

TOPICS IN STOCHASTIC CONTROL WITH APPLICATIONS
TO A TUBULAR REACTOR

By

©

THOMAS JAMES HARRIS, M.ENG.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

Hamilton, Ontario, Canada

January, 1980

copyright © Thomas James Harris, 1980

TOPICS IN STOCHASTIC CONTROL

DOCTOR OF PHILOSOPHY (1980)
(Chemical Engineering)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Topics in Stochastic Control with Applications
to a Tubular Reactor

AUTHOR: Thomas James Harris, B.Sc. (Queen's University)
M.Sc. (McMaster University)

SUPERVISORS: Dr. John F. MacGregor
Dr. Joseph D. Wright

NUMBER OF PAGES: xvii, 310

ABSTRACT

THOMAS JAMES HARRIS: Topics in Stochastic Control with Applications to a Tubular Reactor, Ph.D. Thesis, McMaster University at Hamilton, January (1980).

This thesis represents part of an ongoing study on the modelling and control of a pilot scale packed bed tubular reactor carrying out the hydrogenolysis of n-butane. The use of time series modelling and stochastic control to analyze and develop control strategies for this process was investigated.

The mass and energy balances describing the hydrogenolysis of n-butane in this tubular reactor are a set of nonlinear partial differential equations in time and two spatial co-ordinates. In spite of their complexity, an analytical solution to these equations exists for certain linear combinations of the reaction species. These linear combinations, known as reaction invariants, define the reaction stoichiometry.

A dynamic model of the process suitable for on-line computer control had previously been developed from the material and energy balances. This dynamic model, and a stochastic model for the inherent process disturbances, were used to investigate the optimal location of thermocouples along the central axis of the reactor. The results of this analysis indicated that good state estimation and control of the temperature profile, and effluent concentrations, could be achieved when one or two

thermocouples were located in the vicinity of the hot spot (maximum) temperature. The location of the thermocouples was insensitive to the statistical properties of the disturbances. A canonical variate analysis of the reactor temperatures using this model indicated that the variation of the hot spot temperature and average temperature had significant predictable components. Other linear combinations of the axial temperatures were essentially white noise processes, and therefore unpredictable. Control of the hot spot temperature and average temperature would control most of the predictable variation in the temperature profile, and as a result, most of the predictable variation in the effluent concentrations.

Univariate stochastic controllers designed for processes with deadtime have some very unusual spectral characteristics. The spectral characteristics of these controllers depends on the process deadtime and the structure and parameters of the stochastic disturbance model.

A number of univariate self-tuning regulators were implemented to control the hot spot temperature. The self-tuning controller gave good control over the hot spot temperature when compared to digital proportional plus integral type controllers. This algorithm quickly tuned the controller parameters and was robust to the assumption in its derivation.

Linear quadratic controllers designed for stochastic disturbances are identical to those designed to compensate for

an 'equivalent' class of deterministic disturbances. When the controller structures are identical, it is the manner in which the state variables are reconstructed that determines how well a control strategy will perform. Although controllers can be readily designed to compensate processes subject to deterministic and stochastic disturbances, reconstruction of the state variables in such instances is complicated by the presence of both types of disturbances.

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge my gratitude to my supervisors, Dr. John F. MacGregor and Dr. Joseph D. Wright of the Department of Chemical Engineering. Their unending patience, and guidance were truly appreciated, and made my association with them very rewarding.

I would like to express my appreciation to others who helped:

To Dr. C.M. Crowe for the considerable interest he has shown in this work;

To McMaster University and the Province of Ontario for financial assistance;

To my wife Barb for her words of encouragement and masterful job of typing an illegible manuscript.

CONTENTS

	Page
1. INTRODUCTION	
2. TOPICS IN REACTOR MODELLING	
2.1 Introduction	7
2.2 Reaction Scheme	8
2.3 Reaction Stoichiometry and Independence of Reaction	9
2.4 Description of the Packed Bed Tubular Reactor	15
2.5 Stoichiometry in Tubular Reactors	17
2.6 Temperature Sensitivity and Reactor Runaway	22
2.7 Summary	37
3. STATE SPACE MODELLING AND CONTROL OF TUBULAR REACTORS	
3.1 Introduction	38
3.2 Review of Linear Quadratic Control Theory	38
3.3 State Space Formulation of Tubular Reactors	46
3.4 Control of Tubular Reactors	47
3.5 Development of the State Space Model in Ratio and Total Flow Variables	55
3.6 Wall Temperature Effects	56

