1 3
OF/DE

26598

National Library
of CanadaBibliothèque nationale
du CanadaCANADIAN THESES
ON MICROFICHETHÈSES CANADIENNES
SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR Krishnamurthy Giridharagopal

TITLE OF THESIS/TITRE DE LA THÈSE 'Estimation and Control of Stochastic Switching Systems'

UNIVERSITY/UNIVERSITÉ Carleton University

DEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE Ph.D.

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ 1975

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE Professor C.M. Woodside

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE 14/10/75 SIGNED/SIGNÉ K. Krishnamurthy

PERMANENT ADDRESS/RÉSIDENCE FIXE 3/8 A Multi Storey Building, Peshwa Road, New Delhi,
INDIA 110001

INFORMATION TO USERS

THIS DISSERTATION HAS BEEN
MICROFILMED EXACTLY AS RECEIVED

This copy was produced from a microfiche copy of the original document. The quality of the copy is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Canadian Theses Division
Cataloguing Branch
National Library of Canada
Ottawa, Canada K1A 0N4

AVIS AUX USAGERS

LA THESE A ETE MICROFILMEE
TELLE QUE NOUS L'AVONS RECUE

Cette copie a été faite à partir d'une microfiche du document original. La qualité de la copie dépend grandement de la qualité de la thèse soumise pour le microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

NOTA BENE: La qualité d'impression de certaines pages peut laisser à désirer. Microfilmée telle que nous l'avons reçue.

Division des thèses canadiennes
Direction du catalogage
Bibliothèque nationale du Canada
Ottawa, Canada K1A 0N4

ESTIMATION AND CONTROL
OF
STOCHASTIC SWITCHING SYSTEMS

by
K. GIRIDHARAGOPAL

A thesis submitted to the Faculty of Graduate
Studies in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

Faculty of Engineering

Carleton University

Ottawa, Ontario

June, 1975

© K. Giridharagopal 1976

The undersigned hereby recommend to the Faculty of
Graduate Studies acceptance of this thesis submitted by
K. Giridharagopal in partial fulfillment of the require-
ments for the degree of

Doctor of Philosophy.

C. M. Lonsdale
Supervisor

David Hall
Chairman of the Division

External Examiner

कर्मण्येवाधिकारस्ते मा फलेषु कदाचन ।

कर्मफलहेतुर्भूर्मा ते सङ्गोऽस्त्वकर्मणि ।

(Karmanye vadikarasthe ma phalesu kadachana
karmaphalahethurbhurma the sanghasthavakarmani)

You have control over action alone,
never over its fruits,
live not for the fruits of action,
nor attach yourself to inaction.

- Bhagavad Gita
Verse 48, Chapter 2

ACKNOWLEDGEMENTS

The author takes this opportunity to express his sincere thanks to his thesis supervisor, Dr. C.M. Woodside, for his constant encouragement and guidance throughout the course of this work.

The financial support received from Carleton University and the National Research Council of Canada is gratefully acknowledged.

Thanks are also due to Mrs. Sandra Ryan for her care and patience in typing this thesis.

Finally, the author wishes to place on record his appreciation of the cooperation and encouragement provided by friends and colleagues, notably, E. Baillargeon and S.A. Mahmoud.

ABSTRACT

Estimation and control in discrete-time stochastic systems with two modes of operation are considered. A unified modelling framework for these systems, based on the incorporation of a binary switching variable in the system equations, is presented. The dynamics of the switching process are represented by a continuous model, which is subsequently used for the joint estimation of the state and the switching variable via an extended Kalman filter. This approach provides equally good estimates and requires less storage and computation compared to other known approaches.

For parameter estimation, the difference equation model for a switching system is considered. The coefficients of the difference equation are functions of both the current as well as the past values of the switching variable.

Parameter estimation with this model requires a large computational effort even for small scale systems. A simplification results on assuming that the coefficients of the underlying differential equation switch simultaneously, in which case the coefficients of the corresponding difference equation are functions only of the current value of the switching variable. For this simplified model, modi-

fications of the ordinary least squares and the boot-strap estimators employed for non-switching systems provide acceptably accurate estimates of the parameters.

Adaptive controllers are obtained for two types of switching systems, namely, systems with missing observations and those with switching time constants. These adaptive controllers provide a better performance of the system compared to the certainty equivalent controller while reducing the computational requirements of the optimal stochastic controller. Control, with a model of the switching process embedded in the system equations, results in a performance comparable to that of the certainty equivalent controller.

Prevention or limiting of the divergence of the Kalman filter based on an erroneous model is studied. It is shown via simulation studies that compensation for modelling errors and hence the limiting of divergence can be achieved by the addition of a switching plant noise. This approach to modelling error compensation facilitates the use of a simple model for representing a complex industrial process thus easing the analysis and the design of the process.

Several examples are included to illustrate the use and the scope of the various approaches to estimation and control of switching systems.

LIST OF SYMBOLS

The symbols that have been used throughout the text are listed here. Symbols whose use is restricted to a specific section are defined in the appropriate section. The numbers in the parenthesis refer to the equation number in the text where the symbol appears for the first time.

$()^T$	Transpose of matrix
$ (.) $	Determinant of matrix or absolute value of scalar
$ (.) $	Norm of a vector
$x(k)$	Vector of state variables (dimension n) (2.1)
$U(k)$	Deterministic input to the process (dimension p) (2.1)
$y(k)$	Noise corrupted system output (dimension m) (2.1)
$\xi(k)$	System noise (2.1)
$n(k)$	Measurement noise (2.1)
i	As a subscript, denotes the i th mode (2.2)
A_i, B_i, C_i	System and output matrices for the i th mode (2.2)
$\gamma(k)$	Indicator or switching variable designating the mode (2.3)
q	Initial probability of γ i.e. $p(\gamma_0=1)$
q_{ij}	Transition probabilities of the Markov chain for γ (2.4)

$Q(k)$	Variance of system or plant noise (2.5a)
$R(k)$	Variance of measurement noise (2.5a)
Q_i, R_i	Variances of the plant and the measurement noises for the i th mode (2.2)
I_k	Sequence of γ 's up to time k (2.21)
$\hat{x}_k, \hat{x}_{k k}$	Estimates of x_k based on observations up to time k (2.5)
Y^k	Collection of outputs up to time k (2.5)
U^k	Collection of inputs up to time k (5.6)
$P_k, P_{k k}$	Estimation error covariance (2.19)
k	Discrete time index (used interchangeably as $(.)_k$ or $.(k)$) (2.1)
f	Model function for the prediction of γ_k (3.1)
$\hat{\gamma}_{k k-1}$	Value of γ_k predicted from the model (3.2)
$\gamma_{k k}^*$	Optimal estimate of γ_k (3.7)
$\hat{\gamma}_{k k}$	Estimate of γ_k resulting from the use of the model function f (3.3)
$\psi(k)$	Noise added to the model function f (3.1)
Q_ψ	Variance of $\psi(k)$ (3.16)
$\theta^{(i)}$	Vector of system parameter corresponding to the i th mode (5.5)
J	Criterion function to be minimized (6.2)
N	Length of the identification or control experiment

TABLE OF CONTENTS

	<u>PAGE</u>
CHAPTER 1. INTRODUCTION	1
1.1 Introduction to the Problem	1
1.2 Occurrence and Use of Switching Models	2
1.3 Review of Existing Literature	3
1.4 Objective and Scope of the Thesis	6
1.5 Contributions of This Study	7
1.6 Thesis Organization	8
CHAPTER 2. STATE ESTIMATION FOR SWITCHING SYSTEMS	12
2.1 Introduction	12
2.2 The Problem	13
2.3 Optional Linear Estimator	16
2.4 Optimal Non-linear Estimator	22
2.5 Sub-Optimal Non-linear Estimators	26
CHAPTER 3. MODELLING THE SWITCHING PROCESS	33
3.1 Preliminary Remarks	33
3.2 The Model	33
3.3 Models for Some Special Cases	45
3.4 Pros and Cons of Modelling	49
CHAPTER 4. JOINT DETECTION/ESTIMATION IN SWITCHING SYSTEMS VIA MODELLING	54
4.1 Introduction	54
4.2 The Extended Kalman Filter	55
4.3 Estimation with State Dependent Noise	56
4.4 Constrained Estimation	59
4.5 Examples	60
4.6 Conclusions	73
CHAPTER 5. IDENTIFICATION OF SWITCHING SYSTEMS	83
5.1 Introduction	83
5.2 The Problem	85
5.3 Off-Line Identification	89
5.4 On-Line Identification	93
5.4.1 Uncorrelated Errors	95
5.4.2 Correlated Noise	106
5.5 Conclusions	113

	<u>PAGE</u>
CHAPTER 6. CONTROL OF SWITCHING SYSTEMS	125
6.1 Introduction	125
6.2 The Problem	127
6.3 Effect of Control on Estimation Accuracy	130
6.4 Control of Systems with Missing or Interrupted Observations	132
6.4.1 γ -optimal and Certainty Equivalent Controllers	133
6.4.2 Adaptive Control	135
6.5 Stochastic Control of Systems with Switching Time Constants	143
6.5.1 γ -optimal Controller	145
6.5.2 Adaptive Control	150
6.5.3 Modelling Approach	152
6.6 Conclusions	156
CHAPTER 7. ERROR COMPENSATION VIA SWITCHING NOISE	166
7.1 Introduction	166
7.2 Divergence of the Kalman Filter	167
7.3 Switching Plant Noise	169
7.4 Modelling Boiler Dynamics	177
7.5 Conclusions	184
CHAPTER 8. CONCLUSIONS	201
APPENDIX I. INPUT DESIGN FOR ESTIMATION IN SWITCHING SYSTEMS	207
APPENDIX II. CHOICE OF Q_v - COVARIANCE OF NOISE ADDED TO MODEL	214
APPENDIX III. CONSTRAINED LINEAR ESTIMATION	221
REFERENCES	227

LIST OF FIGURES

<u>FIGURE NO.</u>		<u>PAGE</u>
1.1	A switching system with two modes of operation.	11
2.1	Structure of the optimal estimator for a switching system with two modes.	31
2.2	Structure of the approximate estimators for a switching system with two modes.	32
3.1	Shape of the model for the switching process in the detection example ($q_{10}=0.1$).	51
3.2	Comparison of the mean and the mean squared errors in the detection example.	52
3.3	Shape of the model for the switching process in the case of a switching plant noise ($q_{10}=0.1$).	53
4.1a	Mean and RMS estimation errors in the scalar example with missing observations (modelling and Sawaragi approaches compared with a fixed γ -sequence).	74
4.1b	Mean and RMS estimation errors in the scalar example with missing observations (modelling and Sawaragi approaches compared with a variable γ -sequence).	75
4.1c	State and γ -estimates in a typical run of the scalar system with missing observations (Example 4.1). Modelling and Sawaragi approaches compared.	76
4.2a	Comparison of average estimation errors for the second order example with missing observations. (Q_{ψ} for modelling approach given by equation 3.17.)	77
4.2b	Comparison of the mean squared estimation errors.	77

FIGURE NO.PAGE

4.3a	Estimates of the state and the parameter 'a' in a typical run of the system in Example 4.3. Modelling and the Bruckner approaches compared.	78
4.3b	Comparison of the mean squared errors in the estimation of the state in Example 4.3.	79
4.3c	Comparison of the mean squared errors in the estimation of the parameter 'a' in Example 4.3.	80
4.4a	Comparison of the mean and the mean squared estimation errors for the scalar example with a switching time constant.	81
4.4b	State and γ -estimates in a typical Monte-Carlo run of the system with a switching time constant in Example 4.4.	82
5.1	The structure of the on-line identification scheme for switching systems.	116
5.2	Mean squared errors in the estimation of the parameter 'a ₁ ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal approaches.	117
5.3	Mean squared errors in the estimation of the parameter 'b ₁ ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal approaches.	118
5.4	Mean squared errors in the estimation of the parameter 'a ₀ ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal approaches.	119
5.5	Mean squared errors in the estimation of the parameter 'b ₀ ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal.	120

<u>FIGURE NO.</u>		<u>PAGE</u>
5.6	Variance of estimates vs no. of observations in the smoothing approach for Example 5.1.	121
5.7	Mean squared errors in $\hat{a}_1$ in the scalar example with correlated errors.	122
5.8	Mean squared errors in $\hat{b}_1$ in the scalar example with correlated errors.	122
5.9	Mean squared errors in $\hat{a}_0$ in the scalar example with correlated errors.	123
5.10	Mean squared errors in $\hat{b}_0$ in the scalar example with correlated errors (the modified version of 2SLS, LS and boot-strap approaches compared).	123
5.11	Variance of estimates vs no. of observations for Example 5.2 with boot-strap estimator.	124
6.1	Typical control histories for Example 6.2.	160
6.2	State trajectories for Example 6.2.	161
6.3	Variation of adaptive controller performance with ϵ .	162
6.4	Control histories and estimates of γ in a typical Monte-Carlo run of the scalar system with missing observations (comparison of CE controls with modelling and Sawaragi approaches).	163
6.5	State trajectories in a typical Monte-Carlo run of the scalar system with a switching time constant.	164
6.6	Comparison of control histories and the estimates of γ in a typical Monte-Carlo run of the system with a switching time constant.	165
7.1	Comparison of state estimates in a typical run.	191
7.2	Comparison of the mean and the mean squared estimation errors in example 7.1. The modelling approach compared with the Quigley approach.	192

<u>FIGURE NO.</u>		<u>PAGE</u>
7.3a	Step responses of pressure and temperature comparison of the third and the ninth order models.	193
7.3b	Step responses of water level with the third and the ninth order models.	194
7.4a	Variation of mean squared error in the estimation of the water level, y , with the variance Q_{yy} of the noise added to the model ($Q_{T_s T_s} = 0, Q_{pp} = 0$).	195
7.4b	Variation of mean squared error in the estimation of the pressure, p , with the variance Q_{pp} of the noise added to the model ($Q_{yy} = 10^{-1}, Q_{T_s T_s} = 0$).	196
7.4c	Variation of mean squared error in the estimation of the temperature, T_s , with the variance $Q_{T_s T_s}$ of the noise added to the model ($Q_{yy} = 10^{-1}, Q_{pp} = 10^4$).	197
7.5a	Comparison of the mean squared errors in the estimation of the temperature, T_s , of the boiler via the three approaches: Q-constant, Modelling, Quigley.	198
7.5b	Comparison of the mean squared errors in the estimation of the water level, y , of the boiler via the three approaches: Q-constant, Modelling, Quigley.	199
7.5c	Comparison of the mean squared errors in the estimation of the pressure, p , of the boiler via the three approaches: Q-constant, Modelling, Quigley.	200
II.1	Variation of the best Q_{ψ}, Q_{ψ}^* with $Q \ln(a)/x^2$ $R/\bar{x}^2$ is a parameter.	220

LIST OF TABLES

<u>TABLE NO.</u>		<u>PAGE</u>
5.1	Mean and mean squared errors at the end of 30 time steps (averaged over 30 runs) for Example 5.1.	114
5.2	Mean and mean squared errors at the end of 100 time steps (averaged over 30 runs) for Example 5.2.	115
6.1	Comparison of costs with the various approaches to the control of the system with missing observations in Example 6.2.	158
6.2	Comparison of costs for the modelling, the adaptive and the γ -optimal approaches to the control of a system with a switching time constant.	159
7.1a	Matrix A.	185
7.1b	Matrix B.	186
7.2	The Matrix A^* ; The Matrix B^* .	187
7.3a	Matrix A_d .	188
7.3b	Matrix B_d .	189
7.4	The Matrix A_d^* ; The Matrix B_d^* .	190

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION TO THE PROBLEM

This thesis considers a class of systems called variously as 'Jump Parameter System' (S5), 'Systems with Random Models' (H1), or 'Systems with Multiple Operational Modes' (B1). These are systems where several modes of operation are possible (for example, an operating mode and several different failure modes) and the transition from one mode to another is governed by a random sequence independent of the system variables. Corresponding to each one of these operational modes, there is a distinct set of system model parameters. The phrase 'Stochastic Switching Systems' is used in this dissertation to describe these systems - switching because the system switches modes and stochastic because the switch occurs at random times. The model describing the system dynamics is, in this sense, called a switching model.

The aim of this dissertation is to develop simple schemes for obtaining information about a stochastic switching system - information that is contained in the state and the parameter vectors of the state-space model

of the system and in the statistical properties of the random disturbances influencing the system. This information, which is necessary for the design of a suitable controller, requires state and parameter estimation.

The general problem attacked in this thesis can now be stated: Given the several mathematical models describing the dynamics of a stochastic switching system and the statistical properties of the random process describing the switch from one model to another, estimate, recursively, the states and any unknown parameters of the system and design a controller such that the system response is optimized in some chosen sense.

1.2 OCCURRENCE AND USE OF SWITCHING MODELS

Switching models are appropriate for systems in which sudden changes occur in the characteristics of either the system or the disturbances affecting it. There are several practical contexts in which such models are encountered:

- (i) failure of one or more system components; for example, sudden changes in the gain, the time constant or even the order of the system;
- (ii) failure of the measurement mechanism, i.e., the gain of the measurement transducer switches to zero at random times;

- (iii) sudden changes in the environment (the statistical properties of the random disturbances). The case of burst errors in the measurements also falls into this category.

Switching models can also be used to advantage in approximately describing a system by the discretization of some critical variables and thus ease the design of the controller. Another use of switching models is in the compensation for modelling errors by the incorporation of a switching error term.

1.3 REVIEW OF EXISTING LITERATURE

One of the earliest attempts to analyze switching systems appears to be due to Mishulina^(M4). He considered linear, continuous-time, deterministic systems with a structure alternating randomly between two given models, and obtained closed form expressions for the system response. Yashin^(Y1,Y2) developed state estimation schemes for the same type of systems subject to random disturbances.

In the realm of discrete-time, stochastic switching systems, one of the first types to be considered was that of uncertain (or missing or interrupted) observations. This corresponds to the case of random failure of the

measurement transducer. Nahi^(N1) derived the optimal linear estimator for this type of switching systems. Sawaragi et al^(S3) considered the same problem and showed, via simulation studies, that an approximate non-linear estimator performs better than the optimal linear estimator.

Ackerson and Fu^(A1) considered the problem of switching environments (i.e. the statistics of the random disturbances influencing the system). They derived the optimal non-linear estimator and, recognizing the impractical nature of this solution, developed an approximate non-linear estimator. Their result was later extended by Bruckner and Scott^(B1) and Moose and Wong^(M5) for estimating the states of a general switching system. In this context, Kohlas^(K5) has shown that the problems of Bad-data^(M3) and Rejection of Outliers^(A4) can be posed as estimation problems in switching systems.

Even though there exists a wealth of literature on the identification of non-switching systems, as evident from the extensive bibliography in the survey by Astrom^(A5), the subject of switching systems appears to have attracted no attention. From this point of view, the work reported in this dissertation is believed to be the first attempt in this direction.

Several articles have appeared on the control of switching systems. Swarder et al^(S5) considered continuous-time switching systems and obtained the optimal controller under the assumption that the mode of the system is known exactly. A similar result was obtained by Rosenbrock^(R2), who considered the case of a single, non-reversible switch in the mode of the system during the control period. For the discrete-time case, Aoki^(A6) obtained the optimal Bayes' controller for a system with a switching plant model (i.e. the observation process does not switch). The optimal controller derived by Fujita and Fukao^(F4) for systems with missing observations has since been identified as the separation controller (see Wernersson^(W1) or Fujishige^(F3)).

Some authors have considered the design problem, namely, the effect of introducing the multimodal feature into a system which is inherently of the non-switching type. Kazangey^(K4) represented the variation in the interest rates by a switching model and obtained optimal policies for the housing sector. Swarder and Robinson^(S5) adopted a similar approach to optimize the management of water resources. Quigley^(Q1) introduced a switching plant noise in an attempt to improve the accuracy of estimation

with an erroneous model. These articles suggest that it is possible to improve estimation accuracy and ease the design of a suitable controller, by the incorporation of a switching model to represent some critical variables.

1.4 OBJECTIVE AND SCOPE OF THE THESIS

The various articles in the literature on estimation and control of switching systems are either restrictive in some sense or demand a large computational effort. The aim of this thesis is to obtain simple and easily implementable recursive solutions for the problems of estimation, identification and control of stochastic switching systems and to examine the use of a switching model for improving estimation accuracy.

Attention in this thesis is restricted to the case of two possible modes of operation of the system and hence two possible models. It is further assumed that in either of the two modes the system is described by a discrete-time, linear model and that any random disturbances influencing the system are of the additive, white, Gaussian type. The structure of the system under consideration is shown in Fig. 1.1. In this Figure, γ_k is an indicator variable taking values 0 and 1 indicating the two possible modes of the system. Estimation or control with this

system requires the estimation of γ first and then the decision to use either one of the two models. It is noticed that by a proper choice of the statistics of γ , say $p(\gamma_k=1)=1/k$, the system in Fig. 1.1 reverts to the traditional linear, time invariant system.

1.5 CONTRIBUTIONS OF THIS STUDY

A principal contribution of this study lies in the development of a model describing the dynamics of the switching process. An extended Kalman filter based on this model provides equally good state estimates and requires less storage and computation compared to other known approaches.

The development of parameter estimation schemes for switching systems represents another contribution of this thesis. With a literature apparently devoid of any articles on this subject, the work reported here is believed to be the first step in this direction.

A third contribution lies in the development of adaptive controllers for two typical switching systems, namely, systems with missing observations and those with switching time constants. These controllers provide a better performance of the system compared to the certainty equivalence controller while requiring lesser computation than the optimal stochastic controller.

1.6 THESIS ORGANIZATION

This dissertation consists of four principal parts. Chapters 2, 3 and 4 examine the problem of state estimation. In Chapter 2, a review of the existing literature and the shortcomings of the various approaches known to date are presented. A new concept is introduced in Chapter 3. Briefly, a dynamic model describing the time evolution of the indicator variable (also called the switching variable) γ_k is developed and models for various special situations are indicated. In Chapter 4, this model is employed for the estimation of the states and the performance of the estimator based on the model for γ_k is compared with those of the estimators reported in the literature. Several examples are presented for illustration and comparison.

Chapter 5 considers the identification problem. It is shown that the ordinary least squares and the instrumental variable approaches taken for constant systems^(A5) can be extended to estimate the parameters of a switching system represented by a simple difference equation model. Two illustrative examples are presented and possible extensions to a general difference equation model are outlined.

The problem of controller design is studied in Chapter 6. Following a discussion of the optimal controller, sub-optimal and adaptive controllers are obtained for two typical switching systems, namely, systems with missing observations and those with switching time constants. Performance of the adaptive controllers are compared with that of a controller designed under the assumption that the mode of the system (the value of γ_k) is known exactly.

Chapter 7 explores a possible application for switching models. It is shown that the use of a switching noise at the input to the system can correct for errors in the mathematical model of the system, and the resulting estimation errors are far smaller in magnitude than those obtained with a traditional Kalman filter based on fixed statistics of input or plant noise. This approach provides a substantial reduction in the estimation errors occurring from the use of a third order model to describe the dynamics of a steamraising plant instead of the original ninth order model⁽¹⁾). With this approach, the estimation errors are small enough to make the design of a controller for the boiler, based on the third order model, meaningful.

Conclusions drawn in the various chapters are summarized in Chapter 8. This Chapter also outlines possible extensions of the work reported in this dissertation.

The three Appendices included in this dissertation present ideas of lesser importance but nevertheless useful.

Appendix I presents the results of a Monte-Carlo study of the variance of the fictitious noise added to the model for γ . Appendix II discusses two approaches to the design of input signals with the view of improving the performance of the estimation and the identification schemes considered in this thesis. In Appendix III, the optimal constrained estimator is derived - a result utilized in Chapter 4 for state estimation.

L

FIG. 1.1. A switching system with two modes of operation.

CHAPTER 2

STATE ESTIMATION FOR SWITCHING SYSTEMS

2.1 INTRODUCTION

The advent of high speed data processing machines has resulted in a surge of activity towards the development of advanced techniques for data analysis. One such technique is the subject of this Chapter, namely, the estimation of random, Markov processes that are generated, at any time instant, by either one of two or more possible mechanisms. The type of estimates are minimum mean square error (MMSE) estimates and are related closely to the linear filter developed by Kalman^(K2).

Several techniques for state estimation of switching systems, i.e. systems whose model switches randomly between two given models, have appeared in the literature. Ackerson and Fu^(A1) presented the optimal (conditional mean) non-linear estimator and showed that the resulting estimator can be represented by a "bank" of Kalman filters. Their observation that the number of Kalman filters needed to estimate the state vector doubled with each incoming observation led to a series of articles on approximate non-linear filters^(B1,M5,S3). Nahi^(N1) derived the

optimal linear estimator for a special class of problems, namely, the estimation of the states of systems with missing observations. Though Ackerson and Fu^(A1) refer to an optimal linear estimator for the case of switching environments, there is no derivation presented either in their paper or in the doctoral dissertation^(A2) the article came out of. For this reason, the optimal linear estimator is derived in Section 2.3. The non-linear estimator of Ackerson and Fu and the various approximations leading to sub-optimal estimates are discussed in Sections 2.4 and 2.5, respectively. It is worth noting here that the problem of bad data studied by Merrill and Schweppe^(M3) and Kohlas^(K5) and the problem of state estimation for systems with random models analyzed by Haddad^(H1) can both be posed as problems of state estimation in switching systems.

2.2 THE PROBLEM

The system under consideration is described by a set of linear difference equations:

$$\begin{aligned}x(k+1) &= A x(k) + B u(k) + \xi(k) \\ y(k+1) &= C x(k+1) + \eta(k+1)\end{aligned}\tag{2.1}$$

The usual notation is used in 2.1, viz.,

x is the n -vector of states

U is the m -vector of inputs.

y is the p -vector of observations

ξ is the n -vector of plant disturbances

η is the p -vector of observation noises.

The matrices A , B and C have appropriate dimensions. $\xi(k)$ and $\eta(k)$ are samples from uncorrelated, zero-mean, Gaussian white noise processes with covariances Q and R respectively. They are further assumed independent of the initial state x_0 which has a Gaussian distribution $N(\hat{x}_0, P_0)$.

The system is identical to the familiar linear, minimum variance problem in all but one respect. The significant difference is that the system descriptors, i.e., the set $\mathcal{S} = [A, B, C, Q, R]$, are not uniquely specified. Rather, the system is defined by the set $[\mathcal{S}_i, i=0, s]$, of such descriptors where the individual descriptors $\mathcal{S}_i$ designate the mode of the system.

Introduce an indicator variable γ_k (called the switching variable in the sequel), which takes values $0, 1, 2, \dots, s$. Then, $\gamma_k = i$ implies that the system equations, at time k , are:

$$\begin{aligned} x(k) &= A_i x(k-1) + B_i U(k-1) + \xi_i(k-1) \\ y(k) &= C_i x(k) + \eta_i(k) \end{aligned} \tag{2.2}$$

In the case of two modes ($s=1$) considered in this thesis, the system can be compactly represented by the equations:

$$\begin{aligned} x(k) &= [A_1\gamma_k + A_0(1-\gamma_k)]x(k-1) + [B_1\gamma_k + B_0(1-\gamma_k)]u_{k-1} + \gamma_k\xi_1(k-1) \\ &\quad + (1-\gamma_k)\xi_0(k-1) \\ y(k) &= [C_1\gamma_k + C_0(1-\gamma_k)]x(k) + [\gamma_k\eta_1(k) + (1-\gamma_k)\eta_0(k)] \\ \gamma_k &= 0 \text{ or } 1 \end{aligned} \quad (2.3)$$

$\xi_i(k)$ and $\eta_i(k)$, $i=0,1$ are sample vectors from zero-mean, independent, white, Gaussian processes with variances Q_i and R_i , respectively. The transition from one mode to another is described by a Markov-chain; therefore,

$$p(\gamma_k=i | \gamma_{k-1}=j) = q_{ij} \quad (2.4)$$

where q_{ij} is an element of the transition matrix defining the Markov-chain.

The problem is to obtain an estimate $\hat{x}_{k/k}$, of the state x_k in the system described by (2.3) and (2.4), given the observations up to time k , Y^k , where $Y^k \triangleq \{y_0, y_1, \dots, y_k\}$, such that the criterion

$$E\{||x_k - \hat{x}_{k/k}||^2\} \quad (2.5)$$

is minimized.

Consider first the optimal linear estimator for the system given by (2.3).

2.3 OPTIMAL LINEAR ESTIMATOR

Nahi^(N1) derived the optimal linear filter for the special case of missing observations with a further restriction that γ_k are independently distributed. In this section, the optimal linear filter for the general system (2.3) is obtained. The derivation follows closely the approach taken by Nahi. For the sake of a clear presentation, the following quantities are defined:

$$A(k) = \gamma_k A_1 + (1 - \gamma_k) A_0$$

$$B(k) = \gamma_k B_1 + (1 - \gamma_k) B_0$$

$$C(k) = \gamma_k C_1 + (1 - \gamma_k) C_0$$

$$Q(k) = \gamma_k Q_1 + (1 - \gamma_k) Q_0$$

$$R(k) = \gamma_k R_1 + (1 - \gamma_k) R_0$$

$$\xi(k-1) = \gamma_k \xi_1(k-1) + (1 - \gamma_k) \xi_0(k-1)$$

$$\eta(k) = \gamma_k \eta_1(k) + (1 - \gamma_k) \eta_0(k)$$

$$I_k = \{\gamma_0, \gamma_1, \dots, \gamma_k\} \quad (2.5a)$$

It is now required to obtain a linear estimator of the form:

$$\hat{x}_k = \sum_{j=0}^k \alpha(j) y_j$$

where each $\alpha(j)$ is $n \times p$ matrix, such that the norm

$$E_x E_{I_k} \{ ||x_k - \hat{x}_k||^2 \} \quad (2.6)$$

is minimized. E denotes the expectation operator and the suffixes represent the variables with respect to which expectation is taken.

