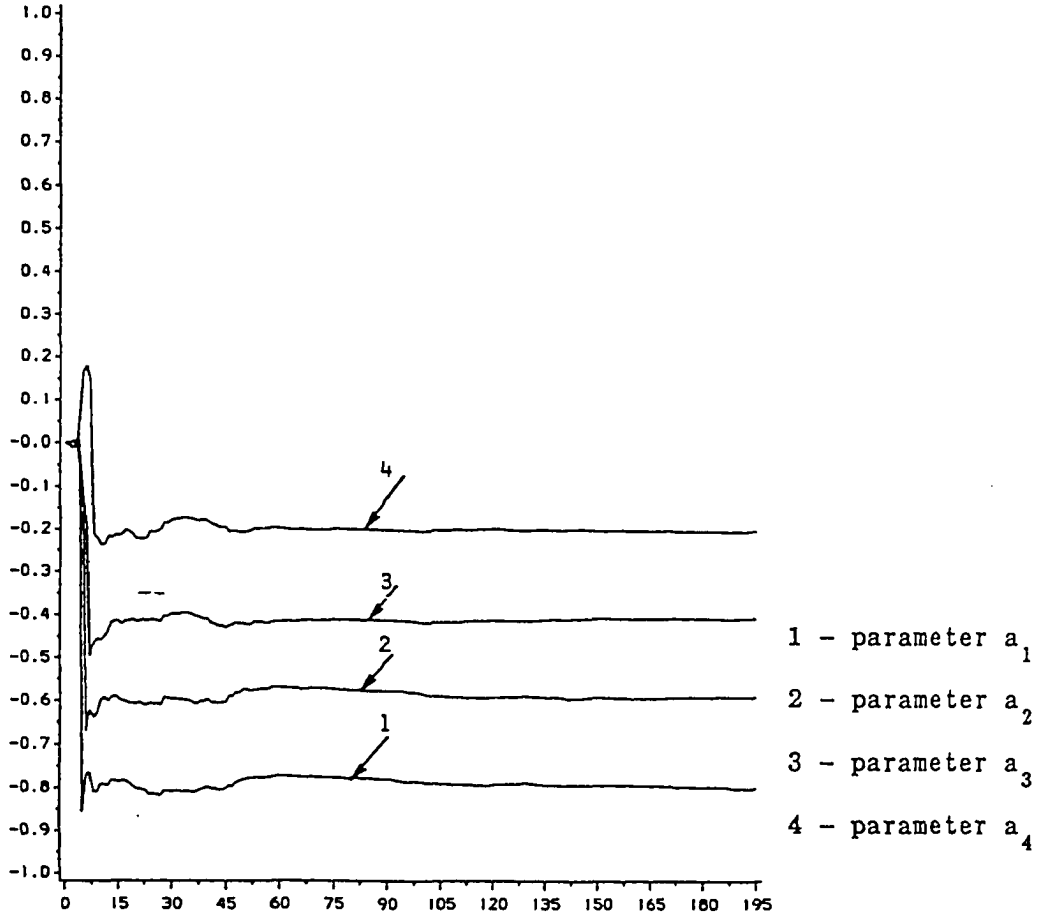

Fig. B.2.1: A parameters of the fourth order system.