	Page
3.7 Estimation of R_w	57
3.8 Summary	60
4. OPTIMAL SENSOR LOCATION IN TUBULAR REACTORS	
4.1 Introduction	62
4.2 Theory on the Optimal Location of Sensors	64
4.3 Distributed Parameter Systems and Tubular Reactors	71
4.4 Description of the Tubular Reactor System and Model	74
4.5 The Optimal Location of Reactor Thermocouples	77
4.6 Summary	85
5. FORECASTING AND DIMENSIONALITY REDUCTION OF MULTIPLE TIME SERIES	
5.1 Introduction	86
5.2 Time Series Models	87
5.3 Dimensionality Reduction of the Output Space	99
5.4 Canonical Variate Analysis of Time Series	113
5.5 Input/Output Pairing	118
5.6 Summary	123
6. FREQUENCY RESPONSE OF STOCHASTIC CONTROLLERS	
6.1 Introduction	125
6.2 Sampling of a Continuous Signal	126

	Page
6.3 Frequency Response of Digital Filters	127
6.4 Process Dynamic and Disturbance Models	134
6.5 Spectral Characteristics of Stochastic Controllers	136
6.6 Stochastic Controllers and Discrete Otto Smith Regulators	148
6.7 Summary	152
7. SELF-TUNING AND ADAPTIVE CONTROLLERS	
7.1 Introduction	153
7.2 Self-Tuning Control	155
7.3 Application to the Control of a Catalytic Chemical Reactor	165
7.4 Discussion	187
7.5 Summary	200
8. MODELLING AND CONTROL OF DETERMINISTIC AND STOCHASTIC PROCESSES	
8.1 Introduction	203
8.2 Modelling of Stochastic and Randomly Occurring Deterministic Disturbances	207
8.3 Control of Processes Subject to Deterministic Disturbances	215
8.4 Reconstruction of the States and Disturbances by Observers	236
8.5 The Certainty Equivalence Property in Stochastic Control	238

	Page
8.6 Control of Processes Subject to Both Deterministic and Stochastic Disturbances	240
8.7 Summary	259
9. CONCLUSIONS	261
REFERENCES	266
APPENDIX 1. RATE EXPRESSIONS FOR THE HYDROGENOLYSIS OF N-BUTANE	
A1.1 Reaction Scheme	278
A1.2 Reaction Rates of the Reaction Species	278
APPENDIX 2. REACTION STOICHIOMETRY IN TUBULAR REACTORS	
A2.1 Introduction	283
A2.2 Tubular Reactor with No Axial Dispersion	283
A2.3 Tubular Reactors with Axial Dispersion	291
APPENDIX 3. PARAMETERS FOR LINEARIZED FINITE DIMENSIONAL REACTOR MODEL	
A3.1 Parameters for the Continuous Model	295
A3.2 Parameters for the Discrete Model	297
APPENDIX 4. SAMPLING PROPERTIES OF AUTO- CORRELATED FUNCTIONS	
A4.1 Introduction	301

	Page
A4.2 Weighted Least Squares Estimation	301
A4.3 Approximate Variances of the Weighted Autocorrelation Function	306

FIGURES

	Page
2.1 Temperature Profiles for Sensitive and Insensitive Reactors	24
2.2 Contours of Maximum Temperature Rise	33
2.3 Contours of Butane Conversion	33
2.4 Contours of Propane Selectivity	33
2.5 Temperature Profiles for Modified Reactor	35
2.6 Contours of the Maximum Temperature Rise: Modified Reactor	36
2.7 Contours of Butane Conversion: Modified Reactor	36
2.8 Contours of Propane Selectivity: Modified Reactor	36
4.1 Typical Temperature Profile	75
4.2 Objective Function v_4 Evaluated at Different Sensor Locations	81
5.1 Average Temperature Profile	95
5.2 Geometry of Principal Components	105
5.3 Sampled Cosine Wave	129
6.1 Frequency Response of the Filter $H(z)=1-\alpha_1 z^{-1}$	129
6.2 Frequency Response of $(1-(-.5+.68j)z^{-1}) \cdot (1-(-.5-.68j)z^{-1})$	135
6.3 Spectral Density of ∇U for Various Levels of Constraining	148