Substitution of $\hat{x}_k$ into the criterion (2.6) and differentiation with regard to elements of $\alpha(j)$ results in a total of $n p k$ equations, which can be expressed compactly as,

$$E_x E_{I_k} \{ (x_k - \hat{x}_k) y_j^T \} = 0, \quad j \leq k \quad (2.7)$$

Following Nahi, it is conjectured that the estimator can be put into the recursive form:

$$\hat{x}_k = F_1(k) \hat{x}_{k-1} + F_2(k) y_k + F_3(k) u_{k-1} \quad (2.8)$$

$$\hat{x}_0 = E_x(x_0)$$

This conjecture will be verified if expressions for $F_1(k)$, $F_2(k)$ and $F_3(k)$ can be found such that the conditions (2.7) are satisfied. To this end, decompose (2.7) into two parts:

$$E_x E_{I_k} \{ (x_k - \hat{x}_k) y_j^T \} = 0, j < k \quad (2.9)$$

$$E_x E_{I_k} \{ (x_k - \hat{x}_k) y_k^T \} = 0 \quad (2.10)$$

Substitution of (2.8) and (2.3) into (2.9) results in:

$$\begin{aligned} E_x E_{I_k} \{ [I - F_2(k)C(k)] A(k)x_{k-1} + [I - F_2(k)C(k)] B(k)U_{k-1} \\ + [I - F_2(k)C(k)] \xi(k-1) - F_1(k)\hat{x}_{k-1} - F_2(k)\eta(k) \\ - F_3(k)U_{k-1} \} y_j^T = 0, j < k \end{aligned} \quad (2.11)$$

Now, (2.7) implies:

$$E_x E_{I_k} \{ (x_{k-1} - \hat{x}_{k-1}) y_j^T \} = 0, j \leq k-1. \quad (2.12)$$

A fundamental result in probability theory (see Doob^(D4)) is that:

$$E_{x,y} [f(x,y)] = E_y \{ E_x [f(x,y) | y] \}$$

Using this result and (2.12) in (2.11) and taking expectations with regard to y_k first (notice that $\xi(k-1)$ and $\eta(k)$ are independent of y_j , $j \leq k-1$), yields:

$$\begin{aligned} E_{y_k} \{ E_x E_{I_{k-1}} \{ [A(k) - F_2(k)C(k)A(k) - F_1(k)] x_{k-1} y_j^T | y_k \} \\ + E_x E_{I_{k-1}} \{ [B(k) - F_2(k)C(k)B(k) - F_3(k)] U_{k-1} y_j^T | y_k \} \} \\ = 0, j < k \end{aligned} \quad (2.13)$$

The conditions (2.13), which are equivalent to (2.9), are seen to be satisfied by:

$$F_1(k) = E_{\gamma_k} [A(k) - F_2(k)C(k)A(k)] \quad (2.14)$$

$$F_3(k) = E_{\gamma_k} [B(k) - F_2(k)C(k)B(k)] \quad (2.15)$$

Similarly, substitution of (2.3) and (2.8) into (2.10) yields:

$$\begin{aligned} E_{\gamma_k} \{ E_x E_{I_{k-1}} [(I - F_2(k)C(k))A(k)x_{k-1} + (I - F_2(k)C(k))B(k)U_{k-1} \\ + (I - F_2(k)C(k))\xi(k-1) - F_1(k)\hat{x}_{k-1} - F_2(k)\eta(k) \\ - F_3(k)U_{k-1}] y_k^T | \gamma_k \} = 0 \end{aligned} \quad (2.16)$$

Define the following expectations:

$$V(k-1) = E_x E_{I_{k-1}} x_{k-1} x_{k-1}^T$$

$$P(k-1) = E_x E_{I_{k-1}} (x_{k-1} - \hat{x}_{k-1})(x_{k-1} - \hat{x}_{k-1})^T$$

$$T(k-1) = E_x E_{I_{k-1}} (x_{k-1} U_{k-1}^T)$$

$$M(k-1) = E_x E_{I_{k-1}} (U_{k-1} U_{k-1}^T)$$

$$(\bar{\cdot}) = E_{\gamma_k}(\bar{\cdot})$$

With these definitions and substitution for y_k in (2.16)

yields, after some algebraic manipulations,

$$\begin{aligned}
 F_2(k) = & E_{Y_k} \left[\{ (A(k) - \overline{A(k)}) (V(k-1)A(k)^T + T(k-1)B(k)^T) \right. \\
 & + (B(k) - \overline{B(k)}) (T(k-1)^T A(k)^T + M(k-1)B(k)^T) \\
 & + \overline{A(k)} P(k-1)A(k)^T + Q(k) \} C(k)^T \Big] \\
 & \cdot \left[E_{Y_k} \{ (C(k)A(k) - \overline{C(k)} \overline{A(k)}) (V(k-1)A(k)^T \right. \\
 & + T(k-1)B(k)^T) C(k)^T + (C(k)B(k) - \overline{C(k)} \overline{B(k)}) (T(k-1)^T A(k)^T \\
 & + M(k-1)B(k)^T) C(k)^T + \overline{C(k)} \overline{A(k)} P(k-1)A(k)^T C(k)^T \\
 & \left. \left. + C(k)Q(k)C(k)^T + R(k) \right\} \right]^{-1} \quad (2.17)
 \end{aligned}$$

It remains now to derive recursive relations for $P(k)$ and $V(k)$. Notice that $T(k-1)$ can be easily evaluated by extrapolating the expectation of the state if the input U is either deterministic or independent of other system variables.

By definition,

$$\begin{aligned}
 V(k) &= E_x E_{I_k} x_k x_k^T \\
 &= E_x E_{I_k} \left[A(k)x_{k-1} + B(k)U_{k-1} + \xi(k-1) \right] \\
 &\quad \left[A(k)x_{k-1} + B(k)U_{k-1} + \xi(k-1) \right]^T \\
 &= E_{Y_k} \left[A(k)V(k-1)A(k)^T + B(k)T(k-1)^T A(k)^T \right. \\
 &\quad \left. + A(k)T(k-1)B(k)^T + B(k)M(k-1)B(k)^T + Q(k) \right] \quad (2.18)
 \end{aligned}$$

$V(0) = E_x E_{I_0} x_0 x_0^T = P(0)$, the covariance of the initial state.

Similarly,

$$\begin{aligned} P(k) &= E_x E_{I_k} (x_k - \hat{x}_k)(x_k - \hat{x}_k)^T \\ &= E_x E_{I_k} (x_k - \hat{x}_k) x_k^T \end{aligned}$$

Substitution for x_k in this expression and some algebraic manipulation, yields:

$$\begin{aligned} P(k) &= E_{Y_k} \left[(A(k) - \overline{A(k)})(V(k-1)A(k)^T + T(k-1)B(k)^T) \right. \\ &\quad \left. + (B(k) - \overline{B(k)})(M(k-1)B(k)^T + T(k-1)^T A(k)^T) \right. \\ &\quad \left. + Q(k) + \overline{A(k)}P(k-1)A(k)^T \right] \\ &\quad - F_2(k)E_{Y_k} \left[(C(k)A(k) - \overline{C(k)A(k)})(V(k-1)A(k)^T \right. \\ &\quad \left. + T(k-1)B(k)^T) + (C(k)B(k) - \overline{C(k)B(k)})(M(k-1)B(k)^T \right. \\ &\quad \left. + T(k-1)^T A(k)^T) + C(k)Q(k) + \overline{C(k)A(k)}P(k-1)A(k)^T \right] \quad (2.19) \end{aligned}$$

Equations (2.8), (2.14), (2.15), (2.17)-(2.19) together define the optimal linear filter for the system (2.3).

If the system (2.3) were time invariant, i.e. say $Y_k = 1 \forall k$, the optimal linear filter transforms to the traditional Kalman filter:

$$\begin{aligned} \hat{x}_k &= F_1(k)\hat{x}_{k-1} + F_2(k)y_k + F_3(k)u_{k-1} \\ F_1(k) &= A_1 - F_2(k)C_1A_1 \end{aligned}$$

$$F_3(k) = B_1 - F_2(k) C_1 B_1$$

or

$$\hat{x}_k = (A_1 \hat{x}_{k-1} + B_1 U_{k-1}) + F_2(k) (y_k - C_1 A_1 \hat{x}_{k-1} - C_1 B_1 U_{k-1})$$

$$F_2(k) = (A_1 P(k-1) A_1^T + Q_1) C_1^T \left[C_1 \{A_1 P(k-1) A_1^T + Q_1\} C_1^T + R_1 \right]^{-1}$$

$$P(k) = \left[I - F_2(k) C_1 \right] \left[A_1 P(k-1) A_1^T + Q_1 \right]$$

Consider next the optimal non-linear estimator.

2.4 OPTIMAL NON-LINEAR ESTIMATOR

The optimal non-linear estimator for the case of switching environments (i.e. only the random processes ξ and η are switching) was derived by Ackerson and Fu^(A1). What follows is an exposition of the optimal estimator for (2.3).

The desired estimator is one which, based on the measurements $y_1, y_2, \dots, y_k$, gives an estimate $\hat{x}_{k|k}$ of the system state x_k , such that the criterion

$$E\{ \|x_k - \hat{x}_{k|k}\|^2 \}$$

is minimized. It is well known^(K2) that the desired estimate is given by:

$$\hat{x}_{k|k} = E(x_k | Y^k), \quad (2.20)$$

where

$$Y^k \triangleq \{y_1, y_2, \dots, y_k\}$$

Define another sequence I_k :

$$I_k = \{\gamma_1, \gamma_2, \dots, \gamma_k\} \quad (2.21)$$

and let $I_k^{(j)}$ be a specific sequence from the space Ω^k of the sequences I_k . Denote the estimate conditioned on the sequence $I_k^{(j)}$ as:

$$\hat{x}_k^{(j)} = E(x_k | Y^k, I_k^{(j)}), \quad j=1, 2, \dots, 2^k \quad (2.22)$$

Let

$$\begin{aligned} I_k^{(j)} &= \{I_{k-1}^{(\ell)}, \gamma_k = \rho\} \\ I_{k-1}^{(\ell)} &= \{I_{k-2}^{(1)}, \gamma_{k-1} = \sigma\} \end{aligned} \quad (2.23)$$

That is, $I_k^{(j)}$ is a sequence which contains $I_{k-1}^{(\ell)}$ as a subsequence and $I_{k-1}^{(\ell)}$ is a sequence that contains $I_{k-2}^{(1)}$ as a subsequence. Given $I_k^{(j)}$, the system is a linear system with known parameters and the state estimates $\hat{x}_k^{(j)}$ are exactly those provided by a Kalman filter ^(A1):

$$\begin{aligned} \hat{x}_{k|k-1}^{(j)} &= [\rho A_1 + (1-\rho)A_o] \hat{x}_{k-1|k-1}^{(\ell)} + [\rho B_1 + (1-\rho)B_o] U_{k-1} \\ P_{k|k-1}^{(j)} &= [\rho A_1 + (1-\rho)A_o] P_{k-1|k-1}^{(\ell)} [\rho A_1 + (1-\rho)A_o]^T + \rho Q_1 + (1-\rho)Q_o \\ G_k^{(j)} &= P_{k|k-1}^{(j)} [\rho C_1 + (1-\rho)C_o]^T \{ [\rho C_1 + (1-\rho)C_o] P_{k|k-1}^{(j)} [\rho C_1 + (1-\rho)C_o]^T \\ &\quad + \rho R_1 + (1-\rho)R_o \}^{-1} \\ P_{k|k}^{(j)} &= P_{k|k-1}^{(j)} - G_k^{(j)} [\rho C_1 + (1-\rho)C_o] P_{k|k-1}^{(j)} \end{aligned}$$

$$\hat{x}_{k|k}^{(j)} = \hat{x}_{k|k-1}^{(j)} + G_k^{(j)} \left[y_k - \{ \rho C_1 + (1-\rho) C_0 \} \hat{x}_{k|k-1}^{(j)} \right] \quad (2.24)$$

Using the smoothing property of conditional expectations (see Doob^(D4)), the state estimates are given as:

$$\begin{aligned} \hat{x}_{k|k} &\triangleq E(x_k | Y^k) \\ &= \sum_{j=1}^{2^k} E(x_k | Y^k, I_k^{(j)}) P(I_k^{(j)} | Y^k) \end{aligned} \quad (2.25)$$

The conditional probability $P(I_k^{(j)} | Y^k)$ is evaluated as follows:

By Bayes' rule,

$$P(I_k^{(j)} | Y^k) = \frac{P(y_k | I_k^{(j)}, Y^{k-1}) P(I_k^{(j)} | Y^{k-1})}{\sum_{m=1}^{2^k} P(y_k | I_k^{(m)}, Y^{k-1}) P(I_k^{(m)} | Y^{k-1})} \quad (2.26)$$

Now, given the transition probabilities q_{ij} , $i, j=0,1$ (see equation 2.4),

$$P(I_k^{(j)} | Y^{k-1}) = q_{\rho\sigma} P(I_{k-1}^{(\ell)} | Y^{k-1}). \quad (2.27)$$

The densities $P(y_k | I_k^{(i)}, Y^{k-1})$ are Gaussian (under the assumption of Gaussian initial state x_0 and additive Gaussian noises), with mean and covariance given by:

$$E(y_k | I_k^{(j)}, Y^{k-1}) = [\rho C_1 + (1-\rho) C_0] \hat{x}_{k|k-1}^{(\ell)}$$

and

$$\begin{aligned} \text{Cov}(y_k | I_k^{(j)}, Y^{k-1}) = & \left[\rho C_1 + (1-\rho) C_o \right] P_{k|k-1}^{(2)} \left[\rho C_1 + (1-\rho) C_o^T \right] \\ & + \rho R_1 + (1-\rho) R_o \end{aligned} \quad (2.28)$$

Thus, the optimal non-linear estimator is specified by (2.24), (2.25) and (2.26). The structure of this estimator is shown in Fig. 2.1.

It is obvious from (2.25), however, that the minimum mean squared error estimator requires an evergrowing amount of memory. Specifically, at the k^{th} step, 2^k covariance matrices $P_{k|k}^{(i)}$ and 2^k state estimates $\hat{x}_{k|k}^{(i)}$ need to be stored. The calculation of the a-posteriori probability densities $P(I_{k+1}^{(j)} | Y^{k+1})$ will also require knowledge of 2^k probabilities $P(I_k^{(i)} | Y^k)$. Thus, the estimator presented in this section, though optimal, is not a very realistic approach because of the infinite memory required for the control of a continuing process.

This observation suggests the necessity for developing estimators (obviously sub-optimal) which are physically realizable while at the same time provide "good" estimates of the states. In the next section, the various approximations to the optimal non-linear estimator derived here are discussed.

2.5 SUB-OPTIMAL NON-LINEAR ESTIMATORS

It is readily seen from equation (2.25) that the density $P(x_k|Y^k)$ is not Gaussian, but rather the sum of 2^k Gaussian densities. Consider the assumption that the conditional density of the state $P(x_{k-1}|Y^{k-1})$ is Gaussian. This assumption is invoked by Ackerson and Fu^(A1) in deriving a sub-optimal estimator. In this case, $P(y_k|Y^{k-1}, \gamma_k=i)$, $i=0,1$ are both Gaussian and (2.26) is changed to:

$$P(\gamma_k=i|Y^k) = \frac{P(y_k|Y^{k-1}, \gamma_k=i)P(\gamma_k=i|Y^{k-1})}{\sum_{j=0}^1 P(y_k|Y^{k-1}, \gamma_k=j)P(\gamma_k=j|Y^{k-1})} \quad (2.29)$$

The new estimate $\hat{x}_{k|k}$ and the new covariance matrix $P_{k|k}$ are calculated from:

$$\hat{x}_{k|k} = \sum_{j=0}^1 P(\gamma_k=j|Y^k) \hat{x}_{k|k}^{(j)} \quad (2.30)$$

$$P_{k|k} = \sum_{j=0}^1 P(\gamma_k=j|Y^k) P_{k|k}^{(j)} \quad (2.31)$$

$\hat{x}_{k|k}^{(j)}$ and $P_{k|k}^{(j)}$ are given by (2.24) with $\hat{x}_{k-1|k-1}^{(l)}$ and $P_{k-1|k-1}^{(l)}$ replaced by $\hat{x}_{k-1|k-1}$ and $P_{k-1|k-1}$, respectively.

The estimator (2.30)-(2.31) requires the evaluation of only two probabilities $P(\gamma_k=j|Y^k)$, $j=0,1$ and the two

state estimates and covariance matrices $\hat{x}_{k|k}^{(j)}$ and $P_{k|k}^{(j)}$, $j=0,1$. This represents a large saving in the computational effort required to calculate the state estimates.

A slightly different assumption was introduced by Bruckner and Scott^(B1) in attempting to extend the results of Ackerson and Fu^(A1) to a general switching system. They assumed that the density of the state conditioned on the observations and the mode (i.e. value of γ) is Gaussian, i.e.

$$P(x_k | Y^k, \gamma_k = i) = N(\hat{x}_{k|k}^{(i)}, P_{k|k}^{(i)}), \quad i=0,1 \quad (2.32)$$

Under this assumption,

$$\begin{aligned} P(x_{k+1} | Y^k, \gamma_k = i, \gamma_{k+1} = j) \\ = N(A_j \hat{x}_{k|k}^{(i)} + B_j U_k, A_j P_{k|k}^{(i)} A_j^T + Q_j) \end{aligned} \quad (2.33)$$

and $P(y_{k+1} | Y^k, \gamma_k = i, \gamma_{k+1} = j)$

$$= N(C_j A_j \hat{x}_{k|k}^{(i)} + C_j B_j U_k, C_j A_j P_{k|k}^{(i)} A_j^T C_j^T + C_j Q_j C_j^T + R_j) \quad (2.34)$$

Equations (2.32), (2.33) and (2.34) lead to the following estimator :

$$\begin{aligned} \hat{x}_{k+1|k+1} &= \frac{1}{\sum_{j=0}^1 P(\gamma_{k+1}=j | Y^{k+1})} \hat{x}_{k+1|k+1}^{(j)} \\ \hat{x}_{k+1|k+1}^{(j)} &= \frac{\sum_{i=0}^1 P(\gamma_k=i, \gamma_{k+1}=j | Y^{k+1}) \hat{x}_{k+1|k+1}^{(i,j)}}{P(\gamma_{k+1}=j | Y^{k+1})} \end{aligned}$$

$$x_{k+1}^{(j)} = \frac{\sum_{i=0}^1 P(\gamma_k=i, \gamma_{k+1}=j | Y^{k+1}) (\bar{P}_{ji} + \tilde{x}_{ji} \tilde{x}_{ji}^T)}{P(\gamma_{k+1}=j | Y^{k+1})}$$

$$\tilde{x}_{ji} = A_j \hat{x}_{k|k}^{(i)} + G_{ji} [y_{k+1} - C_j A_j \hat{x}_{k|k}^{(i)} - C_j B_j u_k]$$

$$G_{ji} = H_{ji} C_j^T [C_j H_{ji} C_j^T + R_j]^{-1}$$

$$H_{ji} = A_j P_{k|k}^{(i)} A_j^T + Q_j$$

$$\bar{P}_{ji} = H_{ji} - G_{ji} C_j H_{ji}$$

$$P_{k+1|k+1}^{(j)} = X_{k+1}^{(j)} - \hat{x}_{k+1|k+1}^{(j)} \hat{x}_{k+1|k+1}^{(j)T} \quad (2.35)$$

The probabilities $P(\gamma_{k+1}=j, \gamma_k=i | Y^{k+1})$ are given as

$$\begin{aligned} & P(\gamma_{k+1}=j, \gamma_k=i | Y^{k+1}) \\ &= \frac{q_{ji} P(y_{k+1} | Y^k, \gamma_k=i, \gamma_{k+1}=j) P(\gamma_k=i | Y^k)}{\sum_{m=0}^1 \sum_{n=0}^1 q_{nm} P(y_{k+1} | Y^k, \gamma_k=m, \gamma_{k+1}=n) P(\gamma_k=n | Y^k)} \end{aligned} \quad (2.36)$$

and

$$P(\gamma_{k+1}=j | Y^{k+1}) = \sum_{i=0}^1 P(\gamma_{k+1}=j, \gamma_k=i | Y^{k+1}) \quad (2.37)$$

Bruckner and Scott^(B1) have showed, via simulated examples, that this estimator performs better than the Ackerson and Fu^(A1) estimator. The reason for this, they suggest, is that (2.32) approximates the tails of the probability density of the state more closely than does the Ackerson and Fu approximation.

Sawaragi et al^(S2), in presenting the non-linear estimator for the problem of missing or interrupted observations (i.e. only the matrix C switches between C_1 and ϕ , the null matrix), utilized the assumption:

$$P(x_k | Y^{k-1}) = N(\hat{x}_k |_{k-1}, P_k |_{k-1}) \quad (2.38)$$

The estimator developed under this assumption is similar in structure to the one developed by Ackerson and Fu. The estimator was shown to perform better than the optimal linear estimator developed by Nahi^(N1) for the same problem (i.e. missing observations). The reason for this is seen to be in the difference between the expressions for the estimation error covariances in the linear and the non-linear estimators. While in the linear estimator the error covariance depends only on the a-priori probability density of γ , the non-linear estimator results in an error covariance which is a function of the a-posteriori density of γ , viz. $P(\gamma_k | Y^k)$.

All the sub-optimal estimators discussed in this section can be seen to have the general form:

$$\hat{x}_{k|k} = \sum_{i=0}^1 x_{k|k}^{(i)} P(\gamma_k=i|Y^k) \quad (2.39)$$

$$P_{k|k} = \sum_{i=0}^1 P_{k|k}^{(i)} P(\gamma_k=i|Y^k) \quad (2.40)$$

where the estimate $\hat{x}_{k|k}^{(i)}$ and the corresponding error covariance $P_{k|k}^{(i)}$ are provided by a Kalman filter. A schematic diagram of the estimator (2.39)-(2.40) is given in Fig. 2.2.

In conclusion, it is noticed that the estimator (2.39) requires the evaluation of two Kalman filters at each stage k , and, for systems of large dimensions, this may be a prohibitive requirement. In the next Chapter, a new approach is presented with a view to reduce the computational effort still further.

FIG. 2.1. Structure of the optimal estimator for a switching system with two modes.

FIG. 2.2. Structure of the approximate estimators for a switching system with two modes.

CHAPTER 3

MODELLING THE SWITCHING PROCESS

3.1 PRELIMINARY REMARKS

This Chapter presents a new approach for the determination of the state estimates, with the view of reducing the computational requirements of the sub-optimal state estimators still further. Briefly, the approach is based on posing γ_k as an additional state by developing a model describing its evolution in time, and estimating the augmented state vector in the resulting non-linear system via an extended Kalman filter^(J2). The prime advantage of this approach, as discussed in the sequel, is that it requires the evaluation of only one (extended) Kalman filter at each stage. In this Chapter, the development of the model for the switching process and the structure of the model for some typical switching systems are presented.

3.2 THE MODEL

It is desired to add the switching variable γ to the state vector x in a model of the following form:

$$\begin{aligned} x_k = & [A_1\gamma_k + A_0(1-\gamma_k)] x_{k-1} + [B_1\gamma_k + B_0(1-\gamma_k)] u_{k-1} \\ & + \gamma_k \xi_1(k-1) + (1-\gamma_k) \xi_0(k-1) \end{aligned}$$

$$\gamma_k = f(\gamma_{k-1}, x_{k-1}) + \psi_k$$

$$y_k = [C_1 \gamma_k + C_0 (1 - \gamma_k)] x_k + \gamma_k \eta_1(k) + (1 - \gamma_k) \eta_0(k) \quad (3.1)$$

γ_k is estimated recursively along with x_k by a single extended Kalman filter. The function $f(\gamma_{k-1}, x_{k-1})$ is a dynamic model to account approximately for the evolution of the switching variable γ over time. Looking ahead briefly, the function f is a smooth function mapping the real line into the continuous range $(0,1)$ (just as the expectation of γ does).

This section develops a procedure for finding the model function f . For notational convenience, define the following terms:

$\hat{\gamma}_k|_{k-1}$ Δ value of γ_k predicted by the model

$\hat{\gamma}_k|_k$ Δ filtered estimate of γ_k obtained from an extended Kalman filter

$\gamma_k^*|_{k-1}$ Δ optimal prediction of $\gamma_k = P(\gamma_k = 1 | Y^{k-1})$

$\gamma_k^*|_k$ Δ optimal estimate of $\gamma_k = P(\gamma_k = 1 | Y^k)$

$\gamma_k^{**}|_{k-1}$ Δ sub-optimal prediction of γ_k (from Bank of Kalman filters)

$\gamma_k^{**}|_k$ Δ sub-optimal estimate of γ_k .

Consider the following Kalman filter equations which calculate the estimate $\hat{\gamma}_{k|k}$ at time t :

$$\hat{\gamma}_{k|k-1} = f(\hat{\gamma}_{k-1|k-1}, \hat{x}_{k-1|k-1}) \quad (3.2)$$

$$\hat{\gamma}_{k|k} = \hat{\gamma}_{k|k-1} + G_k \left[y_k - (C_1 \hat{\gamma}_{k|k-1} + C_0 (1 - \hat{\gamma}_{k|k-1})) \hat{x}_{k|k-1} \right] \quad (3.3)$$

where

$$\begin{aligned} \hat{x}_{k|k-1} = & \left[A_1 \hat{\gamma}_{k|k-1} + A_0 (1 - \hat{\gamma}_{k|k-1}) \right] \hat{x}_{k-1|k-1} \\ & + \left[B_1 \hat{\gamma}_{k|k-1} + B_0 (1 - \hat{\gamma}_{k|k-1}) \right] U_{k-1} \end{aligned} \quad (3.4)$$

The function f will be chosen so as to make $\hat{\gamma}_{k|k-1}$ a good prediction of γ_k , and also to make it compatible with the probable evolution of x and y .

A linear model f is obtained if $\hat{\gamma}_{k|k-1}$ is based on the optimal prediction $\gamma_{k|k-1}^*$ (which minimizes the mean squared error, using observations up to time $k-1$),

$$\begin{aligned} \gamma_{k|k-1}^* &= E(\gamma_k | Y^{k-1}) = P(\gamma_k = 1 | Y^{k-1}) \\ &= q_{11} \gamma_{k-1|k-1}^* + q_{10} (1 - \gamma_{k-1|k-1}^*) \end{aligned} \quad (3.5)$$

where $\gamma_{k-1|k-1}^*$ is the optimal estimate of γ_{k-1} and q_{ij} , $i, j=0,1$ are the transition probabilities of γ .

The corresponding model in the form (3.1) is:

$$\gamma_k = f(\gamma_{k-1}, x_{k-1}) = (q_{11} - q_{10}) \gamma_{k-1} + q_{10} \quad (3.6)$$

The performance of such a linear model is, however, quite inferior except in cases with large measurement noise (large R). Moreover, this linear model is independent of other system variables.

A better estimate, $\gamma_{k|k}^*$, would result if observations up to time k were available:

$$\begin{aligned}\gamma_{k|k}^* &= E(\gamma_k | Y^k) = P(\gamma_k=1 | Y^k) \\ &= \frac{P(y_k | Y^{k-1}, \gamma_k=1)P(\gamma_k=1 | Y^{k-1})}{\sum_{i=0}^1 P(y_k | Y^{k-1}, \gamma_k=i)P(\gamma_k=i | Y^{k-1})}\end{aligned}\quad (3.7)$$

where Bayes' rule is used to obtain (3.7) and

$$P(\gamma_k=1 | Y^{k-1}) = \gamma_{k|k-1}^*$$

and

$$P(\gamma_k=0 | Y^{k-1}) = 1 - \gamma_{k|k-1}^*$$

are available from (3.5).

The conditional probability density functions for y_k , $P(y_k | Y^{k-1}, \gamma_k=i)$, $i=0,1$, have the same growing memory requirement as cited in the previous Chapter and must be approximated if long observation sequences are to be handled. Denoting these approximate densities as $P^*(y_k | Y^{k-1}, \gamma_k=i)$, and similarly $P^*(\gamma_k=i | Y^{k-1})$, the new estimate (necessarily sub-optimal) of γ_k , $\gamma_{k|k}^{**}$, is given as:

$$\gamma_{k|k-1}^{**} = (q_{11} - q_{10}) \gamma_{k-1|k-1}^{**} + q_{10} \quad (3.8)$$

$$\gamma_{k|k}^{**} = \frac{P^*(y_k | Y^{k-1}, \gamma_{k-1}) P^*(\gamma_{k-1} | Y^{k-1})}{\sum_{i=0}^1 P^*(y_k | Y^{k-1}, \gamma_{k-1}) P^*(\gamma_{k-1} | Y^{k-1})} \quad (3.9)$$

where

$$\gamma_{k-1|k-1}^{**} = P^*(\gamma_{k-1} = 1 | Y^{k-1}) \quad (3.10)$$

The model (3.1) is based only on prior data, so an extrapolated value, y_k^* , is substituted for y_k in (3.9) as follows:

$$y_k^* = E^*(y_k | Y^{k-1})$$

where E^* denotes the expectation with respect to the approximate probability density $P^*(y_k | Y^{k-1})$.

Thus,

$$y_k^* = \sum_{i=0}^1 E^*(y_k | Y^{k-1}, \gamma_{k-1}) P^*(\gamma_{k-1} | Y^{k-1}) \quad (3.11)$$

The model function f in the form (3.2) is then:

$$\bar{\gamma}_k = (q_{11} - q_{10}) \hat{\gamma}_{k-1|k-1} + q_{10} \quad (3.12)$$

$$\begin{aligned} \hat{\gamma}_{k|k-1} &= f(\hat{\gamma}_{k-1|k-1}, \hat{x}_{k-1|k-1}) \\ &= \left[1 + \frac{1 - \bar{\gamma}_k}{\bar{\gamma}_k} \frac{P^*(y_k | Y^{k-1}, \gamma_{k-1} = 0)}{P^*(y_k | Y^{k-1}, \gamma_{k-1} = 1)} \right]^{-1} \bigg|_{y_k = y_k^*} \end{aligned} \quad (3.13)$$

The model function f in (3.13) depends on the dynamics of the whole system, not only on the probabilities of the switching process. Notice that (3.12) represents the linear model based on the prediction $\gamma_k^{**}|_{k-1}$ (Equation 3.8). If the approximate densities $P^*(y_k|Y^{k-1}, \gamma_k=i)$, $i=0,1$, are assumed Gaussian, then, for a large observation noise covariance (large R), the model (3.13) reduces to the linear model (3.12). Thus, the linear model is a special case of the model (3.13).

Having specified the model function f (equation 3.13), it now remains to determine the statistics of the noise process $\{\psi_k\}$, added to the model in (3.1). This noise term is necessary to prevent the divergence of the Kalman filter resulting from modelling errors^(F2). Choice of the statistics of ψ_k becomes simpler if the modelling errors are assumed to have a zero-mean, normal distribution. In this case, the variance Q_ψ of the noise ψ_k completely specifies its distribution.

Consider the case of a symmetric transition probability matrix, i.e. $q_{ij}=q_{ji}$, $i,j=0,1$. The transition probability q_{ij} , $i \neq j$, is referred to in the sequel as the "switching probability". In this case,

$$\lim_{k \rightarrow \infty} P(\gamma_k=i) = 0.5, \quad i=0,1 \quad (3.14)$$

Also

$$\begin{aligned}
 \lim_{k \rightarrow \infty} \text{var}(\gamma_k) &= \lim_{k \rightarrow \infty} \{E(\gamma_k^2) - E(\gamma_k)^2\} \\
 &= \lim_{k \rightarrow \infty} \left\{ \left[\sum_{i=0}^1 i^2 P(\gamma_k=i) \right] - \left[\sum_{i=0}^1 i P(\gamma_k=i) \right]^2 \right\} \\
 &= 0.25
 \end{aligned} \tag{3.15}$$

Considering the modelling error per se, the maximum value of the mean squared modelling error,

$$E(\gamma_k - f(\gamma_{k-1}, x_{k-1}))^2$$

is seen to be 1, occurring when $f(\gamma_{k-1}, x_{k-1}) \approx 1 - \gamma_k$.

Thus,

$$0 \leq Q_\psi \leq 1 \tag{3.16}$$

From (3.15), an intuitive choice of Q_ψ would be $Q_\psi = 0.25$. This choice was, however, found insufficient to 'bury' the modelling errors in systems with large random disturbances. In order to get more insight into the choice of Q_ψ , a Monte-Carlo simulation study was undertaken for a simple scalar system with missing observations. Details of this study are given in Appendix B: The "best" Q_ψ , in terms of resulting in minimum mean squared estimation error, was observed to be a function of system parameters,

not only on the statistics of γ . Q_ψ was observed to be close to 0.25 for a slow system with small random influences and approaching unity as the noise to signal ratio increased. For the scalar system considered in Appendix II, an approximate relation between Q_ψ and the system parameters is given as:

$$\begin{aligned}
 Q_\psi &= .25\exp(-\ln(a)/2q_{10}) + 0.5\sin(-2\pi\ln(a)Q/x^2)\exp(-15R/x^2), \\
 &\quad \text{if } \exp(-.5x^2/Q) < a < 1 \\
 &= .25\exp(-\ln(a)/2q_{10}) \quad , \quad \text{otherwise}
 \end{aligned} \tag{3.17}$$

'a' is the pole of the system transfer function and q_{10} is the switching probability. NSR_p and NSR_o are the noise to signal ratios at the plant and at the output respectively, where, by definition,

$$NSR = \frac{\text{Mean Squared Value of Noise}}{\text{Mean Squared Value of Signal}}$$

For systems of higher order described by several state and measurement equations and hence several eigenvalues and noise to signal ratios, a worst case design for Q_ψ could be based on choosing the most unfavourable values for NSR_p , NSR_o , and a . In other words, one chooses the first dominant eigenvalue for a and the largest of the noise to signal ratios at the plant and at the output for NSR_p and NSR_o respectively.

The model for γ is thus completely specified by (3.13) and (3.17). Consider an example.

Example 3.1. Detection Problem

It is desired to estimate the switching variable γ_k from noisy observations y_k , where γ_k may switch with probability q_{10} :

$$y_k = \gamma_k + v_k$$

$$v_k \sim N(0, R)$$

and $\gamma_k = 0$ or 1 , with $P(\gamma_0 = 1) = q$ and $P(\gamma_k \neq \gamma_{k-1}) = q_{10}$.