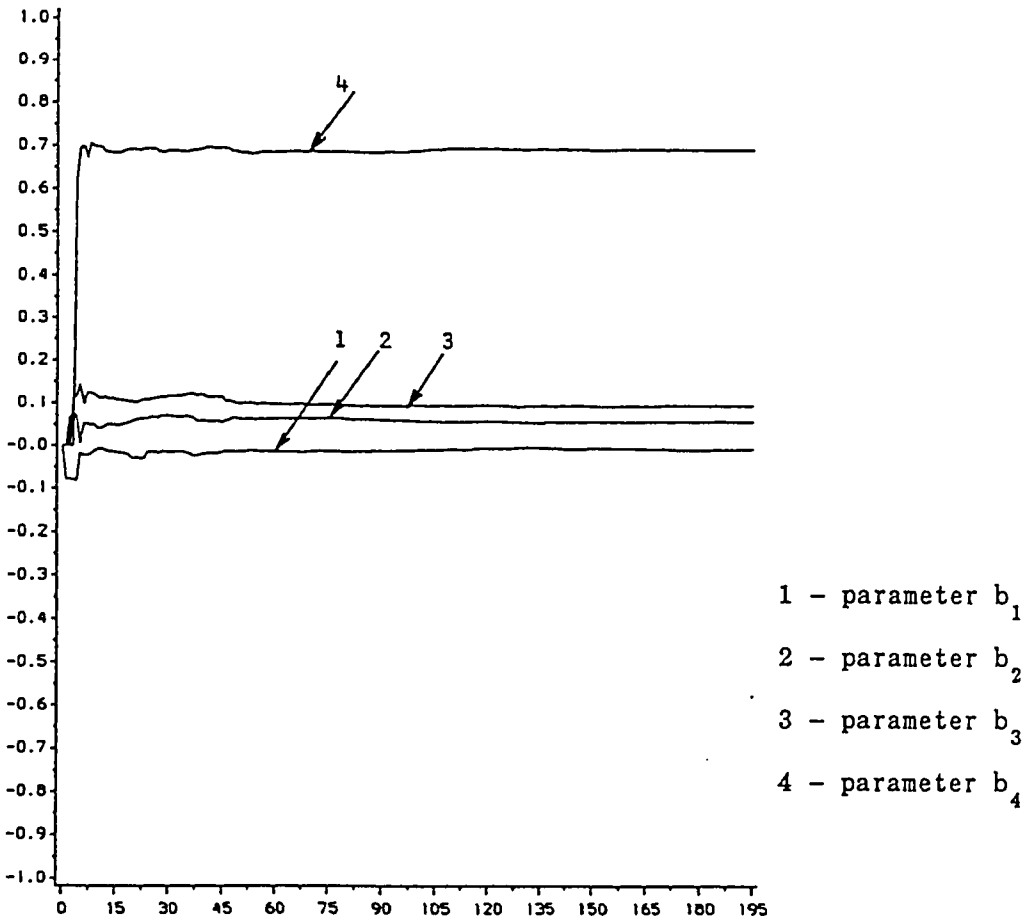

Fig. B.2.2: B parameters of the fourth order system.

In the first run when the parameters are kept constant, they converge to the true values within 45 steps. Keeping in mind that the hypothesis test is optimum only after the Kalman filter has settled down to the steady state, in the second run the jumps are introduced at step