	Page
6.4 Spectral Density of Y for Various Levels of Constraining	147
6.5 Smith Deadtime Compensator	149
7.1 Adaptive Control Scheme	154
7.2 Reactor Control Configuration	167
7.3 Data from Reactor Run 1	171
7.4 Parameter Ratios for Reactor Run 1	172
7.5 Controller Poles for Reactor Run 1	173
7.6 Clarke's Algorithm - An Alternative Run	178
7.7 Operation at Extreme Conditions	180
7.8 Data from Reactor Run 2	182
7.9 Parameter Ratios for Reactor Run-2	183
7.10 Controller Poles for Reactor Run 2	184
7.11 Alternative Run Weighted Autocorrelation Functions for Reactor Run 1	201
8.1 Deadbeat Controller for Load Disturbance	205
8.2 Reconstruction Errors for Kalman Filter: var $\alpha(k) = 1.0$	243
8.3 Reconstruction Errors for Kalman Filter: var $\alpha(k) = .001$	244
8.4 A Distribution for $\alpha(k)$	247
8.5 Estimates for $\underline{x}(k)$	247

TABLES

	Page
2.1 Steady State Simulation Parameters	32
4.1 Sequential Sensor Location	79
4.2 Sequential Sensor Location - Multiple Sensors at the Same Position	79
4.3 Effect of Thermocouple Position on Controller Performance	83
5.1 AIC Statistic for Reactor Data	96
5.2 Predictor Covariance Matrix $\hat{\Sigma}_Y$ for AR(2) Model	100
5.3 Residual Covariance Matrix $\hat{\Sigma}_a$ for AR(2) Model	100
5.4 Eigenvalues and Eigenvectors of $\hat{\Sigma}_Y$	108
5.5 Expected Values of the Eigenvalues of $\hat{\Sigma}_Y$	108
5.6 Eigenvalues and Eigenvectors from Canonical Variate Analysis of Reactor Data	117
5.7 Canonical Variate Analysis for White Noise Input	122
5.8 Canonical Variate Analysis for Autocorrelated Input	122
6.1 The Effect of Constraining on the Variance of Y and ∇U	145

NOMENCLATURE

A	state transition matrix
$\underline{a}(k)$	random shock
\underline{C}	concentration of all reaction species
\underline{c}	concentration of effluent reaction species
C_{pg}	specific heat of gas, $J/(kg \cdot ^\circ K)$
C_{ps}	specific heat of solid, $J/(kg \cdot ^\circ K)$
D_{ea}	effective axial diffusivity (based on empty reactor volume), m^2/sec
D_{er}	effective radial diffusivity (based on empty reactor volume), m^2/sec
E	elemental matrix, Equation (2.10)
G	control matrix
G_m	mass flow rate based on area of empty reactor $kg/(m^2 \cdot s)$
G_0	superficial gas velocity, $m^3 gas/(m^2 reactor s)$
H	measurement matrix
Δh_i	heat of reaction i , $J/mole$
h_w	heat transfer coefficient at reactor wall, $W/(m^2 \cdot ^\circ K)$
$K(k)$	Kalman filter gain, Equation (3.16)
$L(k)$	controller feedback gain, Equation (3.11)
L	length of reactor, m
$\Delta n_{i,j}$	change in moles of species j in reaction i
Δn	change in moles of all species, Equation (2.8)

$N(k)$	disturbance affecting process output at time k
$\hat{N}(k+b/k)$	b step ahead forecast of disturbance, Equation (6.22)
$P(k/k)$	covariance of filtered state estimate, Equation (3.20)
$P(k+1/k)$	covariance of state prediction error, Equation (3.19)
$P_k^{(\alpha, \beta)}$	Jacobi polynomial of order k
$p(\underline{x}(k)/Y(\tau))$	conditional distribution of $\underline{x}(k)$ given all available output data to time τ , Equation (8.92)
Q	positive semidefinite state weighting matrix
R	positive semidefinite control weighting matrix or radius of reactor bed, m
$r_{\phi\phi}(k)$	sample autocorrelation at lag k , Equation (7.39)
$\bar{r}_{\phi\phi}(k)$	weighted sample autocorrelation at lag k , Equation (7.44)
R_v	covariance matrix of measurement errors
R_w	covariance matrix of process noise
S	stoichiometric matrix, Equation (2.7)
T	control interval
$T^{(N)}(r, \dot{z})$	N th order approximating polynomial, Equation (2.38)
T_0	temperature along central axis of reactor, $^{\circ}K$
T_w	wall temperature, $^{\circ}K$
$\bar{U}(k)$	manipulated variable
$\underline{u}(k)$	vector of manipulated variables
$\underline{v}(k)$	measurement noise
v_c	concentration wave velocity $G_0/(E \cdot L)$, $m/(s \cdot m)$