In this case,

$$P(y_k | \gamma^{k-1}, \gamma_k = i) = N(i, R), \quad i = 0, 1.$$

This leads to the optimal (m.m.s.e.) estimate, $\gamma_{k|k}^*$, of γ_k (see equation 3.7)

$$\begin{aligned} \gamma_{k|k}^* &= \frac{\frac{1}{\sqrt{2\pi R}} e^{-(y_k - 1)^2 / 2R} \gamma_{k|k-1}^*}{\gamma_{k|k-1}^* \frac{1}{\sqrt{2\pi R}} e^{-(y_k - 1)^2 / 2R} + (1 - \gamma_{k|k-1}^*) \frac{1}{\sqrt{2\pi R}} e^{-y_k^2 / 2R}} \\ &= \left[1 + \frac{1 - \gamma_{k|k-1}^*}{\gamma_{k|k-1}^*} \exp\left(\frac{1 - 2y_k}{2R}\right) \right]^{-1} \end{aligned} \quad (3.20)$$

where

$$\gamma_{k|k-1}^* = q_{10} + (1 - 2q_{10}) \gamma_{k-1|k-1}^* \quad (3.21)$$

From (3.11),

$$\begin{aligned}
 y_k^* &= \sum_{i=0}^1 E(y_k | Y^{k-1}, \gamma_k=i) P(\gamma_k=i | Y^{k-1}) \\
 &= P(\gamma_k=1 | Y^{k-1}) \\
 &= \gamma_{k|k-1}^*
 \end{aligned} \tag{3.22}$$

The model function f is now given as:

$$\begin{aligned}
 \bar{\gamma}_k &= q_{10} + (1-2q_{10}) \hat{\gamma}_{k-1|k-1} \\
 \hat{\gamma}_{k|k-1} &= f(\hat{\gamma}_{k-1|k-1}) \\
 &= \left[1 + \frac{1-\bar{\gamma}_k}{\bar{\gamma}_k} \exp\left(\frac{1-2\bar{\gamma}_k}{2R}\right) \right]^{-1}
 \end{aligned} \tag{3.23}$$

Fig. 3.1 shows the shape of the function f . It is smoother and more rounded in a system with large random influences and more like a switching characteristic in a system with a large signal to noise ratio (small R). It is also seen that as the noise level increases the model approaches the linear model (equation 3.12).

The system (3.18) was simulated with $q_{10}=0.1$ and with two values of R : $R=0.01$ and $R=0.1$. It is noticed immediately that equation (3.17) can not be used here for the determination of Q_v as there are no plant dynamics

involved. However, two other approaches were attempted:

- (i) Q_ψ is a constant (0.1 and 0.25 were used in the simulation).
- (ii) Q_ψ chosen adaptively by an examination of the residuals π_k , defined as:

$$\pi_k \triangleq y_k - \hat{y}_{k|k-1}$$

$$E(\pi_k) = 0$$

$$\text{cov}(\pi_k) = P_{k|k-1} + R$$

The approach taken in (ii) is due to Quigley^(Q1). His algorithm is given here briefly (details in Chapter 7).

1. Choose two levels of Q_ψ , for example, $Q_1=0.9$ and $Q_2=0.1$.
2. Assume that, at stage $k-1$, Q_ψ was chosen as Q_1 .
3. At stage k , assume $Q_\psi=Q_1$ and evaluate π_k and $\text{cov}(\pi_k)$.
4. If $\pi_k^2 \geq 9(P_{k|k-1} + R)$, $Q_\psi=Q_1$.
5. If $\pi_k^2 \leq P_{k|k-1} + R$, $Q_\psi=Q_2$.
6. Otherwise, Q_ψ remains unchanged.
7. Use the chosen value of Q_ψ to recalculate $P_{k|k-1}$.

The mean and the mean squared errors (computed by averaging the errors in 50 Monte-Carlo runs, each over 20 time

steps) in the estimation of γ_k , with Q_ψ chosen according to the two approaches cited earlier, are compared in Fig. 3.2 with those obtained by (3.20), the optimal estimator.

It is seen that the estimation errors with the modelling approach, with Q_ψ chosen according to either one of the two approaches cited earlier, are within acceptable limits for both values of the observation noise covariance, R , i.e. within $\sqrt{\pm R}$. However, the errors are large at the occurrence of a switch in γ . This follows from the inability of the filter to track the switch. The tracking capability of the filter can be improved by increasing the Kalman gain or the model noise covariance Q_ψ . It is thus seen that $Q_\psi = 0.25$ results in smaller transients in the estimation error than $Q_\psi = 0.1$. However, the larger Q_ψ results in increased errors when γ does not switch. Obviously, an adaptive scheme which results in small Q_ψ , say $Q_\psi = .1$, when γ does not switch and a large Q_ψ , say $Q_\psi = .9$, at the occurrence of a switch in γ is expected to perform better than a constant value of Q_ψ . This feature is provided by Quigley's^(Q1) approach. It is seen from Fig. 3.2 that this adaptive approach provides estimates close to the optimal estimates.

Having established the plausibility of using a model for the estimation of γ , consider next the models for some typical switching systems.

3.3 MODELS FOR SOME SPECIAL CASES

The models obtained in this section are based on the assumption of Ackerson and Fu^(A1) that the density of the state conditioned on the current and the past observations is Gaussian, i.e.:

$$P^*(x_{k-1}|Y^{k-1}) = N(\hat{x}_{k-1|k-1}, P_{k-1|k-1})$$

Then, from (3.1), for $i=0,1$,

$$\begin{aligned} E^*(x_k|Y^{k-1}, \gamma_k=i) &\triangleq \hat{x}_{k|k-1}^{(i)} \\ &= A_i \hat{x}_{k-1|k-1} + B_i U_{k-1} \end{aligned} \quad (3.24)$$

and

$$\begin{aligned} \text{cov}(x_k|Y^{k-1}, \gamma_k=i) &\triangleq P_{k|k-1}^{(i)} \\ &= A_i P_{k-1|k-1} A_i^T + Q_i \end{aligned} \quad (3.25)$$

Similarly,

$$E^*(y_k|Y^{k-1}, \gamma_k=i) = C_i \hat{x}_{k|k-1}^{(i)} \quad (3.26)$$

$$\text{cov}(y_k|Y^{k-1}, \gamma_k=i) = C_i P_{k|k-1}^{(i)} C_i^T + R_i \quad (3.27)$$

The extrapolated value of y_k , y_k^* is now given as:

$$\begin{aligned} y_k^* &= E^*(y_k | Y^{k-1}) \\ &= C_1 \hat{x}_{k|k-1}^{(1)} \bar{\gamma}_k + C_0 \hat{x}_{k|k-1}^{(0)} (1 - \bar{\gamma}_k) \end{aligned} \quad (3.28)$$

where

$$\bar{\gamma}_k = q_{10} + (1 - 2q_{10}) \bar{\gamma}_{k-1|k-1} \quad (3.29)$$

Substitution of (3.24)-(3.29) in (3.13) gives the model function f as:

$$\begin{aligned} \hat{y}_{k|k-1} &= \left[1 + \frac{1 - \bar{\gamma}_k}{\bar{\gamma}_k} \sqrt{\frac{|C_1 P_{k|k-1}^{(1)} C_1^T + R_1|}{|C_0 P_{k|k-1}^{(0)} C_0^T + R_0|}} \right. \\ &\quad \cdot \exp \left\{ -\frac{1}{2} \| y_k^* - C_0 \hat{x}_{k|k-1}^{(0)} \|^2 [C_0 P_{k|k-1}^{(0)} C_0^T + R_0]^{-1} \right. \\ &\quad \left. \left. + \frac{1}{2} \| y_k^* - C_1 \hat{x}_{k|k-1}^{(1)} \|^2 [C_1 P_{k|k-1}^{(1)} C_1^T + R_1]^{-1} \right\} \right] \quad (3.30) \end{aligned}$$

Consider the following special cases:

Case 1. Missing Observations

The system (3.1) can be transformed to one with missing observations by choosing $A_0 = A_1 = A$, $B_0 = B_1 = B$, $Q_0 = Q_1 = Q$, $R_0 = R_1 = R$, $C_0 = \phi$, the null matrix and $C_1 = C$. The model (3.30) now becomes:

$$\hat{Y}_{k|k-1} = \left[1 + \frac{1-\bar{\gamma}_k}{\bar{\gamma}_k} \sqrt{\frac{|C P_{k|k-1} C^T + R|}{|R|}} \exp \left\{ \frac{1}{2} \|C \hat{x}_{k|k-1}\|^2 \right. \right. \\ \left. \left. \left[(C P_{k|k-1} C^T + R)^{-1} (\bar{\gamma}_k - 1)^2 - \bar{\gamma}_k^2 R^{-1} \right] \right\} \right]^{-1} \quad (3.31)$$

It follows from (3.31) that, when the estimation error covariance $P_{k|k-1}$ becomes small such that $C P_{k|k-1} C^T \ll R$, the model can be simplified as:

$$\hat{Y}_{k|k-1} \approx \left[1 + \frac{1-\bar{\gamma}_k}{\bar{\gamma}_k} \exp \left\{ \left(\frac{1-2\bar{\gamma}_k}{2} \right) \|C \hat{x}_{k|k-1}\|^2 R^{-1} \right\} \right]^{-1} \quad (3.32)$$

The shape of the model function (3.32) is very similar to Fig. 3.1, i.e., for small R , the model approaches a switching characteristic while, for large random influences, the model reduces to the ~~linear~~ model (3.29).

Case 2. Noise Bursts

With $A_0 = A_1 = A$, $B_0 = B_1 = B$ and $C_0 = C_1 = C$, the system (3.1) is identical to the case of switching environments of noise bursts considered by Ackerson and Fu^(A1). The model, in this case, is seen to be:

$$\hat{Y}_{k|k-1} = \left[1 + \frac{1-\bar{\gamma}_k}{\bar{\gamma}_k} \sqrt{\frac{|C A P_{k-1|k-1} A^T C^T + C Q_1 C^T + R_1|}{|C A P_{k-1|k-1} A^T C^T + C Q_0 C^T + R_0|}} \right]^{-1} \quad (3.33)$$

A special case of interest here is that of a switching plant noise, i.e. the case above with $R_1 = R_0 = R$. (This case

is considered in Chapter 7 for limiting the divergence of the Kalman filter.) In this case, for a very small estimation error covariance, i.e. $P_{k-1|k-1} \ll \min(Q_0, Q_1)$, the model (3.33) simplifies to:

$$\hat{y}_{k|k-1} = \left[1 + \frac{1-\bar{\gamma}_k}{\bar{\gamma}_k} \sqrt{\frac{|CQ_1C^T+R|}{|CQ_0C^T+R|}} \right]^{-1}$$

which is seen to predict the smaller of the two plant noise covariances. The shape of the model function for a scalar system with $C=1$, $q_{10}=0.1$, and $R \ll \min(Q_0, Q_1)$ is shown in Fig. 3.3 for different values of the ratio Q_1/Q_0 .

Case 3. System Identification - Difference Equation Model

This case is of interest in system identification and is discussed in detail in Chapter 5. Briefly, the following model is considered; for a scalar output y_k ,

$$y_k = \gamma_k m_k \theta^{(1)} + (1-\gamma_k) m_k \theta^{(0)} + e_k \quad (3.34)$$

where

m_k is the row vector of past inputs and outputs, i.e.

$$m_k = [y_{k-1}, y_{k-2}, \dots, y_{k-n}, u_{k-1}, \dots, u_{k-n}]$$

and $\theta^{(1)}$ and $\theta^{(0)}$ are vectors of unknown parameters and are, in general, functions of γ_j , $k-1 \leq j \leq k-n$. e_k is assumed

to have a zero-mean normal distribution with a covariance R . In this case,

$$P(y_k | \gamma_k = 1, m_k, \theta^{(1)}, \theta^{(0)}) = N(m_k(1\theta^{(1)} + (1-1)\theta^{(0)}, R)$$

and

$$\begin{aligned} y_k^* &= E^*(y_k | Y^{k-1}) \\ &= m_k \hat{\theta}_{k-1}^{(1)} \bar{\gamma}_k + m_k \hat{\theta}_{k-1}^{(0)} (1 - \bar{\gamma}_k) \end{aligned} \quad (3.35)$$

where $\hat{\theta}_{k-1}^{(1)}$ and $\hat{\theta}_{k-1}^{(0)}$ are respectively the estimates of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(0)}$ based on observations up to time $k-1$ and

$$\bar{\gamma}_k = q_{10} + (1 - 2q_{10}) \hat{\gamma}_{k-1|k-1}$$

The model is thus given as:

$$\hat{\gamma}_{k|k-1} = \left[1 + \frac{1 - \bar{\gamma}_k}{\bar{\gamma}_k} \exp \left\{ \frac{1}{2R} (1 - 2\bar{\gamma}_k) \left[m_k (\hat{\theta}_{k-1}^{(0)} - \hat{\theta}_{k-1}^{(1)}) \right]^2 \right\} \right]^{-1} \quad (3.36)$$

3.4 PROS AND CONS OF MODELLING

As indicated earlier, joint estimation of x and γ via an extended Kalman filter provides a reduction in the computation time and storage required for the evaluation of the state estimates. In fact, the reduction in computational effort is magnified for systems of large order. For example, with a 10th order system, the modelling approach requires the estimation of 11 variables and their

associated covariances while the sub-optimal estimators discussed in Chapter 2 each require that 20 variables and their covariances be estimated.

The modelling approach, however, converts the system equations into a set of non-linear difference equations and hence an extended Kalman filter or other approximate non-linear estimators have to be used. Moreover, the requirement that the estimate of γ_k lie within the range (0,1) necessitates the development and the use of constrained estimation schemes (see Chapter 4). A further requirement of the modelling approach is the necessity of handling state-dependent noise. This occurs, for example, in systems subject to noise bursts.

In spite of these difficulties, the simulation studies presented in the next Chapter indicate that the modelling approach performs as well as other sub-optimal estimators cited earlier.

FIG. 3.1. Shape of the model for the switching process in the detection example ($Q_{10}=0.1$).

FIG. 3.2. Comparison of the mean and the mean squared estimation errors in the detection example.

LEGEND: --- $Q_p=0.25$ = Adaptive Q_p
 --- $Q_p=0.10$ = Optimal

FIG. 3.3. Shape of the model for the switching process in the case of a switching plant noise ($Q_{10}=0.1$).

CHAPTER 4

JOINT DETECTION/ESTIMATION IN SWITCHING SYSTEMS VIA MODELLING

4.1 INTRODUCTION

The problem posed in this Chapter is that of estimating the state x_k and the switching variable γ_k jointly via a single extended Kalman filter for the system described by (2.3), which can be expressed compactly as:

$$\begin{aligned} z(k) &= g(z(k-1), U(k-1), w(k-1)) \\ y(k) &= h(z(k), n(k)) \end{aligned} \quad (4.1)$$

where

$$\begin{aligned} z(k) &\triangleq \begin{bmatrix} x(k) \\ \gamma(k) \end{bmatrix} \\ w(k) &\triangleq \begin{bmatrix} \xi(k) \\ \psi(k) \end{bmatrix} \sim N(0, Q(k)) \end{aligned}$$

and the functions g and h can be evaluated once the parameters (A_i, B_i, C_i) , $i=0,1$, and the model function f (equation 3.1) are known. $\xi(k)$ and $\psi(k)$ were defined in Chapters 2 and 3 respectively as:

$$\xi(k-1) = \gamma(k)\xi_1(k-1) + (1-\gamma(k))\xi_0(k-1)$$

$$\psi(k) \sim N(0, Q_\psi)$$

$$\xi_1(k-1) \sim N(0, Q_1)$$

$$\xi_0(k-1) \sim N(0, Q_0)$$

The problem now is to estimate, recursively, the augmented state $z(k)$ in the resulting non-linear system (4.1). The estimation can be done via the extended Kalman filter given below.

4.2 THE EXTENDED KALMAN FILTER

The extended Kalman filter is basically the Kalman filter developed for the system (4.1) after linearizing it about a nominal value of the state $z(k-1)$. The nominal value is taken here as the estimate of the state based on past observations, i.e. $\hat{z}_{k-1|k-1}$. The filter equations are given as (J2):

$$\hat{z}_{k+1|k} = g(\hat{z}_{k|k}, U(k))$$

$$P_{k+1|k} = \phi(\hat{z}_{k|k}) P_{k|k} \phi^T(\hat{z}_{k|k}) + Q(k)$$

$$\hat{z}_{k+1|k+1} = \hat{z}_{k+1|k} + G_{k+1}(y_{k+1} - h(\hat{z}_{k+1|k}))$$

$$P_{k+1|k+1} = (I - G_{k+1} H_{k+1}^T) P_{k+1|k} (I - G_{k+1} H_{k+1}^T)^T + G_{k+1} R(k+1) G_{k+1}^T$$

$$G_{k+1} = P_{k+1|k} H_{k+1}^T (H_{k+1} P_{k+1|k} H_{k+1}^T + R(k+1))^{-1} \quad (4.2)$$

where

$$Q(k) = \text{var}(w(k))$$

$$R(k) = \text{var}(\eta(k))$$

$$H_{k+1} \triangleq \left. \frac{\partial h}{\partial z} \right|_{z=\hat{z}_{k+1}|k}$$

$$\Phi(\hat{z}_k|k) \triangleq \left. \frac{\partial g}{\partial z} \right|_{z=\hat{z}_k|k}$$

The initial conditions needed for the estimator are:

$$\hat{z}_0|_{-1} = E(z_0); P_0|_{-1} = \text{var}(z_0).$$

Two conditions have been ignored in the estimator (4.2): first, determination of $Q(k)$ and $R(k)$ which represent the covariances of $w(k)$ and $\eta(k)$ respectively, requires the analysis of a system with state dependent noise since $w(k)$ and $\eta(k)$ are functions of γ , a component of the state vector z . Second, the estimate of γ_k , $\hat{\gamma}_k|k$, must be constrained to lie within the range $(0,1)$, thus requiring the development of a constrained estimator. These two requirements are considered in the sequel.

4.3 ESTIMATION WITH STATE DEPENDENT NOISE

The first attempt to obtain state estimates under state dependent noise situations appears to be due to

McLane^(M1). He considered the continuous-time problem and derived the optimal linear filter. Rajasekaran et al^(R1) obtained the discrete filter for the case of state dependent noise in the measurements alone. A more general treatment, for both discrete and continuous time cases, was presented by Gustafson and Speyer^(G1).

The approach in all these articles is similar to the one taken in Section 2.2 (Chapter 2) for the case of an additive white noise. For this reason, only the final result of Gustafson and Speyer, modified to suit the system (4.1) is given here. The resulting estimator is seen to be identical to the extended Kalman filter (4.2), with $R(k+1)$ and $Q(k)$ obtained as follows.

From (4.2),

$$Q(k) \triangleq E(\omega(k)\omega(k)^T) \\ = \begin{bmatrix} E(\xi(k)\xi(k)^T) & 0 \\ 0 & Q_\psi \end{bmatrix}$$

Recalling that

$$\xi(k) \triangleq \gamma(k)\xi_1(k-1) + (1-\gamma(k))\xi_0(k-1)$$

where

$$\xi_0(k-1) \sim N(0, Q_0)$$

$$\xi_1(k-1) \sim N(0, Q_1)$$

and $\gamma_k = 0$ or 1, is independent of ξ_1 and ξ_0 .

$$\begin{aligned} E(\xi(k)\xi(k)^T) &= E(\gamma(k)^2)Q_1 + E\{(1-\gamma(k))^2\}Q_0 \\ &= E\{\gamma(k)\}^2(Q_1+Q_0) - 2E(\gamma(k))Q_0 + Q_0 \end{aligned} \quad (4.3)$$

Defining the $(n+1)$ vector e_{n+1} as

$$e_{n+1} \triangleq [0 \ 0 \ 0 \ 0 \ 0 \ \dots \ 0 \ 1]^T,$$

(4.3) can be rewritten as:

$$\begin{aligned} E(\xi(k)\xi(k)^T) &= e_{n+1}^T E(z(k)z(k)^T) e_{n+1} (Q_1+Q_0) \\ &\quad - 2e_{n+1}^T E(z(k)) Q_0 + Q_0 \end{aligned} \quad (4.4)$$

The mean and the covariance of the augmented state $z(k)$ propagate according to the equations:

$$E(z(k+1)) = g(E(z(k)), U(k)) \quad (4.5)$$

$$E(z(k+1)z(k+1)^T) = \Phi E(z(k)z(k)^T) \Phi^T + Q(k) \quad (4.6)$$

with the initial conditions,

$$E(z(0)) = \hat{z}_0|_0; \quad E(z(0)z(0)^T) = P_0|_0 + \hat{z}_0|_0 \hat{z}_0^T|_0$$

Similarly,

$$\begin{aligned} R(k+1) &\triangleq E(n(k+1)n(k+1)^T) \\ &= e_{n+1}^T E(z(k+1)z(k+1)^T) e_{n+1} (R_1+R_0) - 2e_{n+1}^T E(z(k+1)) R_0 + R_0 \end{aligned} \quad (4.7)$$

4.4 CONSTRAINED ESTIMATION

Collmeyer and Gupta^(C2) obtained the optimal linear filter for a linear, continuous-time system under the constraint that the mean squared value of the state estimates be less than a pre-specified limit. Jain and Sworder^(J1) considered the same problem and showed that the constrained estimate can be obtained by a simple linear transformation of the unconstrained estimate. Sankaran and Srinath^(S2) showed that the result of Collmeyer and Gupta can be easily obtained by considering the dual problem, namely, the control of a linear, continuous-time system under energy constraints. In Appendix III, the result of Jain and Sworder has been extended to the system under consideration, i.e., a discrete-time system with magnitude constraints on the estimate of a single state variable (y_k). In simple terms, the result obtained in Appendix III indicates that only the estimate of $\hat{y}(k)$ is altered by the incorporation of the constraint, the estimate of the state vector $x(k)$ remaining the same. In other words, the constrained estimates $\hat{x}_{k|k}^c$ and $\hat{y}_{k|k}^c$ of the state vector $x(k)$ and the switching variable $y(k)$ are given as:

$$\hat{x}_{k|k}^c = \hat{x}_{k|k}$$

$$\begin{aligned} \hat{\gamma}_{k|k}^c &= 1 && \text{if } \hat{\gamma}_{k|k} > 1 \\ &= \hat{\gamma}_{k|k} && \text{if } 0 \leq \hat{\gamma}_{k|k} \leq 1 \\ &= 0 && \text{if } \hat{\gamma}_{k|k} < 0 \end{aligned} \quad (4.8)$$

where the unconstrained estimates $\hat{x}_{k|k}$ and $\hat{\gamma}_{k|k}$ are obtained from the extended Kalman filter (4.2).

Thus, the estimator (4.2) with $Q(k)$ and $R(k+1)$ obtained from (4.4) and (4.7) respectively and $\hat{\gamma}_{k|k}$ modified according to (4.8) specifies the required estimator for the system (4.1).

Several examples are considered in the sequel to illustrate the use of this estimator and compare its performance with those of other sub-optimal estimators indicated in Chapter 2.

4.5 EXAMPLES

In this section, results of simulation studies with the estimator (4.2), augmented by equations (4.4), (4.7) and (4.8), are compared with the sub-optimal estimators considered in Chapter 2. The examples are chosen so as to cover the different types of switching systems mentioned in Chapter 1.

Example 4.1. Scalar System with Missing Observations

A scalar switching system is described by the equations:

$$x_{k+1} = 0.95x_k + \xi_k$$

$$y_k = \gamma_k x_k + \eta_k$$

$$\xi_k \sim N(0, 0.64)$$

$$\eta_k \sim N(0, 1.69)$$

$$x_0 \sim N(30, 2)$$

$$\hat{x}_0|_{-1} = 10 \text{ and } P_0|_{-1} = 2.$$

The statistics of γ are given as:

$$p(\gamma_0=1) = 0.9$$

$$p(\gamma_k \neq \gamma_{k-1}) = 0.1.$$

The performance of the estimator (4.2) is compared with that of the approximate non-linear filter developed by Sawaragi et al^(S3) on the basis of sample mean and root mean squared errors. These are obtained as averages of the results of 100 Monte-Carlo runs, each over 30 time steps. Two such experiments are performed: one with a fixed realization of the sequence $\{\gamma_k, k=1, 2, \dots, 30\}$ and

the other with a variable sequence of γ . The mean and rms errors in estimation via the two approaches are shown in Figs. 4.1a and 4.1b. Running values of the state estimates in a typical Monte-Carlo run are shown in Fig. 4.1c. For the modelling approach, Q_ψ , the variance of the noise in the model for γ , was calculated by equation (3.17). From Fig. 4.1c, it is seen that the estimate obtained via the modelling approach compare very well with those provided by the Sawaragi estimator. Moreover, both estimates are reasonably good owing to the large signal to noise ratio during the transient period under investigation. The rms estimation errors, shown in Fig. 4.1a, are seen to increase when a series of observations are missing (i.e. a sequence of 0's appear in the realization of γ). The reason for this is obvious. The estimation errors propagate according to the plant equation and the presence of the noise term ξ_k with a non-zero variance results in larger rms errors as more observations become unavailable (i.e. γ_k continues to be 0).

Example 4.2. Second Order Underdamped System with Missing Observations.

Consider the second order system with missing observations:

$$x_{k+1} = Ax_k + w_k$$

$$y_k = \gamma_k C x_k + v_k$$

$$\gamma_k = 0 \text{ or } 1$$

where w and v are independent, Gaussian noise sources with

$$w \sim N(0, Q)$$

$$v \sim N(0, R)$$

$$\text{and } x_0 \sim N(\hat{x}_0, P_0).$$

Given,

$$A = \begin{bmatrix} 0 & 1 \\ -0.8 & 1.6 \end{bmatrix}, \quad C = [1 \quad 0], \quad Q = \begin{bmatrix} .25 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 0.5,$$

$$\hat{x}_0 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \quad P_0 = \begin{bmatrix} .25 & 0 \\ 0 & 1 \end{bmatrix}, \quad p(\gamma_0 = 1) = 0.9 \text{ and } q_{10} = 0.1,$$

it is required to estimate the state x_k .

It is noticed that the eigenvalues of A are $0.8 \pm j0.4$, leading to a natural frequency of response of about 0.1 cycles/sampling period. Thus, the average rate of switching in the value of γ is approximately equal to the frequency of oscillation of the system.

From equation (3.31), the model function f is given as:

$$\begin{aligned}\hat{y}_{k|k-1} &= f(\hat{y}_{k-1|k-1}, \hat{x}_{k-1|k-1}) \\ &= \left[1 + \frac{1-\bar{y}_k}{\bar{y}_k} \sqrt{\frac{C P_{k|k-1} C^T + R}{R}} \right. \\ &\quad \left. \exp \left\{ -\frac{y_k^{*2}}{2R} + \frac{(y_k^* - C \hat{x}_{k|k-1})^2}{2(C P_{k|k-1} C^T + R)} \right\} \right]^{-1}\end{aligned}\quad (4.9)$$

where

$$\bar{y}_k = q_{10} + (1-2q_{10})\hat{y}_{k-1|k-1}$$

$$\hat{x}_{k|k-1} = A \hat{x}_{k-1|k-1}$$

$$P_{k|k-1} = A P_{k-1|k-1} A^T + Q$$

$$y_k^* = \bar{y}_k C \hat{x}_{k|k-1}$$

The model (4.9) can be simplified as:

$$\begin{aligned}\hat{y}_{k|k-1} &= \left[1 + \frac{1-\bar{y}_k}{\bar{y}_k} \sqrt{\frac{C P_{k|k-1} C^T + R}{R}} \right. \\ &\quad \left. \exp \{ -(C \hat{x}_{k|k-1})^2 [C P_{k|k-1} C^T \bar{y}_k^2 + 2R\bar{y}_k - R] \} \right]^{-1}\end{aligned}\quad (4.10)$$

For a large signal to noise ratio (i.e. $\|C \hat{x}_{k|k-1}\|^2 \gg R$),
and small error covariance $P_{k|k-1}$, the model given by (4.10)

reduces to the relay type characteristic displayed in Fig. 3.1 while, for a small signal to noise ratio, the model becomes linear in γ , i.e.,

$$\hat{\gamma}_{k|k-1} = \bar{\gamma}_k = q_{10} + (1-2q_{10})\hat{\gamma}_{k-1|k-1}.$$

Estimation errors with the modelling approach are compared with those resulting from the use of Bruckner and Scott^(B1)'s estimator in Figs. 4.2a and 4.2b. The errors shown in these figures are the averages of the results of 100 Monte-Carlo runs, each over 30 time steps (approximately 3 cycles of system response), with a fixed realization of the γ -sequence. Simulation studies indicated that a constant value of Q_ψ , the variance of the noise added to the model for γ , was quite unsatisfactory. An optimistic value, say $Q_\psi=0.1$, gave large transients at the occurrence of a switch in γ while, with a relatively pessimistic choice ($Q_\psi=0.25$), there were smaller transients at a switch in γ but larger errors at other sampling instants. The choice of Q_ψ according to equation (3.17), (see also Appendix I), by contrast, performed very well and the resulting estimates of the states were comparable to those obtained by the Bruckner and Scott approach. Further, the two approaches result in estimates comparable in accuracy to

those obtained under the assumption that the γ -sequence is known exactly. Again, as in the previous example, estimation errors increase with a failure of the observation process. It can be shown in this case that, in the absence of any observations, the steady state mean squared errors in the estimation of x_1 and x_2 approach 14 and 16, respectively. In terms of CPU time, the modelling approach was slightly faster (*8%) compared to the Bruckner and Scott estimator (a saving of 10% was predicted in Chapter 3 for a second order system with two modes).

Example 4.3. Joint Estimation of State and Time Constant

This example considers the identification of the time constant of a scalar system with missing observations, defined by the equations:

$$\left. \begin{aligned} x_{k+1} &= a x_k + U_k + w_k \\ y_k &= \gamma_k x_k + v_k \end{aligned} \right\} \quad (4.11)$$

w_k and v_k are samples from independent, zero-mean, Gaussian white noise processes with covariances Q and R , respectively, and γ_k is a binary random variable with the statistics,

$$P(\gamma_0=1)=q; \quad P(\gamma_k \neq \gamma_{k-1})=q_01.$$

Given that $x_0 \sim N(\hat{x}_0, P_0)$, it is required to obtain an estimate $\hat{a}_{k|k}$ of a , based on observations up to time k such that

$$J = \sum_{k=1}^N (a - \hat{a}_{k|k})^2$$

is minimized. N is the length of the identification experiment.

It is readily seen that the system ceases to be observable at time k when $\gamma_k = 0$, implying that the system of equations (4.11) can not be transformed to a difference equation relating the input and the output. The only possible approach to the estimation or identification of the parameter a is to estimate it jointly with the state x . This can be achieved by posing the parameter ' a ' as an additional state with the state equation

$$a_{k+1} = a_k, \text{ (it is assumed here that 'a' is a constant)}$$

and estimating the augmented state vector via an extended Kalman filter. Thus, the system becomes:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{k+1} = \begin{bmatrix} x_1(k) & x_2(k) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} u_k \\ 0 \end{bmatrix} + \begin{bmatrix} w_k \\ 0 \end{bmatrix}$$

$$y_k = \gamma_k \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + v_k \quad (4.12)$$

The estimator of Bruckner and Scott^(B1) (see equation 2.35) can be employed for estimating the state vector with the following modifications:

- 1) the predicted state estimate is:

$$\hat{x}_{k+1|k} = \begin{bmatrix} \hat{x}_{1\ k|k} & \hat{x}_{2\ k|k} \\ \hat{x}_{2\ k|k} & \hat{x}_{1\ k|k} \end{bmatrix} + \begin{bmatrix} u_k \\ 0 \end{bmatrix}$$

- 2) the covariance of the predicted state is:

$$P_{k+1|k} = \Phi P_{k|k} \Phi^T + Q_k$$

where

$$\Phi \triangleq \begin{bmatrix} \hat{x}_{2\ k|k} & \hat{x}_{1\ k|k} \\ 0 & 1 \end{bmatrix}$$

and

$$Q_k \triangleq \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}$$

The system (4.11) was simulated with the following parameters:

$$\hat{x}_0 = 10, P_0 = 0.25, a = 0.9, Q = 0.25, R = 0.5, q = 0.9, q_{10} = 0.1$$

As before, mean and mean squared estimation errors were calculated by averaging the results of 50 Monte-Carlo runs, each over 30 time steps with a fixed realization of the γ -sequence. The performance of the Bruckner and Scott estimator is compared with that of the estimator based on a model for γ_k in Figs. 4.3a, 4.3b and 4.3c. The model for the switching variable is seen to be identical to (4.10), since the system under consideration (4.12) is also a second order system with missing observations. For the purpose of estimation, the initial estimate of 'a' and its covariance were assumed to be:

$$\hat{a}_0|_{-1} = 0, \text{cov}(\hat{a}_0|_{-1}) = 0.1.$$

The results shown in Figs. 4.3a, b and c indicate that the performance of the Bruckner and Scott estimator is inferior to the modelling approach. The reason for this is as follows: In the modelling approach, γ_k is posed as an additional state and hence the error covariance in the estimation of γ_k increased the covariance of the residuals directly. This causes the Kalman gain to be smaller and hence the estimator, in this case, is not very sensitive to the error in the estimation of the state vector. On the other hand, the Kalman gain in the Bruckner and Scott estimator is larger by comparison resulting in

the filter being very sensitive to estimation errors. It is seen from Fig. 4.3a that, at $k=10$, when γ switches to zero, the estimate of a , $\hat{a}_{a|a}$, is largely in error. In the absence of any more observations, the estimation errors propagate according to the plant equation (4.12) and the large error in $\hat{a}$ causes a monotonic increase in the error in the estimation of the state. From Figs. 4.3b and 4.3c, it is seen that the large Kalman gain in the Bruckner and Scott estimator causes large fluctuations in the estimate $\hat{a}_{k|k}$ of the parameter a . By comparison, the modelling approach results in smaller fluctuations in the estimates and hence smaller mean squared errors. It is noticed, however, that given a large string of observations both estimators result in very good estimates of the state x and the parameter a .