90. The jump is detected quickly, and the new values are accurately estimated within 35 to 45 steps after the jump.

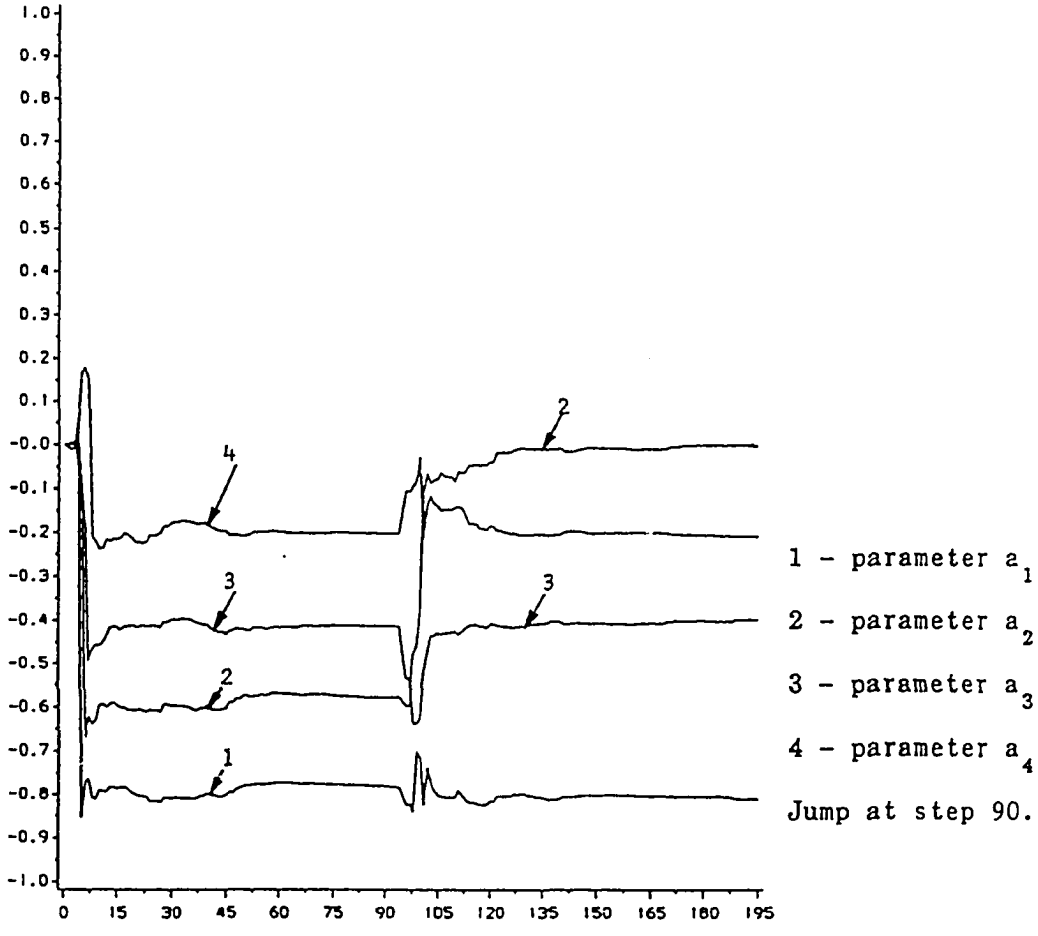

Fig. B.2.3: A parameters of the system with jumps.

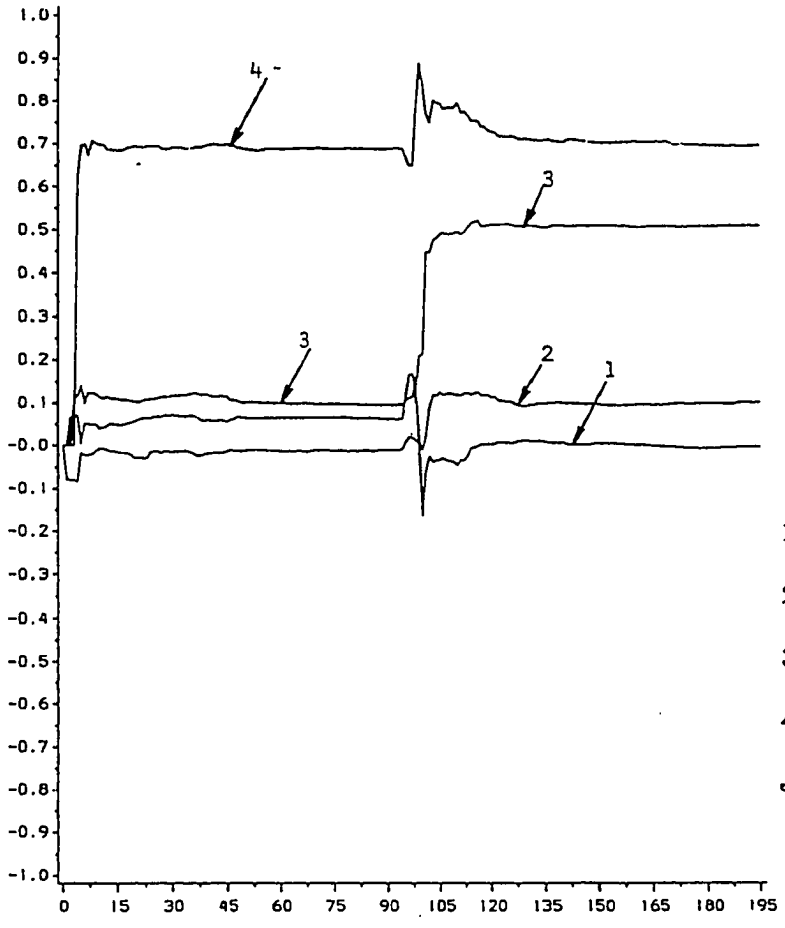

Fig. B.2.4: B parameters of the system with jumps.