v_T	thermal wave velocity, $G_0 C_{pg} \rho_g / (L\bar{c})$, m/(s·m)
$w(k)$	process noise
$x(k)$	state vector at time k
$\hat{x}(k+1/k)$	estimate for $x(k+1)$ given information to time k+1
$\hat{x}(k/k)$	estimate for $x(k)$ given information to time k
$Y(k)$	process output
$y(k)$	vector of process output
$\hat{Y}(k/k-1)$	estimate for $Y(k)$ given information to time k-1
$z(k)$	vector of deterministic disturbances
z^{-1}	backward shift operator such that $z^{-b}Y(k) = Y(k-b)$
\bar{c}	specific heat term, $C_{ps} \rho_B + C_{pg} \rho_g \epsilon$
B_i	Biot number, $h_w R / \lambda_{er}$

Greek Letters

$\underline{\alpha}(k)$	"occasional shock" at time k
$\nu_{i,j}$	stoichiometric coefficient of species j in reaction i
$\delta(z^{-1})$	denominator polynomial in z^{-1} of dynamic transfer function, Equation (6.16)
ϵ	void fraction, m^3 gas in voids/ m^3 empty reactor
$\epsilon', \epsilon'', \lambda$	penalty on control action
$\theta(z^{-1})$	numerator polynomial in z^{-1} of disturbance transfer function, Equation (6.17)
λ_{er}	effective radial thermal conductivity, W/(m·°K)
ρ_B	bulk density of catalyst, kg/m^3
ρ_g	gas density, kg/m^3

ρ_j	sample autocorrelation at lag j
\mathbf{P}	covariance matrix
σ^2	variance
τ_d	transport delay, time
ϕ	phase angle, Equation (6.8)
$\phi(k)$	variable for Clarke's algorithm, Equation (7.20)
$\phi(z^{-1})$	denominator polynomial in z^{-1} of disturbance transfer function, Equation (6.17)
ω	frequency, rad/sec
$\omega(z^{-1})$	numerator polynomial in z^{-1} of dynamic transfer function, Equation (6.16)
ω_s	Nyquist frequency, radians
∇	$1-z^{-1}$

Symbols

$ $	determinant of a matrix or magnitude of a complex number
$ $	Euclidian norm of a matrix
$\arg()$	phase angle of a complex number

Subscripts

∞	converged solution
----------	--------------------

Superscripts

$'$	transpose of a matrix
$()^{-1}$	inverse of a matrix
$()^\#$	pseudo inverse of a matrix
$()^\infty$	converged solution

affecting the process. In the process industries, disturbances are often of a random nature. When a suitable mathematical structure for these disturbances can be postulated, or estimated from experimental data, this additional information can be incorporated in the design of the controller. Hopefully, by designing the controller to guard against the inherent process disturbances, one obtains improved control. Time series modelling and stochastic control provide, in theory at least, a powerful framework for the modelling and control of processes subject to disturbances of a random nature.

The applicability of time series modelling and stochastic control as a means of regulating a complex process is investigated in this thesis. This work represents part of an ongoing study on the modelling and control of tubular reactors. Many topics considered in this thesis are motivated by questions unresolved from previous theoretical and experimental studies on modelling and control of a pilot scale catalytic packed bed tubular reactor carrying out the hydrogenolysis of n-butane. Control of the temperature profile or effluent concentrations has proven to be difficult. As a result, there has been an incentive to explore the use of advanced control schemes to regulate this process. In almost all cases, the problems discussed and their solutions are illustrated with experimental data collected by previous experimentalists, or by application of stochastic control to regulate this pilot scale reactor. The use of real

CHAPTER 1

INTRODUCTION

The control of catalytic packed bed tubular reactors carrying out highly exothermic gas phase reactions represents one of the most challenging control problems in the chemical industry. These reactors are highly non-linear, distributed parameter, multivariable processes. Although it is usually desired to achieve control over the final conversion and product selectivities, on-line measurement of these via gas chromatographs, etc. is often too slow to be of any use in a direct feedback control scheme. Therefore, in the control of industrial reactors one usually relies on temperature measurements throughout the catalyst bed; a common situation being simple feedback from some measure of the hot-spot (maximum) temperature in the bed. This provides an indirect stabilization of the exit conversion and safeguards against dangerous temperature excursions. For more direct control over the exit conversion and product selectivities one must rely upon some form of inferential control based on complex multi-variable models of the reactor.