Example 4.4. Scalar System with Switching Time Constant

A scalar, discrete-time system with a switching time constant is represented by the equations:

$$x_{k+1} = [\gamma_{k+1} a_1 + (1 - \gamma_{k+1}) a_0] x_k + b U_k + w_k \quad (4.13)$$

$$y_k = C x_k + v_k.$$

The system was simulated with the following parameters:

$$C=1, Q=0.5, R=0.25, a_1=0.95, a_0=-0.8, b=1;$$

$$\hat{x}_{0|-1}=5, P_{0|-1}=2, q_{\Delta P}(Y_0=1)=0.9, q_{10\Delta P}(Y_k \neq Y_{k-1})=0.1,$$

$$U_k \equiv 1 \forall k.$$

Estimation errors were calculated as the averages of 40 Monte-Carlo runs, each over 50 time steps. As before, estimation via the Bruckner and Scott approach is compared with that via modelling the switching process. The model function f is given as (see equation 3.13):

$$\hat{Y}_{k|k-1} = \left[1 + \frac{1 - \bar{Y}_k}{\bar{Y}_k} \sqrt{\frac{a_1^2 P_{k-1|k-1} + Q + R}{a_0^2 P_{k-1|k-1} + Q + R}} \right. \\ \left. \exp \left\{ - \frac{(y_k^* - a_0 \hat{x}_{k-1|k-1} - b U_{k-1})^2}{2(a_0^2 P_{k-1|k-1} + Q + R)} \right. \right. \\ \left. \left. + \frac{(y_k^* - a_1 \hat{x}_{k-1|k-1} - b U_{k-1})^2}{2(a_1^2 P_{k-1|k-1} + Q + R)} \right\} \right]^{-1}$$

$$\bar{Y}_k = q_{10} + (1 - 2q_{10}) \hat{Y}_{k-1|k-1}$$

$$y_k^* = [a_0(1 - \bar{Y}_k) + a_1 \bar{Y}_k] \hat{x}_{k-1|k-1} + b U_{k-1} \quad (4.14)$$

Attempts to determine Q_ψ along the lines of equation (3.17) failed in this case, resulting in large estimation errors. The reason for this was observed to be the large value of Q_ψ (approximately 0.65) resulting from equation (3.17). By contrast, the choice, $Q_\psi=0.1$, fared very well and the resulting state estimates were comparable in accuracy to those obtained via the Bruckner and Scott approach. Fig. 4.4a compares the mean and the mean squared estimation errors in the two approaches. In Fig. 4.4b, running values of the estimates of the state x_k and the switching variable γ_k in a typical Monte-Carlo run are plotted. This figure indicates that the modelling approach results in state estimates comparable in accuracy to those of Bruckner and Scott. The estimates of the switching variable are, however, in error when the signal magnitude ($|x_k|$) is small. The reason for this is obvious. The a-posteriori densities $p(y_k|Y^{k-1}, \gamma_k=i)$, $i=0,1$ both have means very close to zero, in this case, making it difficult to decide the value of γ_k . Both the modelling approach and the Bruckner and Scott approach result in large errors at the occurrence of a switch in γ_k owing to the inability of the filters in tracking the switch. This effect is more pronounced when two switches occur consecutively. It

FIG. 4.2a. Comparison of average estimation errors for the second order example with missing observations. (Q_y for modelling approach given by equation 3.17)

FIG. 4.2b. Comparison of the mean squared estimation errors.

LEGEND:
 ○ y-known exactly
 ■ modelling approach
 ▲ Bruckner & Scott

FIG. 4.1a. Mean and RMS estimation errors in the scalar example with missing observations (modelling and Savaragi approaches with a fixed compared γ -sequence).

Fig. 4.1b. Mean and RMS estimation errors in the scalar example with missing observations (modelling and Sawaragi approaches compared with a variable γ -sequence).

FIG. 4.1c. State and y -estimates in a typical run of the scalar system with missing observations (example 4.1). Modelling and Sawaragi approaches compared.

FIG. 4.2a. Comparison of average estimation errors for the second order example with missing observations. (Q_p for modelling approach given by equation 3.17)

FIG. 4.2b. Comparison of the mean squared estimation errors.

LEGEND: + y-known exactly
 x modelling approach
 o Bruckner & Scott

FIG. 4.3a. Estimates of the state and the parameter 'a' in a typical run of the system in Example 4.3. Modelling and the Bruckner and Scott approaches compared.

FIG. 4.3b. Comparison of the mean squared errors in the estimation of the state in the example 4.3.

2 3

OF/DE

In the case of a scalar output, y_k , n such equations are necessary and sufficient for the determination of x_k . In other words, the system (5.1) is observable, at time k , if the matrix S defined by (5.3) is of full rank^(R4):

$$S \triangleq \begin{bmatrix} C(k)\phi(k,k) \\ C(k+1)\phi(k+1,k) \\ \vdots \\ C(k+n-1)\phi(k+n-1,k) \end{bmatrix} \quad (5.3)$$

Notice that S is a function of γ_j , $k \leq j \leq k+n-1$, and hence must be non-singular for all values of γ_j in this period. Needless to say, this requirement is not satisfied for the missing observations case. In the case of a vector output, the matrix S is formed by starting at the bottom, including most recent equations first, and going upwards until n linearly independent rows have been obtained^(R4). With S defined by (5.3), and assuming a scalar output, y_{k+n} is obtained from an equation of the form:

$$y_{k+n} = C(k+n) [F_1 \underline{Y} + F_2 \underline{U}] + e_{k+n} \quad (5.4)$$

where e_{k+n} is an error term and is a function of n and ξ , and F_1 and F_2 are $n \times n$ matrices whose components are functions of the elements of the matrices A_i , B_i , C_i , $i=0,1$ and the past $(n-1)$ values of γ , i.e., γ_j , $k \leq j \leq k+n-1$. $\underline{Y}$ and $\underline{U}$ are $(n-1)$ vectors of past outputs and inputs respectively.

FIG. 4.4a. Comparison of the mean and the mean squared estimation errors for the scalar example with a switching time constant.

FIG. 4.4b. State and y -estimates in a typical Monte-Carlo run of the system with a switching time constant in example 4.4.

CHAPTER 5

IDENTIFICATION OF SWITCHING SYSTEMS

5.1 INTRODUCTION

The need for system identification arises from either one of two requirements: to learn more about the system and its environment or to design a suitable controller for the system. This Chapter analyzes the problem of estimating the parameters of a system with a switching model.

For systems with constant or time varying parameters, there exist a multitude of schemes for parameter estimation. A complete survey of the literature on identification schemes is not presented in this thesis - mainly because there exist several excellent survey articles on this subject (A7,E1). These papers have extensive bibliographies. Systems considered in these articles are assumed to be represented adequately by a single model during the entire period of the identification experiment.

The systems analyzed in this Chapter, on the other hand, are represented by two models, one of them being the correct one at any sampling instant. System identification in this case requires the estimation of the para-

meters of both models. An obvious solution to this problem is to partition the data set, i.e., the set of input and output measurements over the identification period, into two subsets corresponding to the two models which generated them and use these subsets to estimate the parameters of the corresponding models via an appropriate identification scheme. Partitioning of the data set, however, becomes difficult and often inaccurate when the measurements are in error. The identification schemes reported in the literature must therefore be modified in order to account for the uncertainty in the partitioning of the data set.

A brief resume of this Chapter is now presented. Section 5.2 presents the statement of the identification problem and the scope of the work reported in this Chapter. Section 5.3 considers the off-line solution to the problem via the maximum likelihood approach. An approach based on iterating between the estimation of the parameters and improving the partitioning accuracy is suggested. Section 5.4 considers various approaches to the on-line estimation of the parameters of systems into switching models. In particular, the use of the modelling approach (Chapter 3) for the joint estimation of the parameter vector and the switching variable via an extended Kalman filter is considered. It is noticed in this regard that the last example

considered in Chapter 4 is basically an example on system identification. Section 5.4 has two sub-sections, the first considering the problem of uncorrelated error terms while the second analyzing the general problem of correlated errors. In the case of uncorrelated errors, a modification of the ordinary least squares approach, similar to the Bank of Kalman filters discussed earlier (Chapter 2), is shown to give acceptable estimates of the parameters. Similar conclusions are drawn with the use of the modelling approach. This approach is then extended to the more realistic case of correlated error terms by considering a similar modification of the instrumental variable (IV) approach (for details of IV estimators see Pandya^(P1) or Rowe^(R3)). Conclusions arrived at, based on simulation studies presented in Section 5.4, are summarized in Section 5.5.

5.2 THE PROBLEM

A system is assumed to be modelled by the following equation:

$$\begin{aligned} x_k &= \gamma_k [A_1 x_{k-1} + B_1 u_{k-1}] + [1 - \gamma_k] [A_0 x_{k-1} + B_0 u_{k-1}] + \epsilon_{k-1} \\ y_k &= \gamma_k C_1 x_k + (1 - \gamma_k) C_0 x_k + a_k \end{aligned} \quad (5.1)$$

The unknown elements in the matrices $A_i, B_i, C_i, i=0,1,$

and the covariance matrices Q and R need to be estimated.

For the purpose of system identification, it is easier to consider the equivalent difference equation model relating the input and the output directly, since the latter requires fewer parameters to be estimated (see Astrom^(A7)). Transformation to the difference equation model requires the elimination of the state x_k from (5.1). Equation (5.1) represents a time-varying linear system for which the general solution is given as^(D2):

$$y_{k+1} = C(k+1)\phi(k+1,k)x_k + \sum_{m=k}^{k+1-1} C(k+1)\phi(k+1,m+1) [B(m+1)U_m + \xi_m] + \eta_{k+1} \quad (5.2)$$

where

$\phi(k_1, k_2)$ is the state-transition matrix defined as:

$$\phi(k_1, k_2) = \prod_{i=k_2+1}^{k_1} A(i), \quad k_2 < k_1; \quad \phi(k_1, k_1) = I$$

and

$C(i)$, $B(i)$, and $A(i)$ were defined in Chapter 2 as:

$$C(i) = \gamma_1 C_1 + (1-\gamma_1)C_0$$

$$B(i) = \gamma_1 B_1 + (1-\gamma_1)B_0$$

$$A(i) = \gamma_1 A_1 + (1-\gamma_1)A_0$$

In the case of a scalar output, y_k , n such equations are necessary and sufficient for the determination of x_k . In other words, the system (5.1) is observable, at time k , if the matrix S defined by (5.3) is of full rank^(R4):

$$S \triangleq \begin{bmatrix} C(k)\phi(k,k) \\ C(k+1)\phi(k+1,k) \\ \vdots \\ C(k+n-1)\phi(k+n-1,k) \end{bmatrix} \quad (5.3)$$

Notice that S is a function of y_j , $k \leq j \leq k+n-1$, and hence must be non-singular for all values of y_j in this period. Needless to say, this requirement is not satisfied for the missing observations case. In the case of a vector output, the matrix S is formed by starting at the bottom, including most recent equations first, and going upwards until n linearly independent rows have been obtained^(R4). With S defined by (5.3), and assuming a scalar output, y_{k+n} is obtained from an equation of the form:

$$y_{k+n} = C(k+n) [F_1 \underline{Y} + F_2 \underline{U}] + e_{k+n} \quad (5.4)$$

where e_{k+n} is an error term and is a function of n and ξ , and F_1 and F_2 are $n \times n$ matrices whose components are functions of the elements of the matrices A_i , B_i , C_i , $i=0,1$ and the past $(n-1)$ values of y , i.e. y_j , $k \leq j \leq k+n-1$. $\underline{Y}$ and $\underline{U}$ are $(n-1)$ vectors of past outputs and inputs respectively.

The difference equation (5.4) is seen to be linear in $\underline{Y}$, $\underline{U}$ and γ_{k+n} while being non-linear in γ_j , $k \leq j \leq k-n+1$. This general form is difficult to analyze, so a simpler version of (5.4) is considered in the sequel. The matrices F_1 and F_2 are assumed constant. This is a meaningful approximation provided the order of the system, n , and the sampling period are small and the transition probabilities of the Markov chain for γ are such that the probability of a switch occurring over n consecutive sampling instants is small. However, the algorithms developed in the sequel can be easily extended to cover the general case (5.4) as indicated in Section 5.5. The model examined in this Chapter is thus given as:

$$\begin{aligned}
 y_{k+n} = & \gamma_{k+n} \sum_{i=1}^n \left[a_i^{(1)} y_{k+n-i} + b_i^{(1)} u_{k+n-i} + c_i^{(1)} e_{k+n-i} \right] \\
 & + (1 - \gamma_{k+n}) \sum_{i=1}^n \left[a_i^{(0)} y_{k+n-i} + b_i^{(0)} u_{k+n-i} + c_i^{(0)} e_{k+n-i} \right] \\
 & + e_{k+n}
 \end{aligned} \tag{5.5}$$

where e_{k+n} is an error term having a zero-mean Gaussian distribution with variance R .

It is readily seen that the model (5.5) results from the discretization (with a small sampling period) of a differential equation model with coefficients switching

simultaneously. The corresponding state-space model will then be a function of not only the present value of y , but also its past $(n-1)$ values.

The problem posed in this Chapter is to estimate the parameter vectors $\theta^{(1)}$ and $\theta^{(0)}$ defined as:

$$\theta^{(i)} \triangleq \left[a_1^{(i)}, a_2^{(i)}, \dots, a_n^{(i)}, b_1^{(i)}, \dots, b_n^{(i)}, c_1^{(i)}, \dots, c_n^{(i)} \right]^T, i=0,1$$

and the values taken by y , from the input-output data collected for N sampling periods. The available data is represented as $\{Y^N, U^N\}$:

$$Y^N \triangleq \{y_0, y_1, \dots, y_N\}$$

$$U^N \triangleq \{u_0, u_1, \dots, u_N\}$$

In the sequel, off-line and on-line solutions to the problem are discussed.

5.3 OFF-LINE IDENTIFICATION

The approach taken here is to obtain an estimate of the parameters such that the likelihood function is maximized. The likelihood function is defined as the joint probability density of the outputs given the inputs and the parameters (K) . Thus,

$$l_N(\theta^{(0)}, \theta^{(1)}, I_N) = p(y_1, y_2, \dots, y_N | y_0, U^N, \theta^{(0)}, \theta^{(1)}, I_N)$$

where

$$I_N \triangleq \{y_0, y_1, \dots, y_N\}$$

Continued application of Bayes' rule results in:

$$l_N(\theta^{(0)}, \theta^{(1)}, I_N) = \prod_{i=1}^N p(y_i | y^{i-1}, U^{i-1}, \theta^{(0)}, \theta^{(1)}, I_i) \quad (5.6)$$

The density $p(y_i | y^{i-1}, U^{i-1}, \theta^{(0)}, \theta^{(1)}, y_i = j)$, $j=0,1$, is

seen from (5.5) to be Gaussian with a mean

$$\sum_{i=1}^n \left[a_i^{(j)} y_{k-i} + b_i^{(j)} U_{k-i} \right] \text{ and a covariance } R \left(1 + \sum_{i=1}^n C_i^{(j)^2} \right).$$

Thus,

$$l_N(\theta^{(0)}, \theta^{(1)}, I_N) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi R \left(1 + \sum_{k=1}^n C_k^{(j)^2} \right)}}$$

$$\exp \left\{ - \frac{1}{2R \left(1 + \sum_{k=1}^n C_k^{(j)^2} \right)} \right.$$

$$\left. \left[y_i - \sum_{k=0}^n (a_k^{(j)} y_{i-k} + b_k^{(j)} U_{i-k}) \right]^2 \right\}$$

where the superscript (j) denotes the fact that the $(i+1)$ th component of I_N , y_i is equal to j .

From a computational point of view, it is easier to consider the logarithm of the likelihood function, given

as:

$$\begin{aligned}
 L_N(\theta^{(0)}, \theta^{(1)}, I_N) &= \sum_{i=1}^N \frac{1}{2} \ln \left\{ 2\pi R \left(1 + \sum_{j=1}^n c_k^{(j)2} \right) \right\} \\
 &\quad - \sum_{i=1}^N \frac{1}{2R \left\{ 1 + \sum_{k=1}^n c_k^{(j)2} \right\}} \\
 &\quad \left[y_i - \sum_{k=0}^n a_k^{(j)} y_{i-k} + b_k^{(j)} u_{i-k} \right]^2
 \end{aligned} \tag{5.7}$$

Maximization of (5.7) with respect to $\theta^{(0)}, \theta^{(1)}$ and I_N yields the maximum likelihood estimates $\hat{\theta}_{ML}^{(0)}, \hat{\theta}_{ML}^{(1)}$ and $I_{N(ML)}$. If I_N is known, the partitioning of the data set and hence the likelihood function (5.7) into two separate parts and subsequent estimation of $\theta^{(0)}$ and $\theta^{(1)}$ via two separate maximization routines, is seen to be straight forward. However, if I_N is also to be estimated, maximization becomes difficult. The following iterative approach is proposed in this case:

Step 1. Start with estimates $\hat{\theta}^{(0)}, \hat{\theta}^{(1)}$ of $\theta^{(0)}$ and $\theta^{(1)}$ respectively. Maximize $L_N(\hat{\theta}^{(0)}, \hat{\theta}^{(1)}, I_N)$ with regard to I_N yielding $\hat{I}_N$ as $\hat{Y}_1 \Delta((i+1)\text{th component of } I_N)$
 $= 1$ if $L_N(\hat{\theta}^{(0)}, \hat{\theta}^{(1)}, \hat{I}_N | \hat{Y}_1 = 1) > L_N(\hat{\theta}^{(0)}, \hat{\theta}^{(1)}, I_N | \hat{Y}_1 = 0)$
 $= 0$ otherwise.

Or,

$$\hat{y}_i = \frac{1}{2} + \frac{1}{2} \operatorname{Sgn} \left[\frac{\{y_i - (\sum_{k=1}^n a_k^{(1)} y_{i-k} + b_k^{(1)} u_{i-k})\}^2}{1 + \sum_{k=1}^n c_k^{(1)^2}} - \frac{\{y_i - (\sum_{k=1}^n a_k^{(0)} y_{i-k} + b_k^{(0)} u_{i-k})\}^2}{1 + \sum_{k=1}^n c_k^{(0)^2}} \right]$$

for $i=1,2, \dots, N$ where $\operatorname{Sgn}(x)=1$ for $x \geq 0$ and is -1 otherwise.

Step 2. Maximize $L_N(\theta^{(0)}, \theta^{(1)}, \hat{I}_N)$ with regard to $\theta^{(0)}$ and $\theta^{(1)}$. This can be done by two separate maximization routines as mentioned earlier.

Step 3. Go back to step 1, unless the estimates have converged, i.e. change by less than a pre-specified "tolerance limit".

There are two obvious reasons for estimating I_N first and then $\theta^{(1)}$ and $\theta^{(0)}$: firstly, the accuracy of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(0)}$, as mentioned earlier, depends on that in the partitioning of the data set and hence $\hat{I}_N$. Secondly, the estimate $\hat{I}_N$, obtained from the decision approach taken in step 1, is relatively less sensitive to $\hat{\theta}^{(0)}$ and $\hat{\theta}^{(1)}$ compared to the sensitivity of the parameter estimates to the estimate of I_N . The accuracy in the partitioning of the data set

is seen to depend on the signal to noise ratio. This conclusion was supported by the simulation study of a first order system with uncorrelated error terms (i.e. $c_k^{(1)} = 0$, $i=0,1$, $1 \leq k \leq n$). It was observed that for a fixed noise covariance ($R=0.1$) the accuracy of partitioning was very sensitive to y_0 , the initial output. For $y_0=10$, for example, the partitioning was exact while for $y_0=1$ the partitioning was grossly in error.

5.4 ON-LINE IDENTIFICATION

The previous section outlined the difficulties involved in off-line identification of switching systems. Precisely, the same difficulties are encountered in on-line identification.

To simplify the ensuing discussion, consider the problem of estimating a constant signal, x , from uncertain, noisy measurements:

$$y_k = \gamma_k x + e_k$$

where $e_k \sim N(0, R)$ and $\gamma_k = 0$ or 1 indicates whether the observation y_k contains the signal ($\gamma_k = 1$) or not ($\gamma_k = 0$).

The least squares estimate $\hat{x}_{LS}$ obtained by minimizing the sum of squared errors, viz.,

$$J = \sum_{i=1}^N (y_i - \gamma_i x)^2$$

where N is the number of observations,
is given as

$$\hat{x}_{LS} = \frac{\sum_{i=1}^N y_i \gamma_i}{N_1},$$

if γ_i , $i=1, N$ are known. $N_1 = \sum_{i=1}^N \gamma_i$ represents the number of observations containing the signal among the N observations. The estimate $\hat{x}_{LS}$ is an unbiased and consistent estimate of $x^{(A7)}$ and can be put in a recursive form for on-line implementation as

$$\hat{x}_{N+1} = \hat{x}_N + \frac{\gamma_{N+1}}{N_1 + \gamma_{N+1}} (y_{N+1} - \hat{x}_N)$$

where $\hat{x}_k$ is the estimate based on k observations. The estimate is seen to be updated only at sampling instants at which γ is 1.

When γ is not known, the above estimator needs to be modified. Needless to say, the estimate $\hat{x}$ obtained with a modified estimator ceases to be optimal in the least squares sense but is expected to approach the least squares estimate if good estimates of γ are possible. With this background, consider the estimation of the parameters of the system (5.5).

5.4.1 Uncorrelated Errors

Consider the case of model (5.5) with uncorrelated errors (i.e. $c_k^{(1)} = 0$, $i=0,1$, $1 \leq k \leq n$). The parameters to be estimated are:

$$\theta^{(1)} = \left[a_1^{(1)}, a_2^{(1)}, \dots, a_n^{(1)}, b_1^{(1)}, \dots, b_n^{(1)} \right]^T, \quad i=0,1$$

If γ is known, the least squares estimator for $\theta^{(0)}$ and $\theta^{(1)}$ can be put in a recursive form, following the result for time invariant systems (A7) as

$$\hat{\theta}_{k+1}^{(1)} = \hat{\theta}_k^{(1)} + G_{k+1}^{(1)} \left[y_{k+1} - |1 - \gamma_{k+1}| m_{k+1} \hat{\theta}_k^{(1)} \right], \quad i=0,1 \quad (5.8)$$

where

$\hat{\theta}_k^{(1)}$ Δ least squares estimate of $\theta^{(1)}$ based on observations up to time k ,

and $m_{k+1} = [y_k, y_{k-1}, \dots, y_{k-n+1}, u_k, u_{k-1}, \dots, u_{k-n+1}]$.

The gains $G_{k+1}^{(1)}$ are given as:

$$G_{k+1}^{(1)} = \frac{|1 - \gamma_{k+1}| P_k^{(1)T} m_{k+1}}{1 + |1 - \gamma_{k+1}| m_{k+1} P_k^{(1)T} m_{k+1}}$$

$$P_{k+1}^{(1)} = P_k^{(1)} - |1 - \gamma_{k+1}| G_{k+1}^{(1)} m_{k+1} P_k^{(1)}$$

It is seen that the gain $G_{k+1}^{(1)}$ is zero when $\gamma_{k+1} \neq 1$. This implies that y_{k+1} and U_k are used for updating $\hat{\theta}_k^{(1)}$ and $P_k^{(1)}$ only if $\gamma_{k+1} = 1$.

Since γ is not known, the recursive estimator (5.8) must be modified. Define the information available at the end of stage k as

$$I_k^* = \{y^k, U^{k-1}, (\gamma_k = 1 | y^k), \hat{\theta}_k^{(0)}, \hat{\theta}_k^{(1)}\}.$$

At time $(k+1)$, U_k is applied to the system, y_{k+1} is observed and $\hat{\theta}_{k+1}^{(i)}$, $i=0,1$ is required given I_k^* , U_k and y_{k+1} .

Consider the following approaches:

1. Adaptive approach - obtain an estimate $\hat{\gamma}_{k+1}$, of γ_{k+1} based on available information and use $\hat{\gamma}_{k+1}$ instead of γ_{k+1} in (5.8).
2. Decision theoretic approach - decide to accept either one of the two possible values of γ_{k+1} based on $\{I_k^*, U_k, y_{k+1}\}$.
3. Smoothing approach - smooth out in some fashion the effect on $\hat{\theta}_{k+1}^{(i)}$ of the uncertainty in the value of γ_{k+1} .

A fourth approach is possible if direct measurements of γ are available via either a proxy variable or some kind of alarm sounding system. In this case, the estimate of γ_{k+1} can be obtained, independent of the estimator for

$\theta^{(1)}$, $i=0,1$. In the sequel, the more general case, viz. where no direct measurements of γ are possible, is considered.

Adaptive Approach

The minimum mean squared error estimate of γ_{k+1} is given as (see Chapter 2):

$$\begin{aligned}\hat{\gamma}_{k+1} &\triangleq E(\gamma_{k+1} | y_{k+1}, U_k, I_k^*) \\ &= p(\gamma_{k+1}=1 | y_{k+1}, U_k, I_k^*)\end{aligned}$$

This estimate can be obtained by the application of Bayes' rule as:

$$\hat{\gamma}_{k+1} = \frac{p(y_{k+1} | \gamma_{k+1}=1, I_k^*, U_k) p(\gamma_{k+1}=1 | I_k^*, U_k)}{\sum_{j=0}^1 p(y_{k+1} | \gamma_{k+1}=j, I_k^*, U_k) p(\gamma_{k+1}=j | I_k^*, U_k)} \quad (5.9)$$

The conditional densities of the output y_{k+1} were obtained in Section 5.3 and the densities $p(\gamma_{k+1}=j | I_k^*, U_k)$ are given as:

$$p(\gamma_{k+1}=1 | I_k^*, U_k) = q_{11}\hat{\gamma}_k + q_{10}(1-\hat{\gamma}_k)$$

and

$$p(\gamma_{k+1}=0 | I_k^*, U_k) = 1 - p(\gamma_{k+1}=1 | I_k^*, U_k)$$

As indicated in Chapter 2, this estimate, being a probability, takes values in the continuous range (0,1) rather than the discrete values 0 and 1. This estimate is used in place of γ_{k+1} in (5.8).

Decision Approach

Lainiotis et al^(L1) have shown, for a similar problem in communication theory (Joint detection/estimation), that γ_{k+1} represents a sufficient statistic for optimal Bayes' decision. That is, given the hypotheses:

$$H_0: y_{k+1} = m_{k+1} \theta^{(1)} + e_{k+1}$$

$$H_1: y_{k+1} = m_{k+1} \theta^{(0)} + e_{k+1},$$

the optimal decision procedure is to decide H_0 or H_1 , depending on whether $\hat{\gamma}_{k+1} \gtrless f(c)$, respectively, where $f(c)$ is a function of the a-priori costs of each decision. In the particular case of equal costs associated with a miss (i.e. deciding H_1 when H_0 is true) and a false alarm (i.e. accepting H_0 when H_1 is true), the threshold $f(c)$ is 1/2. Again, as indicated in Chapter 2, the accuracy of $\hat{\gamma}_{k+1}$ and hence that of the decision depends on the signal to noise ratio. The result of the decision (i.e. $\gamma_{k+1} = 0$ or 1) is substituted for γ_{k+1} in (5.8).

Smoothing Approach

Rewrite the estimator (5.8) as

$$\hat{\theta}_{k+1}^{(1)} = \hat{\theta}_k^{(1)} + \epsilon_{k+1}(y_{k+1}, \hat{\theta}_k^{(1)}, 1, \gamma_{k+1}, U_k)$$

or
$$\hat{\theta}_{k+1}^{(1)} = \hat{\theta}_k^{(1)} + \epsilon_{k+1}(y_{k+1}, I_k^*, 1, \gamma_{k+1}, U_k)$$

The correction or updating term ϵ is a function of γ_{k+1} . It is obvious that the lack of knowledge of the value of γ_{k+1} is a major source of error in the estimator (5.8). In an attempt to reduce this error, ϵ_{k+1} is obtained as the weighted average of the two correction terms, viz.,

$$\epsilon_{k+1}(y_{k+1}, I_k^*, 1, \gamma_{k+1}=1, U_k) \text{ and}$$

$$\epsilon_{k+1}(I_k^*, y_{k+1}, 1, \gamma_{k+1}=j, U_k).$$

$$\epsilon_{k+1} = \sum_{j=0}^1 w_j \epsilon_{k+1}(I_k^*, y_{k+1}, 1, \gamma_{k+1}=j, U_k) \quad (5.10)$$

The weights w_j are to be chosen so that

$$w_j = \sigma_{1j},$$

where $\sigma_{1j} = 0$ if j

$$= 1 \text{ if } j=1.$$

This requirement is satisfied by the choice:

$$w_j = p(\gamma_{k+1}=j | y_{k+1}, I_k^*, U_k).$$

With this choice of weights, the estimator (5.8) is modified as:

For $i=0,1$,

$$\hat{\theta}_{k+1}^{(i)} = \hat{\theta}_k^{(i)} + \sum_{j=0}^1 w_j G_{k+1}^{(i,j)} (y_{k+1} - \sigma_{ij} m_{k+1} \hat{\theta}_k^{(i)})$$

$$G_{k+1}^{(i,j)} = \frac{\sigma_{ij} P_k^{(i,j)} m_{k+1}^T}{1 + \sigma_{ij} m_{k+1} P_k^{(i,j)} m_{k+1}^T}$$

$$P_k^{(i,j)} = P_k^{(i,j)} - w_j G_{k+1}^{(i,j)} m_{k+1} P_k^{(i,j)} \quad (5.11)$$

Noting that $G_{k+1}^{(i,j)} = 0$ and $P_{k+1}^{(i,j)} = P_k^{(i,j)}$ for $i \neq j$, (5.11) can be simplified as:

$$\hat{\theta}_{k+1}^{(i)} = \hat{\theta}_k^{(i)} + p(\gamma_{k+1} = i | I_k^*, y_{k+1}, U_k) G_{k+1}^{(i)} (y_{k+1} - m_{k+1} \hat{\theta}_k^{(i)})$$

$$G_{k+1}^{(i)} = \frac{P_k^{(i)} m_{k+1}^T}{1 + m_{k+1} P_k^{(i)} m_{k+1}^T}$$

$$P_{k+1}^{(i)} = P_k^{(i)} - p(\gamma_{k+1} = i | I_k^*, y_{k+1}, U_k) G_{k+1}^{(i)} m_{k+1} P_k^{(i)}, \quad i=0,1$$

(5.12)

and

$$p(\gamma_{k+1} = 1 | I_k^*, y_{k+1}, U_k) = \hat{\gamma}_{k+1} \text{ (equation 5.9)}$$

$$p(\gamma_{k+1} = 0 | I_k^*, y_{k+1}, U_k) = 1 - \hat{\gamma}_{k+1}$$

The similarity between (5.12) and the Bank of Kalman filters developed for state estimation (Chapter 2) is not surprising since the least squares estimator is identical to a Kalman filter developed for the system (5.5) with the parameters posed as states^(A7).