The next plot (B.2.5) shows the normalized residual of the system. As seen in the plots of the parameters, there is a detection delay of 3 steps, which is also evident from this plot.

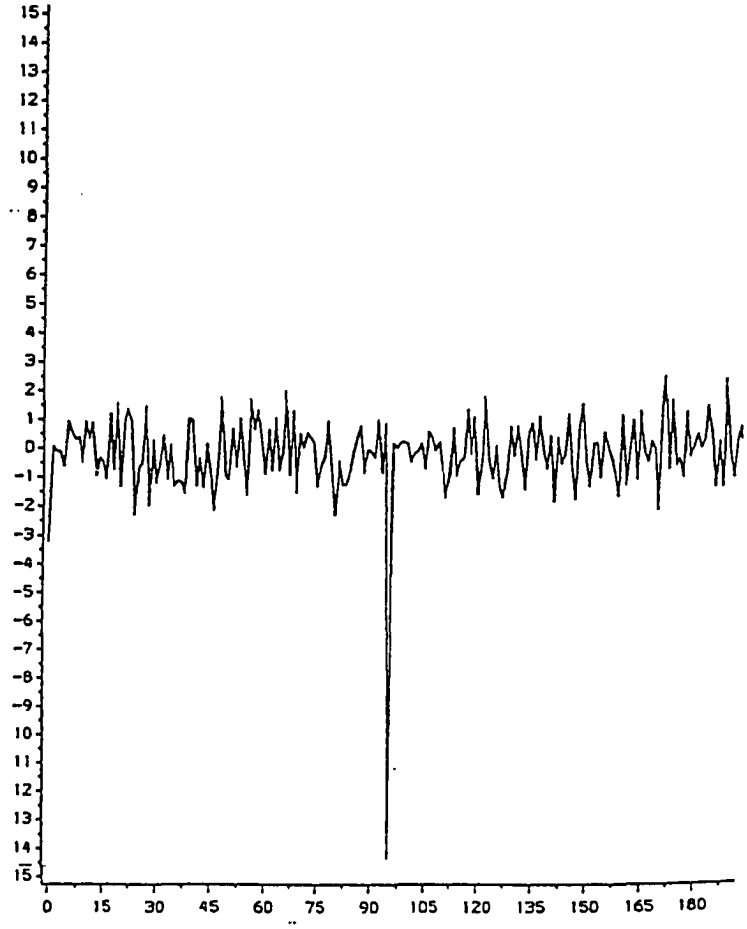

Fig. B.2.5: The normalized residual. The jump is at step 90.

The simulation results described in this appendix confirm the claim that the CFA is efficient in estimating time-varying parameters. The single sample detector is seen to be very fast, although during some runs detection delays are encountered. Some false alarms are also unavoidable. However, false alarm level (α) can be controlled through the choice of the rejection threshold (T_r). The maximum allowable probability of missed jump (β_{max}) can also be controlled via the choice of the acceptance threshold (T_a). Overall, the CFA is seen to be a simple procedure with good performance.

VITA

Fahmida N. Chowdhury was born on July 28, 1955 in Bogra, Bangladesh. She won a scholarship to study Electrical Engineering in the USSR in 1974, and received a Master of Science with Honors from Moscow Power Engineering Institute in 1980. Upon returning to Bangladesh she worked for a short period in the industry, and then taught for two years in the Deaprtment of Electrical and Electronics Engineering at Bangladesh University of Engineering and Technology. She has been a teaching assistant in the Department of Electrical and Computer Engineering at Louisiana State University since August, 1983. Now Mrs. Chowdhury is a candidate for the degree of Doctor of Philosophy in Electrical Engineering.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Fahmida N. Chowdhury

Major Field: Electrical Engineering

Title of Dissertation: On-line Detection and Estimation of Parameter Jumps

Approved:

Major Professor and Chairman

Dean of the Graduate ~~School~~

EXAMINING COMMITTEE:

Date of Examination:

July 28, 1988