The three approaches presented in this Section are seen to provide identical estimates when γ_{k+1} is either 0 or 1, or when applied to non-switching systems (i.e. $p(\gamma_k=1)=1 \forall k$), since all of them reduce to the least squares estimator (5.8). However, the estimator (5.12) appears to be the most appealing one since both the gains $G_{k+1}^{(1)}$ and the error terms $(y_{k+1} - m_{k+1} \hat{\theta}_k^{(1)})$ are identical to the corresponding terms in (5.8), provided $P_k^{(1)}$ are the same in both estimators. A schematic diagram of the various estimators obtained here is shown in Fig. 5.1.

It is worth noting at this point that the approach of Chapter 4, namely, joint detection and estimation based on a model for the switching process, can be taken for system identification as well by posing the identification problem as a state estimation problem and developing a Kalman filter for the estimation of $\theta^{(0)}$ and $\theta^{(1)}$. The necessary model for the switching process was obtained in Chapter 3 (equation 3.36). This approach,

however, does not possess the advantages cited in Chapter 3 since estimates of both parameter vectors $\theta^{(0)}$ and $\theta^{(1)}$ have to be updated at each sampling instant.

Consider an example for illustrating the applicability of the various approaches presented here.

Example 5.1. Uncorrelated Noise - Comparison of Estimators

A scalar system with uncorrelated error terms is modelled as:

$$y_{k+1} = \gamma_{k+1}(a_1 y_k + b_1 U_k) + (1 - \gamma_{k+1})(a_0 y_k + b_0 U_k) + e_{k+1} \quad (5.13)$$

It is required to estimate $\theta^{(0)} \Delta \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}$ and $\theta^{(1)} \Delta \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$.

The input-output data required for the identification experiment was obtained with the following parameters:

$$a_1 = 0.8, b_1 = 1.0, a_0 = -0.8, b_0 = 0.5, R = 0.1 \text{ and } y_0 = 1.$$

The statistics of γ are given as:

$$p(\gamma_0 = 1) = 0.9; p(\gamma_k \neq \gamma_{k-1}) = 0.1.$$

The input U_k , rather than being a PRBS (pseudo-random binary signal) as chosen by Astrom^(A5), is determined adaptively as in Appendix I. Briefly, the input U_k is chosen as +1 or -1 depending on which value results in a

larger signal to noise ratio. Initial estimates of the unknown parameters were all assumed to be zero.

Performance of the various estimators are compared on the basis of the mean squared estimation errors obtained by averaging the results of thirty (30) Monte-Carlo runs, each over 30 time steps. The γ -optimal estimator, obtained under the assumption of perfect knowledge of the γ -sequence, is seen to perform as expected, i.e. update $\hat{a}_0$ and $\hat{b}_0$ when γ_k is 0 and update $\hat{a}_1$ and $\hat{b}_1$ when γ_k is 1.

The decision theoretic approach by comparison gives inferior estimates. The obvious reason for this is the non-zero probability of error in the decision. However, for the low noise level under consideration, the decision theoretic approach provides estimates comparable in accuracy to the γ -optimal estimates.

The modified least squares (smoothing) approach appears to provide larger errors than the decision approach. This difference between the two approaches arises from that in the weights w_j in (5.10). While in the decision approach the weights $w_j \in \{0,1\}$, the weights in the modified least squares approach take values in the continuous range $(0,1)$. This implies that when the decision is correct (i.e. $p(\gamma_k=1 | y_k, u_{k-1}, I_{k-1}^*) \geq 0.5$ when $\gamma_k=1$), the decision

approach results in smaller errors in the estimate than with the modified least squares approach. At $k=10$ in Fig. 5.4, for example, all approaches result in similar errors owing to the fact that the a-posteriori probability density of γ is close to the extreme value (i.e., 0). At other sampling instants, such is not the case and, for this reason, the modified least squares approach results in larger errors than the decision approach. It was, however, noticed that as the noise level increased the decisions were increasingly in error thus providing inferior estimates compared to the smoothing approach. It thus appears that while, for small noise to signal ratios, almost any approach provides good estimates, as the noise level increases the modified least squares approach is to be preferred.

The approach of modelling the switching process along the lines of equation (3.36) and subsequent joint estimation of the parameters and γ via an extended Kalman filter appears to provide estimates whose accuracy, as seen from Figs. 5.2-5.5, lies between those of the decision and the modified least squares approaches. However, as cited earlier, the modelling approach does not provide any computational advantages over other approaches. In fact, the

computational effort required in this case is larger than with other approaches.

A usual characteristic of many identification schemes, as indicated in Astrom^(A7), is that the mean squared estimation error decreases with an increase in the number of observations processed. In other words, the Cramer-Rao lower bound for estimation accuracy becomes, for large N ^(A8),

$$E\{(\hat{\theta}_N - \theta)(\hat{\theta}_N - \theta)^T\} = \frac{1}{N} A,$$

where A is a constant matrix determined by the statistics of the input U_k . Precisely, the same result holds true for switching systems when the γ -sequence is known. The only difference between non-switching and switching systems is that only a part N_1 of the N observations is used to estimate the parameters of the i th model in a switching system. This characteristic is present in the modified least squares approach taken for this example as shown in Fig. 5.6. The reason for this is that for systems with a low switching probability ($q_{10}=0.1$, in this example), long strings of 1's and 0's are expected in the actual realization of the γ -sequence resulting in better estimates as the number of observations increases.

Consider next the problems that arise when the errors are correlated.

5.4.2 Correlated Noise

In this section, an attempt is made to generalize the results of the previous section to the case of correlated errors. Consider the difference equation model (5.5) with the constraints

$$c_k^{(1)} = c_k, \quad i=0,1, \quad k=1,2, \dots, n.$$

Assume $(N+n)$ observations are available. A compact representation of the model is then given as:

$$\underline{Y} = (I - \Delta) M \theta^{(0)} + \Delta M \theta^{(1)} + \underline{V} \quad (5.14)$$

$$\Delta = \text{diag}(y_{n+1}, \dots, y_{n+N})$$

$$M = [\underline{m}_k]$$

where

$$\underline{m}_k = [y_{k-1}, y_{k-2}, \dots, y_{k-n}, u_{k-1}, \dots, u_{k-n}]$$

$$\theta^{(i)} = [a_1^{(i)}, a_2^{(i)}, \dots, a_n^{(i)}, b_1^{(i)}, b_2^{(i)}, \dots, b_n^{(i)}]^T, \quad i=0,1$$

and $\underline{Y}$ and $\underline{V}$ are column vectors of the outputs and the errors respectively.

Even if Δ were known, the correlation between $\underline{V}$ and M results in the least squares estimates of $\theta^{(0)}$ and $\theta^{(1)}$ being biased estimates. There are several approaches to overcome or reduce this bias in the estimates (see, for

example, Astrom^(A7)), but for the purpose of this thesis, only one of these, namely, the instrumental variable (IV) approach^(R3) is considered.

Premultiplying (5.14) by Δ and $(I-\Delta)$ results in:

$$\Delta \underline{Y} = \Delta M \theta^{(1)} + \Delta \underline{V} \quad (5.15)$$

$$(I-\Delta) \underline{Y} = (I-\Delta) M \theta^{(0)} + (I-\Delta) \underline{V} \quad (5.16)$$

Notice that $\theta^{(1)}$ and $\theta^{(0)}$ are estimated from the vector equations (5.15) and (5.16) respectively. Consider the estimation of $\theta^{(1)}$ first.

Using a $2n \times N$ transformation $Z^{(1)T}$ which satisfies the conditions:

$$p \lim_{N \rightarrow \infty} \left[\frac{1}{N} Z^{(1)T} \Delta \underline{V} \right] = \underline{0} \quad (5.17)$$

and

$$p \lim_{N \rightarrow \infty} \left[\frac{1}{N} Z^{(1)T} \Delta \underline{M} \right] \quad (5.18)$$

exists as a non-singular matrix, the model (5.15) is transformed as:

$$Z^{(1)T} \Delta \underline{Y} = Z^{(1)T} \Delta M \theta^{(1)} + Z^{(1)T} \Delta \underline{V}$$

In view of (5.17) and (5.18), the IV estimator is given as:

$$\hat{\theta}_N^{(1)} = \left[Z^{(1)T} \Delta \underline{M} \right]^{-1} Z^{(1)T} \Delta \underline{Y} \quad (5.19)$$

It is readily shown that the condition (5.17) implies the asymptotic unbiasedness of the estimate $\hat{\theta}_N^{(1)}$. A similar set of conditions yield the IV estimator:

$$\hat{\theta}_N^{(0)} = \left[Z^{(0)T} (I - \Delta) M \right]^{-1} Z^{(0)T} (I - \Delta) Y \quad (5.20)$$

The matrix products $\Delta Z^{(1)}$ and $(I - \Delta) Z^{(0)}$ represent the IV matrices for the estimation of $\theta^{(1)}$ and $\theta^{(0)}$ respectively.

Before proceeding further, notice that simple algebraic manipulations lead to the following recursive form for the estimators (5.19) and (5.20):

For $i=0,1$,

$$\begin{aligned} \hat{\theta}_{N+1}^{(i)} &= \hat{\theta}_N^{(i)} + G_{N+1}^{(i)} \left[y_{N+1} - |1 - \gamma_{N+1}| m_{N+1} \hat{\theta}_N^{(i)} \right] \\ G_{N+1}^{(i)} &= \frac{|1 - \gamma_{N+1}| P_N^{(i)} Z_{N+1}^{(i)T}}{1 + m_{N+1} P_N^{(i)} Z_{N+1}^{(i)T} |1 - \gamma_{N+1}|} \\ P_{N+1}^{(i)} &= P_N^{(i)} - |1 - \gamma_{N+1}| G_{N+1}^{(i)} m_{N+1} P_N^{(i)} \end{aligned} \quad (5.21)$$

where m_{N+1} and $Z_{N+1}^{(i)}$ are the $(N+1)$ th rows of M_{N+1} and $Z_{N+1}^{(i)}$ respectively.

The estimator (5.21) is completely specified once the instruments generating the matrices $Z^{(i)}$ are chosen. One particular choice of instruments is considered in the sequel: the choice of estimates of the noise free outputs

as instruments. It is further assumed, with no loss of generality, that $Z^{(1)} \equiv Z^{(0)}$. With this assumption and the stated choice of instruments,

$$\underline{Z}_{N+1}^{(1)T} = \underline{Z}_{N+1}^T = [\hat{x}_N, \dots, \hat{x}_{N-n+1}, u_N, u_{N-1}, \dots, u_{N-n+1}]$$

where

$$\hat{x}_N = \gamma_N \underline{Z}_N^T \hat{\theta}_N^{(1)} + (1 - \gamma_N) \underline{Z}_N^T \theta_N^{(0)} \quad (5.22)$$

(5.21) and (5.22) together specify the IV estimators for $\theta^{(1)}$ and $\theta^{(0)}$. This estimator (based on the choice (5.22) as instruments) is referred to in the sequel as the two-stage least squares (2SLS) estimator.

An improvement in the accuracy of the estimates results from an improvement in the estimate $\hat{x}_N$ of (5.22). For constant systems, Pandya^(P1) suggests the following approach, called the Non-Adaptive Boot-Strap approach;

$$\hat{x}_N = \hat{x}_N' + \alpha_N (y_N - \hat{x}_N') \quad (5.23)$$

where

$$\alpha_N = \left[1 - \left(\frac{N-1}{N} \right)^2 \right]$$

and

$\hat{x}_N'$ is available from (5.22).

Clearly, $\alpha_N \equiv 0$ leads to the estimator (5.22) and $\alpha_N \equiv 1$ yields the ordinary least squares estimator (i.e. $Z=M$). In this

thesis, this approach is extended to the case of switching systems. The Boot-Strap estimator is given as estimator (5.21) with

$$\hat{x}_N = \left[\gamma_N \underline{Z}_N^T \hat{\theta}_N^{(1)} + (1-\gamma_N) \underline{Z}_N^T \hat{\theta}_N^{(0)} \right] \left[\frac{N-1}{N} \right]^2 + \left[1 - \frac{N-1}{N} \right]^2 y_N \quad (5.24)$$

The treatment so far has been for the case of perfectly known γ . Consider the modifications necessary when γ is not known. An approach identical to that taken in Section 5.4.1, i.e. the modified least squares, results in the following modified IV estimator:

For $i=0,1$,

$$\hat{\theta}_{k+1}^{(i)} = \hat{\theta}_k^{(i)} + p(\gamma_{k+1}=1 | I_k^*, y_{k+1}, U_k) G_{k+1}^{(i)} (y_{k+1} - m_{k+1} \hat{\theta}_k^{(i)})$$

$$G_{k+1}^{(i)} = \frac{P_k^{(i)} \underline{z}_{k+1}^T}{1 + m_{k+1} P_k^{(i)} \underline{z}_{k+1}^T}$$

$$P_{k+1}^{(i)} = P_k^{(i)} - p(\gamma_{k+1}=1 | y_{k+1}, U_k, I_k^*) G_{k+1}^{(i)} m_{k+1} P_k^{(i)}$$

with $\underline{z}_{k+1}$ defined by (5.22)

$$\text{and } \hat{x}_k = p(\gamma_k=1 | I_k^*) \underline{z}_k^T \hat{\theta}_k^{(1)} + p(\gamma_k=0 | I_k^*) \underline{z}_k^T \hat{\theta}_k^{(0)} \quad (2SLS)$$

$$\text{or } \hat{x}_k = \left[p(\gamma_k=1 | I_k^*) \underline{z}_k^T \hat{\theta}_k^{(1)} + p(\gamma_k=0 | I_k^*) \underline{z}_k^T \hat{\theta}_k^{(0)} \right] \left[\frac{k-1}{k} \right]^2 + \left[\frac{2k-1}{k^2} \right] y_k \quad (\text{Boot-Strap})$$

The estimator (5.25) is denoted modified two stage least squares (M2SLS) or modified Boot-Strap (MBS) estimator depending upon the approach taken to obtain f . Consider an example to illustrate the use of the estimator (5.25).

Example 5.2

A scalar system is represented by the difference equation model:

$$y_k = \gamma_k [a_k y_{k-1} + b_1 u_{k-1}] + (1 - \gamma_k) [a_0 y_{k-1} + b_0 u_{k-1}] + e_k + c_0 e_{k+1}$$

$$e_k \sim N(0, \lambda^2)$$

The true values of the parameters are:

$$\theta^{(1)} \triangleq [a_1 \ b_1]^T = [0.8 \ 1.0]^T$$

$$\theta^{(0)} \triangleq [a_0 \ b_0]^T = [-0.8 \ 0.5]^T$$

$$c_0 = 0.5, \lambda^2 = 0.5.$$

y_0 is given as 1 and the statistics of γ are given as:

$$p(\gamma_0=1)=0.9, p(\gamma_k=1|\gamma_{k-1}=1)=0.9\sigma_{1j}+0.1(1-\sigma_{1j}), \dots, 1,j=0,1$$

The input U_k is chosen as in the previous example. The initial estimates are all zero.

It is required to estimate $\theta^{(0)}$ and $\theta^{(1)}$ recursively.

Three approaches are considered:

1. Modified least squares (MLS) $\alpha_N \equiv 1$
2. Modified two stage least squares (M2SLS), $\alpha_N \equiv 0$
3. Modified boot-strap (MBS) $0 \leq \alpha_N \leq 1.$

Performances of the three estimators are compared on the basis of mean squared estimation errors computed by averaging the errors in 30 Monte-Carlo runs, each over 100 time steps. These are shown in Figs. 5.7-5.10. The steady state (at $k=100$) mean errors in the estimators are given in Table 5.2. The improvement in estimation accuracy resulting from the use of the MBS approach is obvious from this table. There is a relatively smaller bias in the estimates. The mean squared errors shown in Figs. 5.7-5.10 indicate the same conclusion. As expected, the ordinary least squares approach performs poorly and the modified least squares approach presents an improvement in estimation accuracy.

As in the previous example, the accuracy of estimation is expected to increase with an increase in the number of observations. When γ is known, the variance of the estimates is expected to tend to zero as $1/N$. That this indeed is the case is shown in Fig. 5.11, where the variance of the estimates is plotted for several values of N , the number of observations.

5.5 CONCLUSIONS

An exposition of the various approaches to the estimation of the parameters of switching systems was presented in this Chapter. In particular, a simple difference equation model was considered. A simple modification of the ordinary least squares approach was observed to give good estimates of parameters when the errors in the model were uncorrelated. A similar modification of the bootstrap estimator used by Pandya^(P1) was seen to provide the best estimates in the case of correlated errors.

Extension of these results to the general difference equation model (equation 5.4) is conceptually simple but requires a large increase in the computation time. This is because the smoothing of the estimates (as in MLS) has to be done over all possible realizations of the past n values of y . In the particular case where difference equation formulation is not possible, for example, missing observations, the only possible approach appears to be the joint estimation of the parameters and the states via an extended Kalman filter. An example of this was given in the previous Chapter.

TABLE 5.1
Mean and mean squared errors at the end of 30 time steps.
(Averaged over 30 runs) for Example 5.1.

APPROACH	MEAN ERROR				MEAN SQUARED ERROR			
	a_1	b_1	a_0	b_0	a_1	b_1	a_0	b_0
γ -optimal	5×10^{-3}	5.1×10^{-3}	3.8×10^{-3}	9×10^{-2}	4×10^{-3}	3×10^{-3}	1.5×10^{-3}	10^{-2}
Decision	7.3×10^{-3}	6.7×10^{-3}	11.3×10^{-3}	12×10^{-2}	6×10^{-3}	5×10^{-3}	3×10^{-3}	1.5×10^{-2}
Modelling	8.13×10^{-3}	5.2×10^{-3}	14×10^{-3}	20.8×10^{-2}	10^{-2}	6×10^{-3}	10^{-2}	5×10^{-2}
Modified Least Squares (Smoothing)	11.2×10^{-3}	8.6×10^{-3}	12×10^{-3}	32×10^{-2}	1.9×10^{-2}	9×10^{-3}	1.3×10^{-2}	1.5×10^{-1}

TABLE 5.2
Mean and mean squared errors at the end of 100 time steps
(Averaged over 30 runs) for Example 5.2.

APPROACH	MEAN ERRORS				MEAN SQUARED ERRORS			
	a_1	b_1	a_0	b_0	a_1	b_1	a_0	b_0
MOLS	.1665	-.2464	.1275	.3550	.1437	.6750	.3473	.9483
M2SLS	.1881	.151	.0677	.2879	.1052	.4111	.1486	.9255
MBS	-.0264	.1054	-.0316	.0851	.0036	.0385	.0022	.0110

FIG. 5.1. The structure of the on-line identification scheme for switching systems.

FIG. 3.2. Mean squared errors in the estimation of the parameter ' a_1 ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal approaches.

FIG. 5.3. Mean squared errors in the estimation of the parameter ' b_1 ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal approaches.

FIG. 5.4. Mean squared errors in the estimation of the parameter ' a_0 ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal approaches.

FIG. 5.5. Mean squared errors in the estimation of the parameter ' b_0 ' with uncorrelated noise - comparison of smoothing, modelling, decision and γ -optimal.

FIG. 5.6. Variance of estimates vs no. of observations in the smoothing approach for Example 5.1.

FIG. 5.11. Variance of estimates vs no. of observations for Example 5.2 with boot-strap estimator.

CHAPTER 6

CONTROL OF SWITCHING SYSTEMS

6.1 INTRODUCTION

o In designing a feedback controller for a system when the system and/or its environment are not perfectly known, one is required to perform two tasks:

- a 1) Learning - improving the knowledge of the system and its environment. For example, estimates of the states and any unknown parameters may be needed.
- ii) Control - referring to the design of a suitable controller such that the system response follows a desired trajectory.

For a linear system with known parameters, the Separation Theorem^(K6) shows that these two tasks can be performed independent of each other. The design of the controller in this case is done after replacing all random variables in the system description by their expected values. The optimal controller is seen to consist of an estimator, which provides estimates of the random variables, followed by a controller. This leads to the famous "certainty equivalent principle" (see Aoki^(A5) p. 52) or "Separation Theorem" (see Kushner^(K6), p. 253).

The class of systems considered in this thesis, do not, in general, have this property. A discussion of this is presented in Section 6.3. Looking ahead briefly, this lack of separation between estimation (or learning) and control stems from the uncertainty in the 'mode' of the system (or the value of the switching variable γ). This observation puts the problem within the framework of the 'dual control' problems discussed by Feldbaum^(F1). His solution to the dual control problem demands an exorbitant computational effort (see Mendes^(M2)), thus requiring the development and the use of sub-optimal controllers.

In the sequel, two typical cases of switching systems are considered: (i) systems with missing observations (the certainty equivalence (CE) principle does not hold in this case) and (ii) systems with switching time constants, a case where the CE principle is valid. These two cases are discussed in Sections 6.4 and 6.5 respectively. In each case, following the discussion of the difficulties involved in realizing the optimal controller, sub-optimal controllers are designed and their performance compared with that of the γ -optimal controller (designed under the assumption of perfect knowledge of past values of γ). Section 6.6 summarizes the conclusion drawn from the use of sub-optimal controllers. In the following section, the

statement of the problem and the approach to the solution via dynamic programming are presented.

6.2 THE PROBLEM

A stochastic switching system is assumed to be modelled by the following equations, as before:

$$\begin{aligned}x_{k+1} &= \gamma_{k+1}(A_1 x_k + B_1 U_k) + (1 - \gamma_{k+1})(A_0 x_k + B_0 U_k) + w_k \\y_{k+1} &= \gamma_{k+1}(C_1 x_{k+1}) + (1 - \gamma_{k+1})(C_0 x_{k+1}) + v_k \\ \gamma_{k+1} &= 0 \text{ or } 1\end{aligned}\tag{6.1}$$

The matrices A_i , B_i , C_i , $i=0,1$, the distribution functions for w_k and v_k and the statistics of the switching process $\{\gamma_k\}$ are assumed known. The performance of the system is measured by the cost J :

$$J = \sum_{i=1}^N W_i(x_i, U_{i-1}), \quad W_i \geq 0\tag{6.2}$$

It is required to obtain an optimal control policy which minimizes the expected value, EJ , of the cost J . It is further required that the optimal control policy belong to a class in which each U_i is a function of all previous and current observations $y_0, y_1, \dots, y_i$ and of whatever information is available at the beginning about the stochastic processes $\{w_k\}$, $\{v_k\}$ and $\{\gamma_k\}$. Thus, it is

required to find the functions ϕ_i such that

$$\phi_i(Y^i) = U_i, \quad 0 \leq i \leq N-1$$

where ϕ_i is measurable and $Y^i \triangleq \{y_0, y_1, \dots, y_i\}$.

To the author's best knowledge, no solution is as yet available for the general problem posed here. However, several special cases have been treated in the literature. In the case of only the plant switching, i.e. $C_0 \equiv C_1 = C$, Aoki^(A5) obtained the optimal Bayes' controller. His solution is difficult to implement in practice owing to the growing memory requirement in the estimation of the states. This fact is established by the simple scalar example considered by him. In the case of missing observations, i.e. $A_0 \equiv A_1 = A$, $B_0 \equiv B_1 = B$, $C_0 = \phi$, the null matrix, Fujita and Fukao^(F4) obtained the optimal controller by purporting the separation between the tasks of estimation and control, a claim since disproved by Wernersson^(W1) (see also Fujishige^(F3)). The continuous counterpart of the problem posed here, under the assumption of perfect information on the states and the switching variable γ , was solved by Sworder and Robinson^(S5). Rosenbrock^(R2) obtained the solution to a special case of the systems studied by Sworder, namely, the situation where the A

matrix switches only once during the control experiment.

Aoki's solution can be generalized to obtain the optimal control for the system (6.1) as follows:

Define

$$J_k = \sum_{i=k}^N E(W_i(x_{ij}, U_{i-1}) | Y^{i-1}) \quad (6.3)$$

and

$$J_k^* = \min_{U_{k-1}, U_k, \dots, U_{N-1}} J_k$$

It is seen that minimizing EJ where J is given by (6.2) is equivalent to minimizing J_0 for all possible realizations of Y^N . From the principle of optimality^(A5), it follows that:

$$J_k^* = \min_{U_{k-1}} \{E(W_k | Y^{k-1}, U^{k-1}) + E[J_{k+1}^* | Y^{k-1}, U^{k-1}]\} \quad (6.4)$$

with the terminal condition

$$J_{N+1}^* = 0$$

Determination of the optimal control U_{k-1}^* by performing the minimization in (6.4) becomes difficult, in general, owing to the non-linearity of the second expectation in the control U_{k-1} . This effect appears mainly due to the dependence of the one-step-ahead estimation error covariance, $P_{k|k}$ on the control U_{k-1} . This dependence and the special situations where it vanishes are discussed next.

6.3 EFFECT OF CONTROL ON ESTIMATION ACCURACY

For the system (6.1), the estimation error covariance $P_{k+1|k+1}$, which reflects the accuracy of estimation or learning, is a function of the past controls as follows:

Define the estimation error at time $(k+1)$ as

$$\tilde{x}_{k+1} \triangleq x_{k+1} - \hat{x}_{k+1|k+1}$$

x_{k+1} is obtained from (6.1) and $\hat{x}_{k+1|k+1}$ was obtained in Chapter 2 as

$$\hat{x}_{k+1|k+1} = \int_{I_{k+1} \in \Omega} E(x_{k+1} | Y^{k+1}, I_{k+1}) P(I_{k+1} | Y^{k+1})$$

Consider a typical sequence $\bar{I}_{k+1}$:

$$\bar{I}_{k+1} = \{Y_0 = \bar{Y}_0, Y_1 = \bar{Y}_1, \dots, Y_{k+1} = \bar{Y}_{k+1}\}$$

Then

$$\begin{aligned} \hat{x}_{k+1|k+1}(\bar{I}_{k+1}) &\triangleq E(x_{k+1} | Y^{k+1}, \bar{I}_{k+1}) \\ &= \hat{x}_{k+1|k}(\bar{I}_{k+1}) + G_{k+1}(\bar{I}_{k+1}) \\ &\quad \left[y_{k+1} - \{C_1 \bar{Y}_{k+1} + C_0(1 - \bar{Y}_{k+1})\} \hat{x}_{k+1|k}(\bar{I}_{k+1}) \right] \end{aligned} \quad (6.5)$$

where $G_{k+1}(\bar{I}_{k+1})$ represents the Kalman gain conditioned on $\bar{I}_{k+1}$ and $\hat{x}_{k+1|k}(\bar{I}_{k+1})$ is given as:

$$\begin{aligned}\hat{x}_{k+1|k}(\bar{I}_{k+1}) &= \bar{\gamma}_{k+1} \left[A_1 \hat{x}_{k|k}(\bar{I}_{k+1}) + B_1 U_k \right] \\ &+ (1 - \bar{\gamma}_{k+1}) \left[A_0 \hat{x}_{k|k}(\bar{I}_{k+1}) + B_0 U_k \right]\end{aligned}$$

The error $\tilde{x}_{k+1}$ can now be obtained as:

$$\tilde{x}_{k+1} = \sum_{I_{k+1} \in \Omega} \hat{x}_{k+1}(I_{k+1}) p(I_{k+1} | Y^{k+1})$$

$$\text{where } \tilde{x}_{k+1}(I_{k+1}) = x_{k+1} - \hat{x}_{k+1|k+1}(I_{k+1}) \quad (6.6)$$

Substitution of (6.1) and (6.5) in (6.6) reveals that for the typical sequence $\bar{I}_{k+1}$, the conditional error $\tilde{x}_{k+1}(\bar{I}_{k+1})$ and hence its covariance is a function of U_k . It can similarly be shown that the a-posteriori density $p(\bar{I}_{k+1} | Y^{k+1})$ is dependent on U_k . This indicates that both the estimation error $\tilde{x}_{k+1}$ and its covariance $P_{k+1|k+1}$ are functions of U_k . In fact, this dependence is exploited in the derivation of best inputs for estimation (see Appendix I).

It is readily shown that in switching systems in which either one of the two matrices B, C switches, the estimation error is always a function of the past controls. The reason for this is the difference between the actual and the assumed sequence of γ 's. This difference appears as the coefficient of the input U_k in the expression for

$\bar{x}_{k+1}$. Obviously, if the two matrices C and B do not switch or if the γ -sequence is known exactly, the accuracy of estimation does not depend on past controls. For example, separation between estimation and control exists in the case of switching time constants ($A_0 \neq A_1$, $C_0 \equiv C_1$, $B_0 \equiv B_1$) and in the case of switching environments (see Ackerson and Fu^(A1)). In the particular case of missing observations, considered next, this separation is absent, thus making it difficult to realize the optimal controller.

6.4 CONTROL OF SYSTEMS WITH MISSING OR INTERRUPTED OBSERVATIONS

Under the constraints $A_0 \equiv A_1 = A$, $B_0 \equiv B_1 = B$, $C_1 = C$ and $C_0 = \phi$, the null matrix (6.1) represents a system with missing or interrupted observations:

$$x_{k+1} = A x_k + B U_k + w_k \quad (6.7)$$

$$y_{k+1} = \gamma_{k+1} C x_{k+1} + v_{k+1} \quad (6.8)$$

The control inputs U_k , $k=0,1, \dots, N-1$ are to be chosen so as to minimize the expected value of the quadratic cost:

$$J = \frac{1}{2} (x_N - d_N)^T S_N (x_N - d_N) + \frac{1}{2} \sum_{k=0}^{N-1} \{ (x_k - d_k)^T S_k (x_k - d_k) + U_k^T T_k U_k \} \quad (6.9)$$

Consider the case where, at time k , γ_k and the past values of γ are known exactly. The resulting controller, called the γ -optimal controller, and the certainty equivalent controller (derived under the assumption of separation between estimation and control) are discussed next.

6.4.1 γ -Optimal and Certainty Equivalent Controllers

When the switching variable γ is known, the system of equations (6.7) and (6.8) together represent the familiar linear, discrete-time, stochastic system with known, time varying parameters. The optimal control policy for this type of system is well known (see, for example, Bryson and Ho^(B2), p. 429) and is given as:

$$U_k^* = -\Lambda_k \hat{x}_k | k + \Delta_k$$

$$\Lambda_k = [T_k + B^T W_{k+1} B]^{-1} B^T W_{k+1} A$$

$$\Delta_k = -[T_k + B^T W_{k+1} B]^{-1} B^T V_{k+1}$$

$$W_k = A^T \left[I - (T_k + B^T W_{k+1} B)^{-1} W_{k+1} B B^T \right] W_{k+1} A + S_k$$

$$V_k = A^T \left[I - (T_k + B^T W_{k+1} B)^{-1} W_{k+1} B B^T \right] V_{k+1} - S_k d_k$$

$$W_N = S_N; V_N = -S_N d_N$$

$$\hat{x}_k | k \triangleq E(x_k | Y^k)$$

(6.9)

The control policy derived by Fujita and Fukao^(F4), which in effect is the certainty equivalent control, is identical to (6.9). However, as indicated in Chapter 2 (section 2.4), the minimum mean squared error estimate $\hat{x}_{k|k}$ can not be obtained in practice and, for this reason, an approximate estimate is used in (6.9) to calculate the certainty equivalent controls, U_k^{CE} . Thus, the only difference between the γ -optimal and the certainty equivalent (CE) controllers lies in the estimate $\hat{x}_{k|k}$.

The non-realizability of the optimal state estimates augmented by the observation in the previous section that the estimation error covariance $P_{k|k}$ is a function of the past controls, makes the CE control only an approximation to the γ -optimal control. The accuracy of the CE control depends on that in the estimation of γ . This follows from the discussion in Chapter 2 (Section 2.4) that the accuracy of the state estimate depends on that in the evaluation of $\hat{\gamma}_{k|k}$, the estimate of γ_k . Moreover, the sensitivity of the error covariance $P_{k|k}$ to the past control U_{k-1} decreases with an increase in the accuracy of $\hat{\gamma}_{k|k}$. These observations suggest that, in practice, the CE control may need to be modified when faced with inaccurate estimates of γ . A possible approach is to modify the CE control by the addition of a correction term which reflects the inaccuracy in $\hat{\gamma}_{k|k}$.

6.4.2 Adaptive Control

Wernersson^(W1) suggests the following adaptive approach: "Use the CE control under the assumption that $\hat{\gamma}_{k|k} \Delta p(\gamma_k = 1 | Y^k)$ is close to the stationary value of γ_k . If, now, a perceptible change occurs in the a-posteriori density of γ_k , there is an indication that a jump might have occurred in γ_k . In this case, use a large control to aid in the identification of γ_k ".

This approach raises several questions:

- What is the measure of a perceptible change in $\hat{\gamma}_{k|k}$?
- By what amount should the control be increased?
- Does this increase result in a significant improvement in the accuracy of $\hat{\gamma}_{k|k}$?

Answers to these questions are difficult to provide. Consider an approach which provides intuitive answers to these questions. This approach is motivated by the counter example presented by Wernersson^(W1), where it was observed that when the objective is to control only the terminal state, the initial controls can be large resulting in good estimation of γ and hence x . However, Wernersson chose the initial controls arbitrarily. In the sequel, a possible approach to the choice of the initial controls is presented.

The estimate $\gamma_{k|k}^* \triangleq E(\gamma_k | Y^k) = p(\gamma_k = 1 | Y^k)$ is unbiased, i.e. $E(\gamma_{k|k}^* - \gamma_k) = 0$. Thus, the accuracy of the estimate is determined by its covariance, $E[(\gamma_{k|k}^* - \gamma_k)^2 | Y^k]$. The estimate $\gamma_{k|k}^*$ was obtained in Chapter 3 (equation 3.7) as:-

$$\gamma_{k|k}^* = \frac{p(y_k | Y^{k-1}, \gamma_k = 1) p(\gamma_k = 1 | Y^{k-1})}{\sum_{i=0}^1 p(y_k | Y^{k-1}, \gamma_k = i) p(\gamma_k = i | Y^{k-1})}$$

Following Lainiotis^(L2), the expected value of the standard deviation of $\gamma_{k|k}^*$ is given as:

$$\sigma(\gamma_{k|k}^*) = \sqrt{p_1(1-p_1)\rho}$$

where ρ is the Bhattacharyya coefficient^(K1)

$$\rho = \int \sqrt{f_1(y_k) f_0(y_k)} dy_k$$

$$f_i(y_k) \triangleq p(y_k | Y^{k-1}, \gamma_k = i), \quad i=0,1$$

and $p_1 \triangleq p(\gamma_k = 1 | Y^{k-1})$.

The densities $f_i(y_k)$ are approximately Gaussian (see Moose and Wong^(M5)) with means μ_i and covariances σ_i^2 given as:

$$\mu_1 = 1 - C \hat{x}_{k|k-1}$$

$$\text{and } \sigma_1^2 = 1 - C \left[A P_{k-1|k-1} A^T + Q_{k-1} \right] C^T + R_k$$

where $\hat{x}_{k|k-1} = A \hat{x}_{k-1|k-1} + B U_{k-1}$

In this case, the Bhattacharyya coefficient is given as (L1)

$$\rho = e^{-\beta} \quad (6.10)$$

where

$$\begin{aligned} \beta = & \frac{1}{4} [\mu_1 - \mu_0]^T \left[\frac{\sigma_1^2 + \sigma_0^2}{2} \right]^{-1} [\mu_1 - \mu_0] \\ & + \frac{m}{2} \ln \left[\frac{|\sigma_1^2 + \sigma_0^2|}{2 \sqrt{|\sigma_1^2| |\sigma_0^2|}} \right] \end{aligned}$$

where $|(.)|$ stands for the determinant and m is the dimension of the output vector y_k .

Consider a requirement of the estimate $\gamma_{k|k}^*$:

$$\sigma(\gamma_{k|k}^*) = \sqrt{p_1(1-p_1)}\rho \leq \epsilon \quad (6.11)$$

where ϵ is a chosen "tolerance limit".

In the case of a scalar input U_k , the constraint (6.11) can be transformed to a magnitude constraint on the input or control by substituting for ρ from (6.10). This leads to a control constraint of the form

$$|U_{k-1}| \geq \delta \quad (6.12)$$

where δ is obtained as in Appendix I.

In the multi-input case, general solutions of the form (6.12) are not possible as the controls are required to lie on the surface of an ellipsoid (see Appendix I). The control satisfying (6.12) is denoted U_{k-1}^0 . The following adaptive control for the system (6.7) is then proposed:

$$\begin{aligned} U_{k-1}^{ad} &= U_{k-1}^{CE}, \text{ if } |U_{k-1}^{CE}| \geq \delta \\ &= U_{k-1}^0 \text{ otherwise.} \end{aligned} \quad (6.13)$$

Implementation of the adaptive control (6.13) requires choosing a proper value for the parameter ϵ . Consider the example presented in Wernersson^(W1).

Example 6.1

Consider the scalar system:

$$x_{k+1} = x_k + u_k + w_k$$

$$y_k = \gamma_k x_k + v_k$$

with $x_0=0$, $P_0|_0=Q_0=Q_2=0$, $Q_1=100$, $R=1$ and it is required to minimize the expected cost:

$$J_3 = E \left[100x_3^2 + 0.01(u_0^2 + u_1^2) + u_2^2 \right]$$

γ_k is assumed to be Markov on $\{-1, +1\}$ rather than $\{0, 1\}$.

The initial probabilities of γ , i.e. $p(\gamma_0 = \pm 1)$ are assumed equal (0.5). The switching probability q_{10} is assumed to be 0. This requires only simple modifications of the adaptive control (6.12); for example, the a-posteriori conditional densities of y_k , $f_1(y_k)$, are now defined for $i = +1, -1$, instead of $i = 0, 1$.

Wernersson^(W1) has shown that application of the CE control (Fujita and Fukao^(F4)) to this problem, even with optimal estimates of the states, leads to $U_k = 0$, $k = 0, 1, 2$, and the expected cost is $J_{CK} = 10000$. In an attempt to establish that the CE control is not the optimal control, Wernersson chose the controls arbitrarily as:

$$u_0 = 5, u_1 = -5, u_2 = -y_2 \text{sign}(y_1)$$

and showed that the expected cost with this control was less than 202.

The system was simulated with the adaptive control (6.13) (modified to account for the change in the possible values of γ) and the expected cost was obtained as the average of the costs of 100 Monte-Carlo runs, with ϵ as a parameter. It was observed that an ϵ of 0.01 resulted in an average cost of 205, a result close to that of Wernersson, while as ϵ was increased, the performance of the adaptive controller deteriorated and approached that of

the CE controller. Moreover, as ϵ was decreased from 0.01, the performance again deteriorated. These conclusions are not surprising since a small value of ϵ (say less than 0.01) results in a large control U_{k-1}^0 which, though providing very good estimates of x_k and γ_k , results in a poor controller performance. This is exactly the situation in any dual control problem where the control effort has to be judiciously divided between estimation and control requirements. On the other hand, it is noticed that an upper bound for ϵ is 0.5 as both ρ and p_1 lie in the range (0,1). Thus, for $\epsilon \geq 0.5$, U_k^{CE} always satisfies the constraint (6.12) and hence the adaptive control becomes identical to the CE control.

A further validation of these conclusions is provided by the example considered below.

Example 6.2. Scalar System with Missing Observations

Consider a scalar, discrete-time, stochastic system with missing observations, represented by the equations:

$$x_{k+1} = 0.95x_k + U_k + w_k \quad (6.14)$$

$$y_{k+1} = \gamma_{k+1} x_{k+1} + v_{k+1}$$

$$w_k \sim N(0, 0.5), \quad v_k \sim N(0, 0.1), \quad x_0 \sim N(0, 1).$$

γ_k has the following statistics:

$$p(\gamma_0=1)=0.9, \quad p(\gamma_k \neq \gamma_{k-1})=0.1.$$

It is required to determine the control policy which minimizes the expected value of the cost:

$$J = 100(x_{50}-1)^2 + \sum_{k=0}^{49} 100(x_k-1)^2 + U_k^2 \quad (6.15)$$

Three approaches to the solution are compared with the -optimal control:

- (i) U_k^{CES} : certainty equivalent control with state estimates obtained by the Sawargi approach (see Chapter 2)
- (ii) U_k^{CEM} : CE control with state estimates provided by the modelling approach of Chapter 4, with Q_ψ chosen as in Appendix II
- (iii) U_k^{ad} : adaptive control (6.13).

The expected cost was obtained by averaging the results of 30 Monte-Carlo runs. The average performances of the various control schemes are summarized in Table 6.1. Typical control and state histories are compared in Figs. 6.1 and 6.2. From Table 6.1, the improvement resulting from the use of the adaptive control is obvious. The average and the maximum costs as well as the mean squared deviation of the cost are all smaller than the corresponding quantities with U_k^{CES} . From Fig. 6.1, it is noticed that

U_k^{ad} and U_k^{CES} are very close when the estimates of γ are reasonably accurate, while as the accuracy of the estimate $\hat{\gamma}_{k|k}$ decreases, the adaptive control scheme applies a larger control effort. This increase in control effort is seen from Fig. 6.1 to provide better estimates of γ_k . The adaptive control histories shown in Fig. 6.1 are for $\epsilon=0.274$ or $\epsilon^2=0.075$.

Variation of the average cost (with adaptive control) with respect to ϵ is shown in Fig. 6.3. For the system under consideration, $\epsilon^2=0.075$ gave the minimum average cost. A smaller value of ϵ resulted in larger average costs while, as ϵ was increased to 0.5, the adaptive controller approached the CE controller.

An interesting feature of the optimal value of ϵ ($\epsilon=0.274$) is that it is very close to the standard deviation of $\hat{\gamma}_{k+1|k}$, the a-priori estimate of γ_{k+1} , when $\hat{\gamma}_{k|k}$ is very accurate. In other words, when $\hat{\gamma}_{k|k}$ is close to 1 or 0, $\hat{\gamma}_{k+1|k}$ is close to 0.1 or 0.9 respectively and the standard deviation of $\hat{\gamma}_{k+1|k}$ is:

$$\sigma(\hat{\gamma}_{k+1|k}) = 0.9 \times 0.1 = 0.3.$$

With $\epsilon=0.3$, the average cost is only marginally larger than that with $\epsilon=0.274$. This observation suggests an empirical choice of ϵ as

$$\varepsilon = P(\gamma_k \neq \gamma_{k-1}) P(\gamma_k = \gamma_{k-1}) \quad (6.16)$$

This choice must, however, be tested on other examples.

The use of U_x^{CEM} , CE control with a model for γ_k , resulted in a performance close to that of U_k^{CES} . This is indicated by Fig. 6.3 and Table 6.1. Fig. 6.4 compares the control histories and the estimates of γ obtained via the two CE control approaches for a typical run. The closeness of the performance of the system with U_k^{CES} and U_k^{CEM} is not surprising since, with CE control, accuracy of the control depends on that of the state estimates and the results of Chapter 4 indicated that the estimates obtained with the modelling approach compare very well with those obtained via other sub-optimal approaches.

6.5 STOCHASTIC CONTROL OF SYSTEMS WITH SWITCHING TIME CONSTANTS

Imposing the constraints $B_0 \equiv B_1 = B$ and $C_0 \equiv C_1 = C$ transforms the system (6.1) to a system with switching time constants:

$$x_{k+1} = \gamma_{k+1} (A_1 x_k) + (1 - \gamma_{k+1}) A_0 x_k + B U_k + w_k \quad (6.17)$$

$$y_{k+1} = C x_{k+1} + v_{k+1}$$

The notation in (6.17) is identical to that in (6.1).

Equation (6.17) then represents a system whose coefficient matrix A (or some of its components) switches randomly between two values A_0 and A_1 . The value taken by the switching variable γ_{k+1} determines the "mode" or the value of the coefficient matrix A (A_1 or A_0).

It is required to design a controller which results in a control sequence $U^{N-1} = \{U_0, U_1, \dots, U_{N-1}\}$, such that the expected value of the cost J :

$$J = \sum_{k=1}^{N-1} (x_k - d_k)^T S_k (x_k - d_k) + U_k^T T_k U_k + (x_N - d_N)^T S_N (x_N - d_N) \quad (6.18)$$

is minimized. The matrices S_k and T_k are assumed known.

Following the arguments of Section 6.3, it is seen that the error covariance $P_{k+1|k+1}$ is independent of past controls U^k . This implies 'separation' between estimation and control. In spite of this feature of the problem, solution becomes difficult owing to the uncertainty regarding future plant dynamics. The only available solution to the problem posed here is due to Aoki^(A6). The Bayesian approach taken by him involves evaluating the optimal cost to go for all possible future dynamics of the plant and averaging them. This requirement is identical to the growing memory requirement of the optimal

(min. MSE) state estimators (see Chapter 2). The simple, scalar, three stage ($N=3$) example considered by Aoki^(A6) emphasizes this point. This observation indicates that the optimal controller is difficult to design, especially for long control periods (N large), and, for this reason, suitable sub-optimal controllers need to be arrived at. Consider, first, the controller designed under the assumption that, at time k , y_k and past values of y are known exactly. The controller arrived at under this assumption is denoted the y -optimal controller.

6.5.1 y -Optimal Controller

At time k , it is required to minimize the cost:

$$J_k = E \sum_{j=k}^N W_j \quad (6.19)$$

where

$$W_j \triangleq (x_j - d_j)^T S_j (x_j - d_j) + U_{j-1}^T T_{j-1} U_{j-1}$$

and $U_N \equiv 0$.

$$\text{Now, } J_k = E_{y^{k-1}} \left[E \sum_{j=k}^N (W_j | y^{k-1}, I_{k-1}) \right]$$

where

$$y^{k-1} \triangleq \{y_0, y_1, \dots, y_{k-1}\}$$

$$I_{k-1} \triangleq \{y_0, y_1, \dots, y_{k-1}\}$$

Minimizing J_k is thus equivalent to minimizing the term within braces for all possible Y^{k-1} . (I_{k-1} is assumed known exactly.) Thus, it is required to obtain U_{k-1}^* , the optimal control, by minimizing J_k^* :

$$J_k^* = E \left(\sum_{j=k}^N W_j | Y^{k-1}, I_{k-1} \right)$$

Consider the last stage.

$$J_N^* = E(W_N | Y^{N-1}, I_{N-1})$$

The optimal control U_{N-1}^* , as seen below, depends on the value of Y_{N-1} . For this reason, the two cases, viz.

$Y_{N-1}=i$, $i=0,1$, are considered. Define:

$\lambda_{k,i} \triangleq$ the one stage expected cost with $Y_{k-1}=i$

$$\triangleq E(W_k | Y^{k-1}, I_{k-2}, Y_{k-1}=i)$$

$$J_{k,i}^* \triangleq E \left(\sum_{j=k}^N W_j | Y^{k-1}, I_{k-2}, Y_{k-1}=i \right)$$

The minimal value of $J_{k,i}^*$ is denoted as $J_{k,i}^{**}$.

The optimal control $U_{N-1,i}^*$ is obtained by minimizing $J_{N,i}^*$ (or $\lambda_{N,i}$) with respect to U_{N-1} .

Now, $\lambda_{N,i} = E(W_N | Y^{N-1}, I_{N-2}, Y_{N-1}=i)$

$$= \sum_{j=0}^1 E(W_N | Y^{N-1}, I_{N-2}, Y_{N-1}=i, Y_N=j) P(Y_N=j | Y_{N-1}=i)$$

(6.20)

Defining

$$\hat{x}_{k,i} = E(x_k | Y^k, \gamma_k = i),$$

the expectation in the expression (6.20) is given as:

$$\begin{aligned} & E(W_N | Y^{N-1}, I_{N-2}, \gamma_{N-1} = i, \gamma_N = j) \\ &= E\{(x_N - d_N)^T S_N (x_N - d_N) + U_{N-1}^T T_{N-1} U_{N-1} | Y^{N-1}, I_{N-2}, \gamma_{N-1} = i, \gamma_N = j\} \\ &= (A_j \hat{x}_{N-1,i} + B U_{N-1} - d_N)^T S_N (A_j \hat{x}_{N-1,i} + B U_{N-1} - d_N) \quad (6.21) \\ &+ U_{N-1}^T T_{N-1} U_{N-1} + \text{terms independent of } U_{N-1}. \end{aligned}$$

$$\text{Also, } p(\gamma_N = j | \gamma_{N-1} = i) = q_{ji} \quad (6.22)$$

Substitution of (6.21) and (6.22) in (6.20) and minimization of (6.20) with regard to U_{N-1} yields:

$$U_{N-1,i}^* = -\Lambda_{N-1,i} \hat{x}_{N-1,i} + \Delta_{N-1,i} \quad (6.23)$$

where

$$\Lambda_{N-1,i} = \left[B^T S_N B + T_{N-1} \right]^{-1} B^T S_N \sum_{j=0}^1 A_j q_{ji} \quad (6.24)$$

$$\Delta_{N-1,i} = \left[B^T S_N B + T_{N-1} \right]^{-1} B^T S_N d_N \quad (6.25)$$

Substituting (6.23) in (6.20), the optimal value of $\lambda_{N,i}$ is given as:

$$\lambda_{N,i}^* = \lambda_{N,i}^{**} = \hat{x}_{N-1,i}^T G_{N-1,i} \hat{x}_{N-1,i} + G_{N-1,i}^T \hat{x}_{N-1,i} + H_{N-1,i} \quad (6.26)$$

where

$$F_{N-1,i} = \sum_{j=0}^1 q_{ji} \left[(A_j - BA_{N-1,i})^T S_N (A_j - BA_{N-1,i}) + \Delta_{N-1,i}^T T_{N-1} \Delta_{N-1,i} \right]$$

$$G_{N-1,i}^T = 2 \sum_{j=0}^1 \left[(BA_{N-1,i} - d_N)^T S_N (A_j - BA_{N-1,i}) - \Delta_{N-1,i}^T T_{N-1} \Delta_{N-1,i} \right] q_j$$

$$H_{N-1,i} = \Delta_{N-1,i}^T T_{N-1} \Delta_{N-1,i} + (BA_{N-1,i} - d_N)^T S_N (BA_{N-1,i} - d_N)$$

Equation (6.26) indicates that both $\lambda_{N,0}^*$ and $\lambda_{N,1}^*$ are quadratic functions in the state estimates $\hat{x}_{N-1,0}$ and $\hat{x}_{N-1,1}$ respectively. This observation inspires the assumption that $J_{k,i}^{**}$ is also quadratic in $\hat{x}_{k-1,i}$. This assumption will be justified by induction.

Let

$$J_{k,i}^{**} \triangleq \min_{U_k, U_{k+1}, \dots, U_{N-1}} E \left(\sum_{j=k+1}^N W_j \mid Y^k, I_{k-1}, Y_{k-1} \right) \\ = \hat{x}_{k,i}^T F_{k,i} \hat{x}_{k,i} + G_{k,i}^T \hat{x}_{k,i} + H_{k,i} \quad (6.27)$$

By definition,

$$F_{N,i} = 0$$

$$G_{N,i} = 0$$

$$H_{N,i} = 0$$

for $i=0,1$.

Now, from the principle of optimality,

$$\begin{aligned}
 J_{k,i}^{**} &= \min_{U_{k-1}} \left[\lambda_{k,i} + E(J_{k+1,m}^{**} | Y^{k-1}, I_{k-2}, \gamma_{k-1}=1) \right] \\
 &= \min_{U_{k-1}} \left[\lambda_{k,i} + \sum_{m=0}^1 E(J_{k+1,m}^{**} | Y^{k-1}, I_{k-2}, \gamma_{k-1}=1, \right. \\
 &\quad \left. \gamma_{k-m}=m) q_{mi} \right] \quad (6.28)
 \end{aligned}$$

Substitution for $\lambda_{k,i}$ and $J_{k+1,m}^{**}$ and performing the necessary minimization leads to:

$$U_{k-1,i}^* = -\Lambda_{k-1,i} \hat{x}_{k-1,i} + \Delta_{k-1,i} \quad (6.29)$$

where

$$\begin{aligned}
 \Lambda_{k-1,i} &= \left[T_{k-1} + \sum_{m=0}^1 q_{mi} \{ B^T (F_{k,m} + S_k) B \} \right]^{-1} B^T \\
 &\quad \left[\sum_{m=0}^1 (F_{k,m} + S_k) A_m q_{mi} \right] \quad (6.30)
 \end{aligned}$$

$$\begin{aligned}
 \Delta_{k-1,i} &= \left[T_{k-1} + \sum_{m=0}^1 q_{mi} \{ B^T (F_{k,m} + S_k) B \} \right]^{-1} B^T \\
 &\quad \left[S_k d_k - \frac{1}{2} \sum_{m=0}^1 G_{k,m} q_{mi} \right] \quad (6.31)
 \end{aligned}$$

Substitution of the control (6.29) in (6.28) indicates that $J_{k,i}^{**}$ is indeed quadratic in $\hat{x}_{k-1,i}$. Comparison of the resulting expression for $J_{k,i}^{**}$ with the quadratic form (6.27) (with $k+1$ replaced by k) results in the following

recursive relation for F and G:

$$F_{k-1,i} = \sum_{m=0}^1 \left[(A_m - B\Lambda_{k-1,i})^T (F_{k,m} + S_k) (A_m - B\Lambda_{k-1,i}) + \Lambda_{k-1,i}^T T_{k-1} \Lambda_{k-1,i} \right] q_{mi} \quad (6.32)$$

$$G_{k-1,i}^T = 2 \sum_{m=0}^1 \left[(B\Delta_{k-1,i} - d_k)^T (F_{k,m} + S_k) (A_m - B\Lambda_{k-1,i}) + (d_k^T F_{k,m} + \frac{1}{2} G_{k,m}^T) (A_m - B\Lambda_{k-1,i}) - \Delta_{k-1,i}^T T_{k-1} \Delta_{k-1,i} \right] q_{mi} \quad (6.33)$$

Thus, the -optimal control is given by the relations (6.29) to (6.33) with the terminal conditions:

$$F_{N,i} = 0$$

$$G_{N,i} = 0$$

$$i=0,1.$$

6.5.2 Adaptive Control

In practice, γ is seldom known exactly. Moreover, observations of Chapter 2 indicate that the conditional mean state estimates can not be evaluated in practice. These two observations together imply that the controller defined by (6.29)-(6.33) needs to be modified. It is noticed

that if the conditional mean state estimates were available and the estimates of γ exact, i.e.

$$p(\gamma_k = k | Y^k, U^{k-1}) = 1 \quad \text{for } k=0 \text{ or } 1,$$

the above control is still optimal. Thus, the choice between $U_{k,0}^*$ and $U_{k,1}^*$ at stage k depends on the decision regarding the value of γ_k and the accuracy of the control depends on that of the decision.

Rather than choosing between $U_{k,0}^*$ and $U_{k,1}^*$ depending on γ_k (or its estimate), consider the following adaptive control:

$$U_k^* = \sum_{i=0}^1 U_{k,i}^* p(\gamma_k = i | Y^k, U^{k-1}) \quad (6.34)$$

The control U_k^* is obtained as the weighted sum of $U_{k,0}^*$ and $U_{k,1}^*$, the weights being the a-posteriori probabilities of γ_k . Notice that if the a-posteriori probabilities of γ_k tend to 1 or 0, the control (6.34) approaches the γ -optimal control. Moreover, for a non-switching system, i.e. say $\gamma_k = 1, \forall k$, $U_k^* \equiv U_{k,1}^*, \forall k$. The control (6.34) is identical in structure to the adaptive control postulated by Deshpande et al for time invariant systems with a discrete parameter space. The control (6.34) can also be written as:

$$U_k^* = U_{k,1}^* \hat{\gamma}_k | k + U_{k,0}^* (1 - \hat{\gamma}_k | k) \quad (6.35)$$

where

$$\begin{aligned}\hat{\gamma}_{k|k} &\triangleq E(\gamma_k | Y^k, U^{*k-1}) \\ &= p(\gamma_k = 1 | Y^k, U^{*k-1})\end{aligned}$$

The a-posteriori probability density of γ_k is given in Chapter 2. Consider, next, the use of a model for γ .

6.5.3 Modelling Approach

There are two possible approaches to the control of the system (6.17) when a model describing the time-evolution of γ is postulated:

- (i) Use the control (6.35) with estimates of the states and the switching variable γ provided by the modelling approach taken in Chapter 4. Based on the conclusions of Chapter 4, that estimates provided by the modelling approach compare reasonably well with those obtained by other sub-optimal approaches, use of the control (6.35) with $\hat{\gamma}$ and $\hat{x}$ provided by either of the approaches is expected to result in very similar performance of the system.
- (ii) Given a model for γ , future values of the coefficient matrix A can be obtained by extrapolation. The design of the controller for the resulting non-linear system is straightforward (see, for example, Tse et al. (11)). In other words, defining

the coefficient matrix at time k , $A(k)$, as

$$A(k) = A_1 \gamma_k + A_0 (1 - \gamma_k)$$

future values of A , i.e. $A(k+j)$, $j > 0$, are obtained by extrapolating the expectations as:

$$\begin{aligned} A(k+j) &= E(A(k+j) | \gamma^k, U^{*k-1}) \\ &= A_1 P(\gamma_{k+j} = 1 | \gamma^k, U^{*k-1}) + A_0 P(\gamma_{k+j} = 0 | \gamma^k, U^{*k-1}) \end{aligned}$$

The model for γ basically extrapolates the probability density of γ_k to obtain the densities of γ_{k+j} .

Consider the two models postulated in Chapter 3,

viz.,

$$\gamma_{k+1} = (1 - 2q_{10})\gamma_k + q_{10} \quad (6.36)$$

$$\gamma_{k+1} = f(\gamma_k) \quad (6.37)$$

where q_{10} is the switching probability, i.e. $q_{10} = P(\gamma_k \neq \gamma_{k-1})$.

It was observed in Chapter 3 that the stationary points of the non-linear model is close to the extreme values (0 or 1) while the only stationary point for (6.36) is at 0.5. This means that for a sufficiently large j , $A(k+j) \approx \frac{1}{2} (A_1 + A_0)$ with a linear model. For a scalar system with $a_0 = -a_1$, for example, this implies $a_{k+j} \approx 0$. On the other

hand, the non-linear model (6.37) indicates that, for $q_{10} < 0.5$, $\gamma_{k+j} = \gamma_{k+j-1}$ or $A(k+j) = A(k+j-1)$. This result is meaningful in the sense that the joint probability of γ_j , $k \leq j \leq N$, is maximum when $\gamma_{k+j} = \gamma_{k+j-1}$ (for $q_{10} < 0.5$). A similar conclusion is arrived at for $q_{10} > 0.5$. In the sequel, the control derived with (6.37) is denoted U_k^{MOD} .

Consider an example to illustrate the use of the adaptive control and the control with a model for γ .

Example 6.2. Scalar System with Switching Time Constant

The system:

$$x_{k+1} = 0.95\gamma_{k+1}x_k - 0.8(1-\gamma_{k+1})x_k + U_k + w_k$$

$$y_{k+1} = x_{k+1} + v_{k+1}$$

$$w_k \sim N(0, 0.5)$$

$$v_k \sim N(0, 0.25)$$

$$x_0 \sim N(d, 2.0)$$

The objective: Obtain U_k for $k=0$ to 50 such that the expected cost

$$J = E \{ 10(x_{50}-d)^2 + \sum_{k=0}^{49} 10(x_k-d)^2 + U_k^2 \}$$

is minimized.

Two values of d were chosen: $d=1$ and $d=5$. The expected cost was obtained by averaging the costs of 50th Monte-Carlo runs. The following approaches are compared:

- (i) U_k^{opt} : γ -optimal control
- (ii) U_k^{ad} : adaptive control with state and γ estimate obtained via Bruckner and Scott approach
- (iii) U_k^{mod} : control with model for γ , with $Q_\psi=0.1$.

Figs. 6.5 and 6.6 compare the typical state trajectories and control histories respectively for the three controllers considered here ($d=1$). It is observed that the control history for the control U_k^{mod} follows the U_k^{opt} control except for an increase in magnitude. There is a close resemblance in the state trajectories obtained via either of the three approaches. Table 6.2 summarizes the costs involved in the use of the three control approaches. It is observed that the modelling approach compares very well with the adaptive controller. Notice also that, as d increases to 5, i.e. the control of the state is desired around a point in state space further away from the origin, the adaptive controller as well as the control U_k^{mod} result in a performance closer to that with U_k^{opt} . The reason for this is obvious. As the state moves farther away from the

origin, the signal to noise ratio improves and this results in a better detection of γ than with $d=1$. The large control used in the modelling approach can be thought of as compensating for the error in the extrapolation of plant dynamics, and thus result in "good" estimation of γ . Needless to say, both U_k^{ad} and u_k^{mod} result in larger costs than the γ -optimal for the simple reason that in the former case an extra control effort is needed to aid in the identification of γ .

6.6 CONCLUSIONS

This Chapter presented the problem of controlling a system with two possible modes of operation. Two special cases were considered, viz., systems with missing observations and systems with switching time constants. In the case of missing observations, an adaptive control was developed based on the requirement that the expected standard deviation of the estimate of γ be less than a specified limit. A judicious choice of this limit results in an adaptive control which performs better than the certainty equivalent control.

In the case of switching time constants, the γ -optimal control was derived and two approximate control approaches were suggested: (1) adaptive control based on averaging the controls of the two modes and (2) control developed

for the non-linear system resulting from the use of a model for γ . The two control approaches were observed to perform equally well for a simulated system. They approached the γ -optimal control when the state was controlled about a point far removed from the origin of the state space. Thus, the results of this Chapter indicate that the performance of the approximate controllers is linked to the accuracy in the estimation of γ , or, in other words, the signal to noise ratio both at the plant and at the output.

TABLE 6.1.

Comparison of costs with the various approaches to the control of the system with missing observations in Example 6.2.

APPROACH	Y-OPTIMAL u_k^{opt}	ADAPTIVE u_k^{ad} ($\epsilon^2=0.075$)	CE CONTROL u_k^{CES}	CE CONTROL u_k^{CEM}
Average Cost	5267.72	5901.74	7046.52	7526.3
Maximum Cost in 30 runs	11484.8	12742.8	14990.7	16529.9
Rms Deviation of Cost	1292.17	1431.94	1476.0	1510.3

TABLE 6.2
Comparison of costs for the modelling, the adaptive and the γ -optimal approaches to the control of a system with a switching time constant.

d_k		γ -OPTIMAL	ADAPTIVE	MODELLING
$d_k=1$	Max. Cost	1624.36	2267.23	2413.31
	Std. Dev. of Cost	203.4	321.6	343.1
	Av. Cost	905.192	1206.74	1212.43
$d_k=5$	Max. Cost	8052.64	9801.84	9886.7
	Std. Dev. of Cost	1201.7	1473.8	1475.1
	Av. Cost	6500.91	7107.56	7116.19

FIG. 6.1. Typical control histories for Example 6.2.

FIG. 6.2. State trajectories for Example 6.2.

FIG. 6.3. Variation of adaptive controller performance with ϵ .

FIG. 6.4. Control histories and estimates of y in a typical Monte-Carlo run of the scalar system with missing observations (comparison of CE controls with modelling and Saveragi approaches).

FIG. 6.5. State trajectories in a typical Monte-Carlo run of the scalar system with a switching time constant.

FIG. 6.6. Comparison of control histories and the estimates of γ in a typical Monte-Carlo run of the system with a switching time constant.

CHAPTER 7

ERROR COMPENSATION VIA SWITCHING NOISE

7.1 INTRODUCTION

The proceeding Chapters presented the analysis problem, i.e. the study of systems which are inherently of the switching type. This Chapter exposes the design problem, namely, the effect of introducing a switching feature into a system which is represented by a single model, on the estimation errors. To be specific, the use of a switching plant noise in the prevention or limiting of divergence in the Kalman filter is studied. The occurrence of divergence in the Kalman filter is indicated by the estimation error covariance becoming unbounded (i.e. approaching infinity) or reaching a large steady state value. Several excellent articles on this feature of the Kalman filter have appeared in the literature (see for example Fitzgerald^(F2) or Schlee et al^(S4)) and, for this reason, only a brief outline of this phenomenon and means of preventing it are presented in the sequel. It is shown that the approach of Quigley^(Q1) for limiting the divergence can be viewed as a switching environment problem (Ackerson and Fu^(A1)). A simple example is presented in Section 7.3 to evaluate the

relative performances of the approaches taken by Quigley, and Ackerson and Fu and the modelling approach of Chapter 4. A more detailed example, viz., the compensation for the errors in the modelling of boiler dynamics^(D1) is considered in Section 7.4.

7.2 DIVERGENCE OF THE KALMAN FILTER

Often, after an extended period of operation of a Kalman filter when employed with an inaccurate model of the system, the estimation errors eventually diverge to values entirely out of proportion to the root mean square values predicted by the filtering equations^(F2). This phenomenon is termed "Divergence". One explanation offered for this phenomenon is that the computed covariance matrix becomes unrealistically small, leading to undue confidence in the model and subsequent measurements are effectively ignored. This results in the errors being propagated according to the plant equation and, if in addition, the model is in error, the estimation errors either become unbounded or bounded but too large to allow the estimates to be of any use. Based on this, Fitzgerald^(F2) defines two types of divergence:

1. true or mathematical divergence in which the rms errors can actually be shown to approach infinity with increasing time

2. apparent or practical divergence in which the error covariance reaches a steady state but the associated errors are too large.

In the sequel, only practical divergence is considered and the term divergence henceforth refers only to this type. Several approaches to the prevention of this undesirable phenomenon have been reported in the literature. Many of these are discussed in Jazwinski^(J2). An approach which has attracted considerable attention is termed 'adaptive estimation'. This approach is based on estimating, recursively, the variance of a fictitious plant noise added to the model. The logic behind this approach is the observation that modelling errors manifest themselves as errors in the prediction of the output (often called the residuals) and that, by processing one or more samples of the residuals, the plant noise covariance required to 'bury' the modelling errors can be estimated. A considerable saving in the computational effort required for adaptive estimation can be achieved by limiting the search of the suitable noise covariance to the choice between two assumed values. This is the approach taken by Quigley^(Q1). A discussion of his approach and its correspondence to the switching environment problem studied by Ackerson and Fu^(A1) follows.

7.3 SWITCHING PLANT NOISE

Consider a discrete-time, linear, dynamic system modelled by the equations:

$$x_{k+1} = A x_k + B U_k \quad (7.1)$$

$$y_k = C x_k + v_k \quad (7.2)$$

$$v_k \sim N(0, R)$$

It is assumed that the model (7.1) is in error, i.e., some elements of the matrices A and B are in error. Compensation for this error is done by adding a noise term w_k so that the model becomes:

$$x_{k+1} = A x_k + B U_k + w_k \quad (7.3)$$

w_k is assumed to have a zero-mean, Gaussian distribution, independent of v_k , and with a covariance of $Q(k)$. The problem is to determine a value for $Q(k)$ suitable for the purpose, i.e. sufficient enough to 'bury' the modelling errors. The Kalman filter for the system (7.2, 7.3) is given as:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + G_{k+1} (y_{k+1} - C \hat{x}_{k+1|k})$$

$$\hat{x}_{k+1|k} = A \hat{x}_{k|k} + B U_k$$

$$G_{k+1} = P_{k+1|k} C^T [C P_{k+1|k} C^T + R]^{-1}$$

$$P_{k+1|k} = A P_{k|k} A^T + Q(k)$$

$$P_{k+1|k+1} = P_{k+1|k} - G_{k+1} C P_{k+1|k} \quad (7.4)$$

As indicated before, a small value for $Q(k)$ leads to small error covariance matrix $P_{k+1|k}$ thus presenting the possibility of divergence of the filter (7.4). Quigley's (Q1) approach for the determination of $Q(k)$ was outlined in Chapter 3.

In simple terms, one starts to use a large Q ($Q=Q_1$) if the observed errors are larger than three standard deviations and continues to do so until the errors fall below one standard deviation, in which case the small value of Q ($Q=Q_0$) is used. The decision to use either Q_0 or Q_1 , at stage k , is thus based on the filter performance, measured by the normalized forecast errors J :

$$J = (y_{k+1} - C \hat{x}_{k+1|k})^T (C P_{k+1|k} C^T + R)^{-1} (y_{k+1} - C \hat{x}_{k+1|k}) \quad (7.5)$$

This approach can be reformulated in terms of a switching variable γ as follows:

Assume that the system (7.3) is modelled as:

$$x_{k+1} = A x_k + B u_k + \gamma_{k+1} \xi_k + (1 - \gamma_{k+1}) \eta_k \quad (7.6)$$

$$y_{k+1} = C x_{k+1} + v_{k+1}$$

where $\eta_k \sim N(0, Q_0)$, $\xi_k \sim N(0, Q_1)$ and $\gamma_{k+1} = 1$ or 0.

The choice of $Q(k)$ then reduces to the choice of γ_{k+1} .

Define

$$P_{k+1|k}(\gamma_{k+1}) = A P_{k|k} A^T + \gamma_{k+1} Q_1 + (1 - \gamma_{k+1}) Q_0 \quad (7.7)$$

and

$$J(\gamma_{k+1}) = ||y_{k+1} - C \hat{x}_{k+1|k}||^2 [C P_{k+1|k}(\gamma_{k+1}) C^T + R]^{-1} \quad (7.8)$$

Quigley's approach can then be reformulated, in terms of γ as,

1. Let $\gamma_{k+1} = \gamma_k$.
2. Compute $P_{k+1|k}(\gamma_{k+1})$ and $J(\gamma_{k+1})$.
3. If $J(\gamma_{k+1}) > 9$, $\gamma_{k+1} = 1$.
4. If $J(\gamma_{k+1}) < 1$, $\gamma_{k+1} = 0$.
5. Compute $P_{k+1|k}(\gamma_{k+1})$ and use this as the prediction error covariance in the filter equations (7.4).

or simply,

$$\gamma_{k+1} = \frac{1}{2} + \frac{1}{2} \operatorname{sgn} [J(\gamma_k) - 9 + 8\gamma_k] \quad (7.9)$$

$$P_{k+1|k} = A P_{k|k} A^T + \gamma_{k+1} Q_1 + (1 - \gamma_{k+1}) Q_0 \quad (7.10)$$

Notice that γ_k in the above formulation is simply an indicator variable and the switching from γ_k to γ_{k+1} is deterministic (equation 7.9). Moreover, γ_{k+1} given by (7.9) takes only 0 or 1 values.

Consider now the framework of Ackerson and Fu^(A1). They attack the same problem, namely, state estimation with the model (7.6). However, γ_k is now assumed to be a Markov chain defined by the transition probabilities q_{ij} :

$$p(\gamma_k = i | \gamma_{k-1} = j) = q_{ij}, \quad i, j = 0, 1 \quad (7.11)$$

Following Ackerson and Fu, the minimum mean squared error estimate of γ_{k+1} , based on the observations up to time $(k+1)$, $\hat{\gamma}_{k+1|k+1}$, is seen to be

$$\hat{\gamma}_{k+1|k+1} = E(\gamma_{k+1} | Y^{k+1}) = P(\gamma_{k+1} = 1 | Y^{k+1})$$

Using Bayes' rule and the approximation of the density of the state (see Chapter 2), this estimate is given by:

$$\hat{\gamma}_{k+1|k+1} = \left[1 + \frac{1 - \hat{\gamma}_{k+1|k}}{\hat{\gamma}_{k+1|k}} \sqrt{\frac{|C P_{k+1|k}(1) C^T + R|}{|C P_{k+1|k}(0) C^T + R|}} \exp \left\{ -\frac{1}{2} J(0) + \frac{1}{2} J(1) \right\} \right]^{-1} \quad (7.12)$$

where $J(1)$, $P_{k+1|k}(1)$, $i=0,1$ are defined by (7.7) and (7.8) respectively.

The estimate of γ_{k+1} , given by (7.12), takes values in the continuous range $(0,1)$, rather than the extreme values as in (7.9). Another difference between the approaches of Quigley and Ackerson and Fu lies in the evaluation of

the state estimates. As indicated in Chapter 2, the Ackerson and Fu estimator provides the state estimate as the weighted sum of the estimates provided by two Kalman filters (equation (2.30), Chapter 2). This indicates that the Ackerson and Fu approach would require approximately twice the computational effort demanded by the approach of Quigley.

The modelling approach taken in Chapter 4, by contrast, requires only one Kalman filter but requires the estimation of an additional variable (γ_k). The demand on the digital computer presented by the modelling approach is thus slightly larger than that of Quigley's approach. The model function was given in Chapter 3 (equation (3.33), see also Fig. 3.3) as:

$$\begin{aligned}\hat{\gamma}_{k+1|k} &= f(\hat{\gamma}_{k|k}, \hat{x}_{k|k}) \\ &= \left[1 + \frac{1 - \bar{\gamma}_{k+1}}{\bar{\gamma}_{k+1}} \sqrt{\frac{|C P_{k+1|k}(1)C^T + R|}{|C P_{k+1|k}(0)C^T + R|}} \right]^{-1} \\ \bar{\gamma}_{k+1} &= q_{11}\hat{\gamma}_{k|k} + q_{10}(1 - \hat{\gamma}_{k|k})\end{aligned}\quad (7.13)$$

This model fails when the system under consideration has only the plant noise switching, as is the case here. The reason for this is the fact that the model (7.13) is independent of the state estimate $\hat{x}_{k|k}$. This, together with

the observation that γ_k is not observed at the output, results in a zero Kalman gain in the estimator for γ_k and the filter thus fails to track variations in the value of γ_k . The filter can be 'opened' up by introducing a fictitious, non-zero, Kalman gain. Consider a different approach. A Taylor series expansion of (7.12) about the point γ_{k+1}^* ,

$$y_{k+1}^* = C \hat{x}_{k+1|k}$$

and discarding 3rd and higher order terms results in an estimate $\hat{\gamma}_{k+1|k+1}$:

$$\begin{aligned} \hat{\gamma}_{k+1|k+1} = & [1+D]^{-1} + D[1+D]^{-2} (y_{k+1} - C \hat{x}_{k+1|k})^T \\ & \left[\{C P_{k+1|k}(0)C^T + R\}^{-1} - \{C P_{k+1|k}(1)C^T + R\}^{-1} \right]^{-1} \\ & (y_{k+1} - C \hat{x}_{k+1|k}) \end{aligned} \quad (7.14)$$

$$D = \frac{1 - \bar{\gamma}_{k+1}}{\bar{\gamma}_{k+1}} \sqrt{\frac{|C P_{k+1|k}(1)C^T + R|}{|C P_{k+1|k}(0)C^T + R|}}$$

Equation (7.14) is of the form:

$$\hat{\gamma}_{k+1|k+1} = k_1 + k_2 (J(0) - J(1)) \quad (7.15)$$

In addition, the estimate of γ_{k+1} needs to be in the range (0,1) and hence the estimate $\hat{\gamma}_{k+1|k+1}$ takes extreme values

if (7.15) results in values outside the range. Differences between (7.15) and (7.9) are obvious: the former requires the evaluation of both $J(0)$ and $J(1)$ while the latter requires calculating either $J(0)$ or $J(1)$ depending on the value of γ_k . Moreover, (7.15) gives $\hat{\gamma}_{k+1|k+1}$ in the continuous range $(0,1)$ rather than the extreme values as in (7.9). As seen in the example considered below, this results in $Q(k)$ taking values in the range (Q_0, Q_1) .

Example 7.1

For illustration, consider a simple situation wherein the true system equation is:

$$x_{k+1} = x_k + 1 + w_k \quad (7.16)$$

where x_k is a scalar state variable. The measurement equation is:

$$y_k = x_k + v_k \quad (7.17)$$

It is assumed that the Kalman filter is developed with a model of the system given as:

$$x_{k+1} = x_k + \xi_k \quad (7.18)$$

This example is from Quigley (Q1).

Given $w_k \sim N(0, Q_0)$, $v_k \sim N(0, R)$, $x_0 \sim N(\hat{x}_0, P_0)$ and assuming that $\xi_k \sim N(0, Q(k))$, it is required to estimate $Q(k)$ so that the

errors in estimation with the model (7.18) are bounded.

The system was simulated with the following data:

$$R=2.0; Q_0=0.01; x_0=0; P_0=1.01.$$

In order to compare the performances of the modelling approach and the approach of Quigley, the following data is assumed:

$$Q_1=Q_0+R=2.01; q_{10}=q_{01}=0.1; q_{11}=q_{00}=0.9.$$

Performances of the two approaches are compared on the basis of the calculated mean and mean squared estimation errors, obtained as averages of 50 Monte-Carlo runs, each over 30 time steps. Fig. 7.1 shows running values of the state and state estimates in a typical Monte-Carlo run. It is seen that the standard filter (with $Q_{(k)}=Q_0=0.01$) diverges. By comparison, both Quigley's and the modelling approach perform very well and the estimates are seen to be acceptable. However, Quigley's approach, which corresponds to a 'Bang-Bang' excitation of the filter, results in large oscillations in the estimation error. The modelling approach, on the other hand, presents a smoother variation in $Q_{(k)}$ and results in smaller oscillations of the estimation error. The mean and the mean squared errors shown in Fig. 7.2 indicate that the errors with the modelling

3

OF/DE

3

approach are far smaller than those with Quigley's approach. The choice of the transition probabilities did not appear to be crucial in this example, as similar results were obtained with $q_{11}=q_{10}=0.5$. However, it is emphasized that more experimentation is needed before any ground rules for the choice q_{ij} can be laid down.

7.4 MODELLING BOILER DYNAMICS

One of the first attempts at modelling the dynamics of a boiler appears to be due to Chien et al^(C1). Their analysis indicated that the boiler dynamics can be represented by seventeen (17) linear, ordinary, first order differential equations. Several publications have since appeared^(A3,D1,N2) on the reduction of the order of the model representing boiler dynamics, so that the design of a suitable controller would be easier. Discussion of the various approaches taken for model reduction is beyond the scope of this dissertation. However, it is noticed that all the reduced order models are approximations and hence there is a modelling error introduced, however small. In the presence of input and measurement disturbances, these modelling errors might lead to the divergence of the Kalman filter implemented for the estimation of the state vector. In the sequel, the third order model obtained by Davison^(D1) is considered and it is shown that

the introduction of a switching plant noise vector can substantially reduce the errors in the estimation of the states with this model.

Starting from the ninth order model developed by Nicholson^(N2), Davison^(D1) obtained a third order model by retaining only the first three dominant eigenvalues and the corresponding eigenvectors of the ninth order model. Thus, the original and the reduced models are given as:

$$\text{Original } \dot{x} = Ax + BU \quad 9 \text{ equations} \quad (7.19)$$

$$\text{Reduced } \dot{x}^* = A^* x^* + B^* U \quad 3 \text{ equations} \quad (7.20)$$

The matrices A and B are given in Tables 7.1a and 7.1b, and A^*, B^* are given in Table 7.2. The state and the input vectors of the two models are defined as:

$$x^T = (\rho_s, T_s, T_{ws}, x, w, p, T_{wb}, T_w, y)$$

$$x^{*T} = (T_s, p, y)$$

$$U^T = (w_s, w_f, w_i)$$

where

w_s : output steam flow

w_f : input fuel flow

w_i : input water flow

ρ_s : density of output steam

T_s : temperature of output steam

T_{ws} : temperature of superheater

x : quality of steam

w : water flow in riser

p : pressure

T_{wB} : temperature of riser

T_w : temperature of water in boiler

y : water level.

It is assumed that noisy measurements of the water level (y) and the pressure (p) are available:

$$z = Cx + v \quad (7.21)$$

where

$$z^T = (p \ y)$$

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$v \sim N(0, R)$$

In order to study the use of Quigley's approach and the modelling approach for reduction in estimation errors, the models (7.19) and (7.20) were discretized by sampling every 1/2 sec. (This corresponds to less than half the smallest time constant of the two models.) The resulting discretized models are:

$$\text{Original } x_{k+1} = A_d x_k + B_d U_k \quad (7.22)$$

$$\text{Reduced } x_{k+1}^* = A_d^* x_k^* + B_k^* U_k \quad (7.23)$$

and the measurement equation is:

$$z_k = C x_k + v_k \quad (7.24)$$

The suffix d denotes the discrete equivalents of the corresponding matrices of the continuous time model. The matrices A_d , B_d are given in Tables 7.3a and 7.3b and A_d^* , B_d^* in Table 7.4.

It is assumed that the initial states x_0 , x_0^* are both zero, i.e., they are assumed perfectly known implying that

$$\hat{x}_{0|0} = \hat{x}_{0|0}^* = 0$$

$$P_{0|0} = P_{0|0}^* = 0$$

where $\hat{x}$ and P are the mean and the covariance of the state vector. The inputs w_1 , w_f and w_g are assumed to be step functions with magnitudes (see Chien et al (C1)):

$$w_1 = 32.6 \text{ lbs/sec}$$

$$w_f = 2.38 \text{ lbs/sec}$$

$$w_g = 32.6 \text{ lbs/sec}$$

Measurement errors in the variables p and y are assumed normally distributed with zero means, and ± 20 points corresponding to ± 200 psf and ± 0.02 ft respectively. This gives the covariance matrix R of the noise v_k as:

$$R = \begin{bmatrix} 10^4 & 0 \\ 0 & 10^{-4} \end{bmatrix}$$

It is required to compare the performances of the Kalman filters developed for the original and the reduced models. It is immediately noticed that in the absence of any noise added to the models (7.22) and (7.23), and with a zero covariance of the initial state, the Kalman filters ignore the observations (Kalman gain is zero).

The state estimates then evolve according to the model equations (7.22) and (7.23) respectively. Running values of these estimates are compared in Figs. 7.3a and 7.3b, for the three state variables T_s , p and y . From this Figure, the errors resulting from model reduction are seen to be very large (approximately 10% in T_s and p and almost 100% in y).

The modelling errors and the resulting estimation errors can be reduced by the use of a switching plant noise, as in the previous example. Consider the case of a constant noise covariance first. In order to get some insight into the choice of this noise covariance, a Monte-

Carlo study was conducted with several values of Q , the noise covariance. The best Q was chosen as the one leading to minimum mean squared errors (averaged over 10 Monte-Carlo runs, each over 200 time steps). The variation of the mean squared errors with Q (the three diagonal elements varied independently) is shown in Figs. 7.4a,b,c. The best Q is seen from these figures to be

$$Q = Q^* = \text{diag}(10, 10^4, 10^{-4}).$$

For Quigley's approach and the modelling approach, the two values of Q were taken as:

$$Q_0 = \text{diag}(0., 0., 0.) \quad (7.25)$$

$$Q_1 = Q^* = \text{diag}(10, 10^4, 10^{-4}) \quad (7.26)$$

Ten (10) Monte-Carlo runs of both the reduced and the original systems were obtained each over 2000 time steps (corresponding to 16.6 minutes of real time). Three approaches are compared based on the mean and the mean squared errors:

- 1) $Q = \text{constant} = Q^*$
- 2) Quigley's approach with $Q_0=0$ and $Q_1=Q^*$
- 3) Modelling approach with the same two values of Q , and with $q_{10}=0.1$, $q_{11}=0.9$.

The mean squared errors resulting from the use of these three approaches are compared in Figs. 7.5a-c. Figs. 7.3a-7.3c indicate that in the absence of a plant noise, the steady state mean squared errors in the estimates of the three states under consideration are:

$$p:0.4 \times 10^7, T_s:100, y:0.25$$

From Figs. 7.5(a-c), it is seen that even a constant non-zero Q ($Q=Q^*$), substantially reduces the errors. In fact, at the end of 10 minutes in the system response, the errors are:

$$p:2 \times 10^4, T_s:30, y:4 \times 10^{-4}.$$

By comparison, both the modelling approach and the method of Quigley result in still smaller errors. The reduction in the mean squared errors is seen to be almost one order of magnitude compared to those with a constant Q . It is noticed, however, that the modelling approach does not provide a significantly superior performance compared to Quigley's approach as in the first example considered. It is quite possible that with a different choice of the transition probabilities, the modelling approach may give improved results.

7.5 CONCLUSIONS

This Chapter presented an investigation of the use of a switching plant noise in the reduction of estimation errors accruing from an erroneous model. It was shown that in spite of the arbitrariness involved in the choice of the two levels of the plant noise covariance and in the choice of the transition probabilities, the estimation errors with a switching plant noise were far smaller in comparison to those with a constant noise covariance (0 or non-zero). It was, however, noticed that an extensive investigation is necessary to lay down ground rules for the choice of the transition probabilities. In addition to justifying the necessity for analyzing switching systems, this Chapter has presented a potential use of these types of systems and models.

TABLE 7.1a

MATRIX A

-3.925042 0 0	-0.00315463 0.0000402537	0 0	0 0
367.8026 0 0	-3.051455 -0.00377204	3.030725 0	0 0
27.43095 0 0	0.0786945 -0.000281321	-0.0596402 0	0 0
-0.0646776 -0.00000335212 0	-0.0000519827 0.000000360429	0 0.00006335696	-0.2558487 0.0001939473
3851.207 -2.911237 0	17.25898 -0.1051962	-12.81674 12.72677	-12617.88 43.08814
22434.67 -0.000104 0	18.03119 -0.4136099	0 90.07017	-35.601 56.89565
0 0 0	0 0.000222186	0.002339145 -0.202623	0 0
0 0.00100585 0	0 0.00007862347	0 0	-1.272016 -0.0717133
-2.20356 -0.000110586 0	-0.001771046 0.00001383153	0 0.001489088	-8.440267 0.006018683

NOTE: Elements of matrix are in order row by row.

TABLE 7.1b

MATRIX B

-1.0032×10^{-2}	0	0
0	0	0
0	1.5615	0
0	0	-5.1296×10^{-6}
5.2020×10^1	8.2755	-1.5541
0	0	-1.77793
0	2.3259	0
0	0	-2.4541×10^{-2}
0	0	2.9417×10^{-5}

TABLE 7.2.

THE MATRIX A*

-8.3025×10^{-3}	6.6985×10^{-7}	0
4.0560×10^{-1}	-8.4537×10^{-3}	0
-1.6992×10^{-6}	4.8129×10^{-8}	0

THE MATRIX B*

-7.6478×10^{-2}	1.3759	7.2218×10^{-4}
-1.5595×10^1	3.7869×10^2	-8.5899
-1.0960×10^{-4}	-2.2608×10^{-3}	2.5943×10^{-4}

TABLE 7.3a

MATRIX A_d

.2712320	.2013630x10 ⁻³	-.7127585x10 ⁻³	-.2133357x10 ⁻³	0
.5579213x10 ⁻⁸	.6897696x10 ⁻⁵	.2802738x10 ⁻³	.1801262x10 ⁻³	0
-.1726302x10 ²	.1760247	.7314068	.1748117x10 ⁻¹	0
.1242953x10 ⁻⁵	.2286472x10 ⁻³	-.2527889x10 ⁻¹	-.1602897x10 ⁻¹	0
.6521819x10 ¹	.1216239x10 ⁻¹	.9895985	.1532009x10 ⁻²	0
-.6851791x10 ⁻⁷	-.6282239x10 ⁻⁴	-.1972952x10 ⁻²	-.1271633x10 ⁻²	0
-.1203594x10 ⁻¹	-.2187018x10 ⁻⁶	-.8537396x10 ⁻⁵	.8832635	0
-.7465998x10 ⁻⁶	-.1118671x10 ⁻⁸	.2663654x10 ⁻⁴	.7879459x10 ⁻⁴	0
.3049082x10 ³	-.4486928	.6400034	-.2874636x10 ⁴	0
.2827320	-.1720166x10 ⁻¹	.2188874x10 ¹	.9807945x10 ¹	0
.3845696x10 ⁴	-.1396548x10 ¹	.4084007x10 ¹	-.3161642x10 ²	0
.4738875x10 ⁻²	.8788591	.3937128x10 ²	.2574022x10 ²	0
.3904900	.3121670x10 ⁻³	.1111977x10 ⁻²	-.1584019x10 ⁻²	0
.1441706x10 ⁻⁶	.9705097x10 ⁻⁴	.9058803	.1439616x10 ⁻²	0
.3934547	.1358356x10 ⁻²	-.9422136x10 ⁻³	-.1635502x10 ¹	0
.2445125x10 ⁻³	.2760300x10 ⁻⁴	.1765628x10 ⁻²	.9688600	0
-.4079011	-.2001976x10 ⁻⁵	-.2944081x10 ⁻³	-.3850948x10 ¹	1
-.2466211x10 ⁻⁴	.6311254x10 ⁻⁶	.6103020x10 ⁻³	.2424015x10 ⁻²	1

TABLE 7.3b

MATRIX B_d

$-.1822657 \times 10^{-2}$	$.5398565 \times 10^{-4}$	$.5099220 \times 10^{-5}$
$-.2769619$	$.3883505$	$-.4901228 \times 10^{-3}$
$-.2266523 \times 10^{-1}$	$.7720937$	$-.3538673 \times 10^{-4}$
$.3783315 \times 10^{-4}$	$.1365414 \times 10^{-4}$	$-.2466411 \times 10^{-5}$
$.1045173 \times 10^2$	$.2845712 \times 10^1$	$-.4838950$
$-.1781018 \times 10^2$	$.2426261 \times 10^2$	$.6536699$
$-.4949797 \times 10^{-3}$	$.1107338 \times 10^1$	0.3956254×10^{-4}
$.4072448 \times 10^{-2}$	$.1307031 \times 10^{-2}$	$-.1218614 \times 10^{-1}$
$.1301989 \times 10^{-2}$	$.2874390 \times 10^{-3}$	$.1942529 \times 10^{-4}$

TABLE 7.4

THE MATRIX A_d^*

.9958574	.3335249x10 ⁻⁶	0
.2019522	.9957821	0
-.8454057x10 ⁻⁶	.2401357x10 ⁻⁷	1

THE MATRIX B_d^*

-.3816104x10 ⁻¹	.6865557	.3596243x10 ⁻³
-.7784910x10 ¹	.1890150x10 ³	-.4285849x10 ¹
-.5487150x10 ⁻⁴	-.1128416x10 ⁻²	.1296632x10 ⁻³

FIG. 7.1. Comparison of state estimates in a typical run.

FIG. 7.2. Comparison of the mean and the mean squared estimation errors in Example 7.1. The modelling approach compared with the Quigley approach.

FIG. 7.1a. Step responses of pressure and temperature - comparison of the third and the ninth order models.

FIG. 7.3b. Step response of water level with the third and the ninth order models.

FIG. 7.4a. Variation of mean squared error in the estimation of the water level, y , with the variance Q_{yy} of the noise added to the model ($Q_{TsTs}=0$, $Q_{pp}=0$).

FIG. 7.Ab. Variation of mean squared error in the estimation of the pressure, p , with the variance Q_{pp} of the noise added to the model ($Q_{yy}^{-10^{-1}}$, $Q_T T_0^{-0}$).

FIG. 7.4c. Variation of mean squared error in the estimation of the temperature, T_s , with the variance $Q_{T_s T_s}$ of the noise added to the model ($Q_{yy} = 10^{-4}, Q_{pp} = 10^4$).

FIG. 7.5a. Comparison of the mean squared errors in the estimation of the temperature, T_b , of the boiler via the three approaches: Q-constant, Modelling, Quigley.

FIG. 7.5b. Comparison of the mean squared errors in the estimation of the water level, y , of the boiler via the three approaches: Q-constant, Modelling, Quigley.

FIG. 7.5c. Comparison of the mean squared errors in the estimation of the pressure, p , of the boiler via the three approaches: Q-constant, Modelling, Quigley.

CHAPTER 8

CONCLUSIONS

This thesis presented a study of the various aspects of systems with multiple operating modes, with particular reference to two modes of operation. A unified modelling framework for switching systems was introduced and the various practical situations where such systems are encountered were discussed.

The problem of estimating the state vector of stochastic, switching systems was analyzed in considerable detail. Optimal linear and non-linear estimators were derived and it was indicated that the optimal linear filter, characterized by a deterministic error covariance matrix, failed to account accurately for estimation errors resulting from wrong detection of the mode of the system. The optimal non-linear filter provided better estimates, at the expense of an increased computational effort. The development and the use of a model describing the dynamics of the switching process resulted in estimates comparable in accuracy to those obtained from non-linear estimators, with a smaller demand on the digital computer. The CPU time and storage requirements of the modelling approach

were seen to be marginally increased over those of the optimal linear filter. The examples considered in Chapter 4 validate these conclusions.

The development of the model itself was shown to be straightforward. The model is obtained as the constant term in a Taylor series expansion of the a-posteriori probability density of y_k , the switching variable, about some chosen estimate of y_k , the output at time k . The model is fairly general and can represent all types of switching systems, as long as switches in component values occur simultaneously. Kalman filtering with this model requires the choice of the statistics of the fictitious noise added for modelling error compensation. Several approaches to the choice of the variance of this noise were presented. In the case of missing observations, the approach of Appendix II gave very good results while, in the case of switching time constants, a constant variance of 0.1 appeared to result in acceptably accurate estimates.

The problem of estimating the parameters of a stochastic switching system represented by a simple (switching) difference equation model was then considered. The maximum likelihood approach for off-line identification was shown to have the same growing memory requirement as cited earlier and, in an attempt to circumvent this difficulty, a simple

iterative scheme was outlined. On-line identification schemes for both correlated and uncorrelated error terms were then discussed. It was shown that simple modifications of existing approaches for time invariant systems (OLS, IV, etc.) result in reasonably good estimates of the parameters. In the case of uncorrelated errors, the modelling approach of Chapter 4, when adapted to the identification problem, gave estimates comparable to those obtained by a least squares approach based on perfect knowledge of the mode of the system. The modified identification schemes presented in this thesis are justified intuitively based on two reasons: first, they reduce to the existing identification schemes when the probability of a switch is zero and second, they have a close resemblance to the Bank-of-Kalman filters approach taken for state estimation (Chapter 2).

Design of the optimal stochastic controller for switching systems was considered next. It was shown that the optimal controller is difficult to arrive at in practice owing to the lack of 'separation' between estimation and control. Even in situations where this separation exists, optimal control becomes infeasible because of the large CPU time and memory requirements in the computation of the optimal cost. This observation necessitated the development of adaptive controllers in the two typical cases considered: systems with missing observations and those with switching time constants. In the former case,

the adaptive controller was developed by requiring that the expected inaccuracy (measured by the expected standard deviation) in the estimation of the mode of the system be smaller than a chosen 'tolerance limit' while, in the latter situation, the adaptive control was simply obtained as the weighted average of the mode conditional controls. Since the optimal control is infeasible, the control obtained under the assumption of perfect knowledge of the mode of the system was used as a standard for comparison (this is called the γ -optimal control). The examples considered in this dissertation indicate that the adaptive controllers perform reasonably well and that their performance approaches that of the γ -optimal controller as the signal to noise ratio increases. Moreover, separation control (i.e. control under the assumption of separation) with state estimates obtained in the modelling approach was seen to perform as well as that with estimates provided by the Bank-of-Kalman filters approach.

A switching plant noise was shown in this thesis to be useful in the sense of providing smaller estimation errors compared to the case of constant noise statistics and thus limit or prevent the divergence of the Kalman filter resulting from modelling inaccuracies. This effect was observed in the boiler example considered in this thesis

where reduction in the order of the model results in errors which may be large depending on the accuracy and the philosophy of the reduction algorithm. This simplifies the analysis and the design of complex industrial processes since a simple model complemented by a switching plant noise appears to result in reasonably good state estimates.

Some of the problems posed in this dissertation need a more detailed investigation. For example, further study is needed in the choice of Q_ψ , the variance of the noise added to the model, and ϵ , the tolerance limit on the inaccuracy in the estimation of the mode of the system. Extensions of the work reported here to systems with more than two modes of operation (say by the incorporation of a single multivalued switching variable or several binary switching variables) and to those in which parameters switch independent of each other and at different times, is straightforward. An interesting extension of this research work would be to estimate the transition probabilities q_{ij} adaptively. It is hoped that the model for the switching process developed in this dissertation would be useful in this case since the transition probabilities appear as parameters of the model.

In conclusion, this thesis has presented solutions to the problems of estimation and control for systems with two modes of operation - solutions which are simple and require lesser computation compared to existing approaches while being comparable in accuracy.

APPENDIX IINPUT DESIGN FOR ESTIMATION IN SWITCHING
SYSTEMS

The aim in designing the input is to improve the accuracy of estimation. When a signal or a parameter is to be estimated from noisy measurements, the accuracy of estimation depends on the signal to noise ratio (S/N ratio) which is defined as the ratio of the r.m.s. values of the signal and the noise component of the measurements. Obviously, an increase in the signal magnitude (or a decrease in the noise level) and hence a larger S/N ratio results in more accurate estimates.

In switching systems, there is an additional variable to be estimated: the switching variable γ_k . Considering that the Kalman filter provides the optimal (in m.m.s.e. sense) estimates of the states of a linear system, and remembering that the switching systems considered in this thesis are represented by a linear model in any given mode, the accuracy of state estimation is seen to depend on that in the detection of the mode of the system. The design of the input signal, U_k , is thus based on obtaining the best possible estimates of the mode of the system (value of γ_k).

Two approaches for the design of the input signal are considered here:

Case 1. U_k is constrained, in magnitude, say $|U_k| \leq V$.

Case 2. U_k is designed so that the expected inaccuracy in the estimation of γ_{k+1} is less than a pre-specified tolerance limit (see Chapter 6, Section 6.4.2).

These two cases are discussed below.

Case 1. Magnitude Constraint on U_k

The minimum mean squared error estimate of γ_{k+1} , $\gamma_{k+1}^*|_{k+1}$ is given as (Chapter 2):

$$\begin{aligned} \gamma_{k+1}^*|_{k+1} &\triangleq E(\gamma_{k+1} | Y^{k+1}) \\ &= \left[1 + \frac{P(\gamma_{k+1}=0 | Y^k)}{P(\gamma_{k+1}=1 | Y^k)} \pi(y_{k+1}) \right]^{-1} \end{aligned} \quad (1.1)$$

where

$$\pi(y_{k+1}) = P(y_{k+1} | Y^k, \gamma_{k+1}=0) / P(y_{k+1} | Y^k, \gamma_{k+1}=1)$$

Given the transition probabilities, q_{ij} , of the Markov chain for γ , the a-priori densities $P(\gamma_{k+1}=i | Y^k)$, $i=0,1$ are obtained as:

$$P(\gamma_{k+1}=1 | Y^k) \triangleq \gamma_{k+1}^*|_k = (q_{11} - q_{10}) \gamma_k^* + q_{10} \quad (1.2)$$

where $\gamma_{k|k}^*$ is the m.m.s.e. estimate of γ_k based on Y^k ,
and $p(\gamma_{k+1}=0|Y^k) = 1-p(\gamma_{k+1}=1|Y^k)$.

It is noticed from (I.1) that when the y_{k+1} is observed,
only $\lambda(y_{k+1})$ is affected. Moreover, assuming $q_{10} \neq 0$,

$$\lambda(y_{k+1}) \ll 1 \rightarrow \gamma_{k+1|k+1}^* \rightarrow 1$$

$$\lambda(y_{k+1}) = 1 \rightarrow \gamma_{k+1|k+1}^* = \gamma_{k+1|k}^*, \text{ i.e. } y_{k+1} \text{ is effectively ignored}$$

$$\text{and } \lambda(y_{k+1}) \gg 1 \rightarrow \gamma_{k+1|k+1}^* \rightarrow 0$$

It is thus seen that the requirement that $\gamma_{k+1|k+1}^*$
be as accurate as possible implies that $\gamma(y_{k+1})$ be as far
away from unity as possible within the constraint imposed
on U_k . It is further noticed that irrespective of the
conditional distribution of y_{k+1} , only the first moment
of the densities $p(y_{k+1}|Y^k, \gamma_{k+1}=i)$, $i=0,1$, are functions
of U_k . It is thus obvious that when the difference between
the means of the two conditional densities of y_{k+1} reaches
a maximum value, $\lambda(y_{k+1})$ is farthest from unity. Hence,
 U_k is chosen by performing the maximization:

$$\max_{|U_k| \leq V} ||E(y_{k+1}|Y^k, \gamma_{k+1}=1) - E(y_{k+1}|Y^k, \gamma_{k+1}=0)||^2 \quad (I.3)$$

Using the results of Chapter 2, this becomes:

$$\max_{|U_k| \leq V} \left| (C_1 A_1 - C_0 A_0) E(x_k | Y^k) + (C_1 B_1 - C_0 B_0) U_k \right|^2$$

which yields, for a system with p inputs,

$$U_k = MV$$

where M is a $p \times p$ diagonal matrix with elements m_{ij} ,

$$\begin{aligned} m_{ij} &= 0 \quad i \neq j \\ &= \text{sign} \left[\text{ith component of the } p \text{ vector} \right. \\ &\quad \left. (C_1 B_1 - C_0 B_0)^T (C_1 A_1 - C_0 A_0) E(x_k | Y^k) \right] \end{aligned} \quad (I.4)$$

For the case of a single input, simple output system

(I.4) becomes:

$$U_k = V \text{ sign} \left\{ \frac{(C_1 A_1 - C_0 A_0) E(x_k | Y^k)}{(C_1 B_1 - C_0 B_0)} \right\}$$

Case 2. In this case, the input U_k is chosen so that the expected standard deviation of $Y_{k+1}^* | k+1$, $\sigma(Y_{k+1}^* | k+1)$, is less than a prespecified tolerance limit, ϵ , i.e.

$$\sigma(Y_{k+1}^* | k+1) \leq \epsilon$$

In Section 6.4.2 of Chapter 6, this standard deviation was shown to be given by:

$$\sigma(Y_{k+1}^* | k+1) = \sqrt{p_1(1-p_1)} \rho$$

where

$$P_1 = p(y_{k+1}=1|Y^k) = \gamma_{k+1|k}^*$$

and ρ is the Bhattacharyya distance defined as:

$$\rho = \int \sqrt{p(y_{k+1}|Y^k, \gamma_{k+1}=1)p(y_{k+1}|Y^k, \gamma_{k+1}=0)} dy_k$$

Now, assuming as in Ackerson and Fu, that the conditional density of the state is Gaussian, i.e.

$$p(x_k|Y^k) = N(\hat{x}_{k|k}, P_{k|k}),$$

the conditional densities of y_{k+1} are given as:

$$p(y_{k+1}|Y^k, \gamma_{k+1}=i) = N(\mu_i, \sigma_i^2), \quad i=0,1$$

where

$$\mu_i = C_i A_i \hat{x}_{k|k} + C_i B_i U_k$$

and

$$\sigma_i^2 = C_i \left[A_i P_{k|k} A_i^T + Q_i \right] C_i^T + R_i$$

In this case, the distance ρ is

$$\rho = \alpha e^{-\beta}$$

where

$$\alpha = \left\{ \frac{|\sigma_1^2 + \sigma_0^2|}{2|\sigma_1^2||\sigma_0^2|} \right\}^{m/2}$$

m is the dimension of the output vector y_{k+1}

and

$$\beta = \frac{1}{8} ||\mu_1 - \mu_0||^2 \left[\frac{1}{2}(\sigma_1^2 + \sigma_0^2) \right]^{-1}$$

Hence, the requirement

$$\sigma(\gamma_{k+1|k+1}^*) \leq \epsilon$$

implies

$$\alpha \sqrt{\gamma_{k+1|k}^* (1 - \gamma_{k+1|k}^*)} e^{-\beta} \leq \epsilon$$

or

$$\beta \geq \ln \left\{ \frac{\alpha \sqrt{\gamma_{k+1|k}^* (1 - \gamma_{k+1|k}^*)}}{\epsilon} \right\} = \epsilon_2 \quad (I.5)$$

For the multiple input case, no general solutions to (I.5) appear possible as the equation $\beta = \epsilon_2$ represents an ellipsoid, but for the case of a single input, the solution is straightforward. Define:

$$\epsilon_3 = ||(C_1 A_1 - C_0 A_0) \hat{x}_{k|k}||^2 \left[\sigma_1^2 + \sigma_2^2 \right]^{-1}$$

$$\epsilon_4 = (C_1 B_1 - C_0 B_0)^T \left[\sigma_1^2 + \sigma_2^2 \right]^{-1} (C_1 A_1 - C_0 A_0) \hat{x}_{k|k}$$

$$\epsilon_5 = (C_1 B_1 - C_0 B_0)^T \left[\sigma_1^2 + \sigma_2^2 \right]^{-1} (C_1 B_1 - C_0 B_0)$$

Then, (I.5) implies:

$$\epsilon_5 U_k^2 + \epsilon_4 U_k + \epsilon_3 \geq 4\epsilon_2$$

which yields, after simple algebraic manipulations,

$$U_k = -\frac{\epsilon_4}{2\epsilon_5} + \text{sign}(\epsilon_4) \frac{\sqrt{\epsilon_4^2 - 4\epsilon_5(\epsilon_3 - 4\epsilon_2)}}{2\epsilon_5} \quad \text{if } \epsilon_3 \geq 4\epsilon_2$$

= 0 otherwise. (I.6)

APPENDIX II

CHOICE OF Q_ψ - COVARIANCE OF NOISE ADDED TO MODEL

It was indicated in Chapter 3 that the choice of Q_ψ is crucial. This is because Q_ψ is chosen so as to be sufficient to 'bury' any modelling errors and while a small value does not reflect modelling errors, a large value overcompensates for these errors. In an attempt to obtain more insight into the choice of Q_ψ , a Monte-Carlo study was conducted with a simple scalar system:

$$\begin{aligned} x_{k+1} &= a x_k + U_k + w_k \\ \gamma_k &= \gamma_k x_k + v_k \end{aligned} \tag{II.1}$$

where $w_k \sim N(0, Q)$ and $v_k \sim N(0, R)$

and U_k is a control input chosen such that x_k is regulated around the point $\bar{x}$. γ_k is a binary random variable with a switching probability q_{10} . It is required to append a model for γ of the form:

$$\gamma_{k+1} = f(\gamma_k, x_k) + \psi_k$$

where $\psi_k \sim N(0, Q_\psi)$.

In the steady state, there are five independent factors influencing the choice of Q_ψ :

q_{10} , a , Q , R and $\bar{x}$.

An empirical study for the choice of Q_ψ thus requires the variation of these five factors. It is possible to reduce the number of independent parameters by a dimensional analysis because dimensionless form of equations is simpler. With x as the basic dimension and from the system equations, we have:

$$\bar{x} = d = x \text{ (i.e. } \bar{x} \text{ has dimensions of } x \text{)}$$

$$a = d = 1 \text{ (a is dimensionless)}$$

$$q_{10} = d = 1$$

$$Q = d = x^2$$

$$R = d = x^2$$

Thus, dimensionless factors are a , q_{10} , $Q/\bar{x}^2$, $R/\bar{x}^2$ and Q_ψ itself is dimensionless.

A different analysis can be obtained on noticing that a change ΔT in the time step produces changes in a , q_{10} and Q which are not independent, provided there is an underlying continuous system. In this underlying continuous system,

$$T = \text{time constant} = -t/ha$$

$$q_{10} = \text{probability of a switch, per unit time,} = q_{10}/\Delta t$$

Q_c = variance of a white noise process = $Q\Delta t$

R = variance of measurement noise

and $\bar{x}$ = mean value of x .

There are now two basic dimensions x and t . Thus,

$$\bar{x} = d = x$$

$$T = d = t$$

$$\pi_{10} = d = t^{-1}$$

$$Q_c = d = x^2 t \text{ because } Q_c = Q\Delta t = d = x^2 t$$

$$R = d = x^2$$

In this case, suitable dimensionless factors are:

$$\frac{\bar{x}^2}{R}, \frac{\bar{x}^2}{Q} T, \text{ and } \pi_{10} T$$

or in terms of the discrete time equivalents

$$\frac{\bar{x}^2}{R}, -\frac{\bar{x}^2}{Q \ln(a)}, \text{ and } -\frac{q_{10}}{na}$$

one thus observes Q_ψ to have the functional form:

$$Q_\psi = f\left(\frac{\bar{x}^2}{R}, -\frac{\bar{x}^2}{Q \ln(a)}, -\frac{q_{10}}{\ln a}\right) \quad (II.2)$$

In an attempt to determine this functional form, a Monte-Carlo study was conducted with several sets of values for the above three dimensionless factors and in each case

the best Q_ψ was obtained as the one giving the minimum mean squared error in the estimates of x . The mean squared error was obtained by averaging over 100 runs, each over 30 time steps. Variation of Q_ψ with $x^2/Q \ln a$ and with $\bar{x}^2/R$ as a parameter and $q_{10}=0.9$, $a=0.95$ is shown in Fig. II.1. The following observations are made from this figure:

- i) there appears to be a 'bias' in Q_ψ independent of $Q/x^2 \ln a$ and R/x^2
- ii) the Q_ψ required appears to decrease exponentially with a decrease in the signal level (x)
- iii) Q_ψ appears to have a peak approximately at $Q \ln(a)/x^2 = 0.25$
- iv) for R/x^2 or $Q \ln(a)/x^2 > 0.5$, Q_ψ appears to be a constant.

Similar observations were made with several other values for $q_{10}/\ln a$. In addition, the 'bias' in Q_ψ appeared to decrease with a decrease in $-q_{10}/\ln a$. Curve fitting with the data obtained indicated the following form for Q_ψ :

$$Q_\psi = f\left(\frac{Q}{x^2 \ln a}\right) e^{-15R/\bar{x}^2} + \frac{1}{2} e^{-\ln a/2q_{10}}$$

The function f was observed to have a parabolic stage, approximately represented by the positive half of a sine wave with $Q \ln(a)/x^2 = 0.25$, corresponding to the $\pi/2$ point.

or the peak. Thus,

$$Q_{\psi} = .25\exp(-\ln(a)/2q_{10}) + 0.5\sin(-2\pi\ln(a)NSR_p)\exp(-15NSR_o), \quad (II.3)$$

$$= .25\exp(-\ln(a)/2q_{10}), \quad \text{if } \exp(-1/2NSR_p) < a < 1$$

$$\text{otherwise}$$

It is noticed that R/x^2 and Q/x^2 are measures of the noise to signal ratios at the output and the plant respectively. For small values of these noise to signal ratios, Q_{ψ} is determined by the second term in (II.3) and thus varies from 0 to 0.25 depending on q_{10} and a . It was observed in Chapter 4 that, for large S/N ratio, $Q_{\psi}=0.1$ or 0.25 gives acceptable results, thus justifying the result (II.3).

The choice of Q_{ψ} from (II.3) was tested on two randomly generated scalar systems and the results are given below:

System 1

Given: $a=0.98$, $\bar{x}^2=2$, $q_{10}=0.153$, $R=0.425$, $Q=0.097$

Q_{ψ} (predicted from II.3) = 0.236

Q_{ψ} (actual) = 0.236 ± 0.025

System 2

Given: $a=0.955$, $\bar{x}^2=2$, $q_{10}=0.0315$, $Q=0.086$, $R=0.07$.

Q_{ψ} (predicted) = 0.25

Q_{ψ} (actual) = 0.225 ± 0.025

In these simulations, Q_ψ (actual) was the value resulting in the minimum value of the mean squared error in the state estimates. These two simulations suggest that II.3 gives a reasonably good value for Q_ψ .

For higher order systems (say n states and m outputs), extensions of (II.3) can be envisaged by choosing the worst case situations for Q/x^2 , R/x^2 and a . For example, one could choose $Q \ln(a)/x^2$ as the value closest to 0.25 (corresponding to the peak in Q_ψ) out of the n possible N/S ratios at the plant and the largest R/x^2 out of the noise to signal ratios at the output. Similarly, the largest eigenvalue (corresponding to the largest bandwidth) could be chosen for a . This extension was taken for the second order example (Example 4.2) in Chapter 4.

FIG. 11.1. Variation of the best Q_p^* with $\frac{Q \ln(a)}{x^2}$; $\frac{R}{x^2}$ is a parameter.

APPENDIX III

CONSTRAINED LINEAR ESTIMATION

The system considered here is identical to (4.1) and is represented by the equations:

$$\begin{aligned} z_{k+1} &= g(z_k, u_k) + w_k \\ y_k &= h(z_k) + \eta_k \end{aligned} \quad (\text{III.1})$$

where z_{k+1} is $(n+1)$ vector of states, defined as:

$$z_k \triangleq \begin{bmatrix} x_k \\ y_k \end{bmatrix}$$

x_k being the state vector of the original system and y_k is a binary $(0,1)$ random variable. The functions g and h are known and w_k and v_k are independent white Gaussian sequences with zero means and with covariances Q_k and R_k respectively.

It is required to obtain a constrained linear estimate $\hat{z}_{k|k}^C$ of the state vector z_k , based on measurements up to time k , such that the criterion

$$J = E(z_k - \hat{z}_{k|k}^C)^T (z_k - \hat{z}_{k|k}^C) \quad (\text{III.2})$$

is minimized, subject to the constraint:

$$0 \leq y_{k|k}^C \leq 1 \quad (\text{III.3})$$

Notice that γ_k is the $(n+1)$ th component of z_k and can be obtained as

$$\gamma_k \triangleq e_{n+1}^T z_k \quad (\text{III.4})$$

where

$$e_{n+1} \triangleq [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \dots \ 0 \ 1]^T \quad (\text{III.5})$$

Let $\hat{z}_{k|k}$ denote the unconstrained linear estimate of z_k , i.e. $\hat{z}_{k|k}$ minimizes the criterion (III.2) without considering the constraint (III.3).

Since both $\hat{z}_{k|k}^C$ and $\hat{z}_{k|k}$ are linear estimates of z_k , they must be related by a linear transformation:

$$\hat{z}_{k|k}^C = L_k \hat{z}_{k|k} \quad (\text{III.6})$$

where L_k is an $(n+1) \times (n+1)$ matrix.

Introduction of the Lagrange multipliers λ_k and μ_k transforms the criterion (III.2) to:

$$J^* = E(\hat{z}_{k|k}^C - z_k)^T (\hat{z}_{k|k}^C - z_k) + \lambda_k (e_{n+1}^T \hat{z}_{k|k}^C - 1) - \mu_k e_{n+1}^T \hat{z}_{k|k} \quad (\text{III.7})$$

• where

$$\lambda_k = \begin{cases} 0 & \text{if } \hat{\gamma}_{k|k}^C < 1 \\ > 0 & \text{otherwise} \end{cases}$$

and

$$\mu_k = \begin{cases} 0 & \text{if } \hat{\gamma}_{k|k}^C > 0 \\ > 0 & \text{otherwise.} \end{cases}$$

Viewing $J^*(L_k)$ as a functional of L_k and $\delta J^*(L_k)$ as the first variation of J^* for a small arbitrary change δL_k in L_k , the necessary condition for minimum is given as:

$$\delta J^*(L_k) = 0$$

or

$$\text{Trace}\{\delta L_k E(\hat{z}_{k|k} \hat{z}_{k|k}^T) L_k^T - \delta L_k E(\hat{z}_{k|k} z_k^T) + \frac{\lambda_k - \mu_k}{2} \delta L_k \hat{z}_{k|k} e_{n+1}^T\} = 0 \quad (\text{III.11})$$

Since $\hat{z}_{k|k}$ is a linear estimate of z_k , satisfied the orthogonality conditions (see Chapter 2, Section 2.2)

$$E(z_k - \hat{z}_{k|k}) \hat{z}_{k|k}^T = 0$$

or

$$E(z_k \hat{z}_{k|k}^T) = E(\hat{z}_{k|k} \hat{z}_{k|k}^T) \quad (\text{III.12})$$

Using (III.6) and (III.12) in (III.11),

$$\text{Trace}\{(\delta L_k) \left[E(\hat{z}_{k|k} \hat{z}_{k|k}^C) - E(\hat{z}_{k|k} \hat{z}_{k|k}^T) + \frac{\lambda_k - \mu_k}{2} \hat{z}_{k|k} e_{n+1}^T \right]\} = 0$$

Which implies, for an arbitrary δL_k ,

$$E(\hat{z}_{k|k} \hat{z}_{k|k}^C) - E(\hat{z}_{k|k} \hat{z}_{k|k}^T) + \frac{\lambda_k - \mu_k}{2} \hat{z}_{k|k} e_{n+1}^T = 0 \quad (\text{III.13})$$

On using (III.6) in (III.13), L_k and hence $\hat{z}_{k|k}^C$ are obtained as:

$$\hat{z}_{k|k}^C = \hat{z}_{k|k} + \frac{\mu_k - \lambda_k}{2} e_{n+1}^T \hat{z}_{k|k} \{E(\hat{z}_{k|k} \hat{z}_{k|k}^T)\}^{-1} \hat{z}_{k|k} \quad (\text{III.14})$$

Q.E.D.

It remains to determine λ_k and μ_k . They are both zero when the constraint (III.3) is satisfied. When the estimate of γ_k goes outside the range (0,1), the constrained minimum of J^* is seen to be achieved at the boundary, i.e.

$\hat{\gamma}_{k|k}^C = 0$ or 1. Thus, λ_k and μ_k are given by:

$$(i) \text{ if } 0 < e_{n+1}^T \hat{z}_{k|k} < 1, \lambda_k = 0, \mu_k = 0$$

$$(ii) \text{ if } e_{n+1}^T \hat{z}_{k|k} > 1, \mu_k = 0$$

$$\lambda_k = 2(e_{n+1}^T \hat{z}_{k|k} - 1) \left[\hat{z}_{k|k}^T \{E(\hat{z}_{k|k} \hat{z}_{k|k}^T)\}^{-1} \hat{z}_{k|k} \right]^{-1}$$

$$(iii) \text{ if } 0 > e_{n+1}^T \hat{z}_{k|k}, \lambda_k = 0$$

$$\mu_k = -2e_{n+1}^T \hat{z}_{k|k} \left[\hat{z}_{k|k}^T \{E(\hat{z}_{k|k} \hat{z}_{k|k}^T)\}^{-1} \hat{z}_{k|k} \right]^{-1}$$

Substitution of these values of λ_k and μ_k in (III.14) and some algebraic manipulations leads to the constrained estimate:

$$\hat{x}_{k|k}^C = \hat{x}_{k|k}$$

$$\hat{\gamma}_{k|k}^C = \hat{\gamma}_{k|k}$$

$$= 1$$

$$= 0$$

$$\text{if } 0 \leq \hat{\gamma}_{k|k} \leq 1$$

$$\text{if } \hat{\gamma}_{k|k} > 1$$

$$\text{if } 0 > \hat{\gamma}_{k|k}$$

(III.15)

This simple result can be explained geometrically: the function (III.7) is quadratic in $\hat{z}_{k|k}^C$ and is therefore a convex function with a single minimum. When the estimate of γ_k , $\hat{\gamma}_{k|k}^C$, lies outside the specified range, the solution vector moves along the γ axis until the constraint is satisfied, i.e. at the boundary. The remaining components of the solution vector are unaffected.

REFERENCES

- A1. Ackerson, G.A. and Fu, K.S., "On-State Estimation in Switching Environments", IEEE Tr. on Automatic Control, Vol. AC-15, No. 1, pp. 10-17, February 1970.
- A2. Ackerson, G.A., "Control and Estimation in Markov Switching Environments", Ph.D. Dissertation, School of Elec. Engrg., Purdue University, Lafayette, Indiana, June 1968.
- A3. Anderson, J.H., "Geometrical Approach to Reduction of Dynamical Systems", Proc. IEE, Vol. 114, No. 7, pp. 1014-1018, July 1967.
- A4. Anscombe, F.J., "Rejection of Outliers", Technometrics, Vol. 2, No. 2, pp. 123-147, May 1960.
- A5. Aoki, M., Optimization of Stochastic Systems, N.Y., Academic Press, pp. 52, 1967.
- A6. Aoki, M., "Optimal Control Policies for Dynamical Systems Whose Characteristics Change Randomly at Random Times", 3rd IFAC Congress, London, pp. 29A.1-29A.8, June 1966.
- A7. Astrom, K.J. and Eykhoff, P., "System Identification - A Survey, Automatica, Vol. 7, No. 2, pp. 123-162, March 1971.
- A8. Astrom, K.J., "Lectures on the Identification Problem - The Least Squares Method", Report 6806, Lund Institute of Technology, Division of Automatic Control, Lund, Sweden, September 1968.
- B1. Bruckner, J.M.H. and Scott, R.W., "State Estimation for Systems with Multiple Operational Modes", Proc. of the 1972 IEEE Conference on Decision and Control, Paper No. WPG-1, pp. 201-205, 1972.
- B2. Bryson, A.E. and Ho, Y.C., Applied Optimal Control, Blaisdell Publishing Co., 1969.

- C1. Chieu, H.L., Ergin, E.I., Ling, C. and Allyn Lee, "Dynamic Analysis of a Boiler", 4th ASME Instruments and Regulators Division Conference, University of Delaware, Newark, Del., 1958.
- C2. Collmeyer, A.J. and Gupta, S.C., "Linear Filtering with Constraints", JACC Preprints, pp. 137-143, 1968.
- D1. Davison, E.J., "A Method of Simplifying Linear Dynamic Systems", IEEE Tr. on Automatic Control, Vol. AC-11, No. 1, pp. 93-101, January 1966.
- D2. DeRusso, P.M. Roy, R.J. and Close, C.M., State Variables for Engineers, John Wiley, N.Y., pp. 442, 1967.
- D3. Deshpande, J.G., Upadhyay, T.N. and Lainiotis, D.G., "Adaptive Control of Linear Stochastic Systems", Automatica, Vol. 9, pp. 107-115, 1973.
- D4. Doob, J.L., Stochastic Processes, John Wiley, N.Y., 1953.
- E1. Eykhoff, P., "Process Parameter and State Estimation", Automatica, Vol. 4, pp. 205-233, 1968.
- F1. Feldbaum, A.A., Optimal Control Systems, Academic Press, N.Y., 1965.
- F2. Fitzgerald, R.J., "Divergence of the Kalman Filter", IEEE Tr. on Automatic Control, Vol. AC-16, No. 6, pp. 736-747, February 1971.
- F3. Fujishige, S., "Remarks on Optimal Stochastic Control for Discrete-Time Linear Systems with Interrupted Observations", Automatica, Vol. 10, No. 2, pp. 213-214, March 1974.
- F4. Fujita, S. and Fukao, T., "Optimal Stochastic Control for Discrete-Time Linear Systems with Interrupted Observations", Automatica, Vol. 8, No. 4, July 1972.
- G1. Gustafson, D.E. and Speyer, J.L., "Linear Minimum Variance Estimation in Systems with State Dependent Noise", Proc. of the 2nd Symposium on Non-Linear Estimation Theory, San Diego, California, 1971.
- H1. Haddad, A.H. and Cruz, J.B., "Non-Linear Filtering for Systems with Random Models", Proc. of the 2nd Symposium on Non-Linear Estimation Theory, San Diego, California, 1971.

- J1. Jain, A. and Swarder, D.D., "Constrained Linear Filtering as a Multistage Process", Proc. JACC, pp. 135-139, 1971.
- J2. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, N.Y., 1970.
- K1. Kailath, T., "The Divergence and Bhattacharyya Distance Measures in Signal Selection", IEEE Tr. Comm. Technology, Vol. COM-15, pp. 52-60, February 1967.
- K2. Kalman, R.E. and Bucy, R.S., "New Results in Linear Filtering and Prediction Theory", Trans. ASME, J. Basic Eng., Vol. 83, pp. 95-108, 1961.
- K3. Kashyap, R.L., "Maximum Likelihood Identification of Stochastic Linear Systems", IEEE Tr. on Automatic Control, Vol. AC-15, No. 1, February 1970.
- K4. Kazangey, T. and Swarder, D.D., "Effective Federal Policies for Regulating Residential Housing", Proc. of the 1971 Summer Computer Simulation Conference, pp. 1120-1128, July 1971.
- K5. Kohlas, J., "On Bad Data Suppression", IEEE Tr. on Automatic Control, Vol. AC-17, No. 6, pp. 827-828, December 1972.
- K6. Kushner, H., Introduction to Stochastic Control, Holt, Rinehart, and Winston, 1971.
- L1. Lainiotis, D.G. and Park, S.K., "On Joint Detection, Estimation and System Identification: Discrete Data Case", International J. of Control, Vol. 17, No. 3, pp. 609-633, 1973.
- L2. Lainiotis, D.G., "Relationship Between Optimal Detection and Bhattacharyya Coefficient", IEEE Tr. on Int. Theory, Vol. IT-15, pp. 504-505, 1969.
- M1. McLane, P.J., "Optimal Linear Filtering for Linear Systems with State Dependent Noise", Int. J. of Control, Vol. 10, No. 1, pp. 41-51, 1969.
- M2. Mendes, M., "An On-Line Adaptive Control Method", Automatica, Vol. 7, No. 3, pp. 323-332, May 1971.

- M3. Merril, H.M. and Schweppe, F.C., "Bad Data Suppression in Power System Static State Estimation", IEEE Trans. on Power, Apparatus and Systems, Vol. PAS90, pp. 2718-2725, 1971.
- M4. Mishulina, O.A., "Accuracy of Linear Control Systems with Random Structure Change", Engineering Cybernetics, No. 1, 1970.
- M5. Moose, R.L. and Wong, P.P., "An Adaptive Estimator with Learning for a Plant Containing Semi-Markov Switching Parameters", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-3, No. 3, pp. 277-280, May 1973.
- N1. Nahi, N., "Optimal Recursive Estimation with Uncertain Observations", IEEE Trans. on Information Theory, Vol. IT-15, No. 4, pp. 457-462, July 1969.
- N2. Nicholson, H., "Dynamic Optimization of a Boiler", Proc. IEE, Vol. 111, No. 8, pp. 1479-1499, 1964.
- P1. Pandya, R., "A Class of Boot-Strap Estimators for Linear System Identification", Int. J. of Control, Vol. 15, No. 6, June 1972.
- Q1. Quigley, A.L.C., "An Approach to the Control of Divergence in Kalman Filter Algorithms", Int. J. of Control, Vol. 17, No. 4, pp. 741-746, 1973.
- R1. Rajasekharan, P.K., Satyanarayana, N. and Srinath, M.D., "Optimal Linear Estimation of Stochastic Signals in the Presence of Multiplicative Noise", IEEE Trans. on Aero. and Elect. Systems, Vol. AES-7, pp. 462-468, May 1971.
- R2. Rosenbrock, H.H., "An Example of Optimal Adaptive Control", J. Electronics and Control, Vol. 13, pp. 557-567, 1964.
- R3. Rowe, I.H., "A Boot-Strap Method for the Statistical Estimation of Model Parameters", Int. J. of Control, Vol. 12, No. 4, April 1970.
- R4. Rowe, I.H., "A Statistical Model for the Identification of Multi Variable Stochastic Systems", IFAC Symposium on Multi Variable Control Processes, Dusseldorf, October 1968.

- S1. Sage, A.P. and Melsa, J.L., "Estimation Theory with Applications to Communications and Control", McGraw Hill, 1971.
- S2. Sankaran, V. and Srinath, M.P., "Duality of Linear Estimation and Control with Constraints", J. of Opt. Theory and Applications, Vol. 10, No. 1, pp. 41-50, 1972.
- S3. Sawaragi, Y., Katayama, T. and Fujishige, S., "Sequential State Estimation with Interrupted Observations", Information and Control, Vol. 21, pp. 56-71, 1972.
- S4. Schlee, F.H., Standish, C.J. and Toda, N.F., "Divergence in the Kalman Filter", AIAA Journal, Vol. 5, No. 6, pp. 1114-1120, 1967.
- S5. Szwed, D.D. and Robinson, V.G., "Feedback Regulators for Jump Parameter Systems", Proc. of the 1972 IEEE Conference on Decision and Control, pp. 441-445, 1972.
- T1. Tse, E., Bar-Shalom, Y. and Miler, L., "Wide Sense Adaptive Dual Control for Non-Linear Stochastic Systems", IEEE Trans. on Automatic Control, Vol. AC-18, No. 2, pp. 98-108, April 1973.
- W1. Wernersson, A., "Comments on: Optimal Stochastic Control for Discrete-Time Linear Systems with Interrupted Observations", Automatica, Vol. 10, No. 1, pp. 113-114, January 1974.
- Y1. Yashiu, A.I., "Relative Semi-Invariants in the Filtration of Processes with Step Components", Aut. and Remote Control, No. 2, pp. 188-194, February 1970.
- Y2. "Filtration of Jump Processes", Auto. and Remote Control, No. 5, pp. 725-730, May 1970.

END

190377

FIN