The availability of these models does not insure that one will be able to exert satisfactory control over the variables of interest. The performance of a controller is influenced by the adequacy of the dynamic model and the types of disturbances

data and verification of theory on a difficult process to control, demonstrate the usefulness of time series modelling and stochastic control. A synopsis of the thesis follows.

In Chapter 2, the modelling of the pilot scale reactor studied in this project is reviewed. Attention is focused on the concept of independence of reaction. In spite of the complexity of the mass and energy balances describing this process, certain linear combinations of the concentrations within the reactor are found to be constant. These linear combinations define the reaction stoichiometry. The sensitivity of the steady state temperature profile to perturbations in the inlet flowrates is examined. The results of this analysis help explain some of the difficulties previously encountered in controlling this reactor.

The basic control theory used throughout this thesis is presented in Chapter 3. Experimental and theoretical applications of modern control theory on chemical reactors are reviewed. A number of extensions to the model developed by Jutan (1976) for the reactor studied in this work are put forth. These results are used in subsequent chapters.

The optimal location of measurement sensors is examined in Chapter 4. The location and number of measurement sensors influences the quality of control in processes subject to stochastic disturbances. The theory proposed in the literature to locate measurement sensors is critically reviewed and a number

of modifications proposed. This theory is illustrated with an application of the optimal location of temperature sensors along the central axis of the butane hydrogenolysis reactor. The results in this chapter were not experimentally verified, but rather relied on the dynamic model of the process developed by Jutan (1976).

Many times a dynamic model developed from mechanistic arguments, that is suitable for control, is unavailable. This may be due to the time and effort required to obtain this model. An empirical model of the process may be constructed from input/output data collected from a designed experiment. The use of multivariate autoregressive models as a means of obtaining a model suitable for control is investigated in Chapter 5. In a model fit to reactor data there are seven outputs (temperatures along the central axis of the reactor) and two inputs (flow rates of butane and hydrogen). Since the temperatures must satisfy the mass and energy balances, we would not anticipate that perturbations in the manipulated variables would result in independent variation in all the temperatures. Nor would we suppose that all the temperatures can be predicted equally well. Principal component analysis is used to establish the number of degrees of freedom in the reactor temperature profile. Canonical variate analysis of time series is used to discover which temperatures can be predicted the most accurately. The use of autoregressive model fitting in conjunction with these multivariate statistical

techniques are powerful methods of analyzing dynamic data.

There is considerable interest in the design and implementation of controllers for single input, single output dynamic stochastic systems. This stems in part from the relative ease (compared to multivariate systems) for which the structure and parameters of the model can be estimated. The resulting controllers have some very unusual spectral characteristics and these are examined in Chapter 6. The manner in which these stochastic controllers compensate for deadtime arising from process transport delay is compared to classical techniques for deadtime compensation.

Of all the modern control theories, those which have perhaps the most industrial applications are adaptive versions of the controllers discussed in Chapter 6. These adaptive controllers are known as self-tuning regulators. In Chapter 7, the pertinent theory for industrial application of these regulators is reviewed. A number of different self-tuning algorithms are applied to control a pilot scale chemical reactor. Additional theory is developed to explain a number of experimental observations.

In Chapter 8, an attempt is made to unify many of the concepts involved in control of stochastic and deterministic processes. The use of stochastic difference equations to model deterministic and random disturbances are investigated. Problems in controlling processes subject to both stochastic and deter-

ministic disturbances, a common industrial situation, are discussed. A number of solutions are proposed but these are of theoretical interest only, as they are too complex to implement.

Conclusions and recommendations are discussed in Chapter 9.

Although, at first glance, the topics may appear to be unrelated, the central theme is the application of statistics and stochastic control in the design and analysis of control schemes. Throughout this work, the merit of mechanistic models and empirical models are compared. The overriding conclusion from the work is that the successfulness of a control scheme depends on adequately modelling the dynamics and disturbances affecting the process.

CHAPTER 2

TOPICS IN REACTOR MODELLING

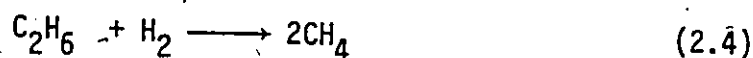
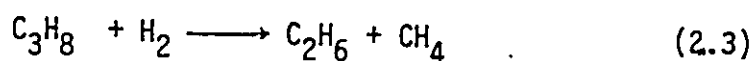
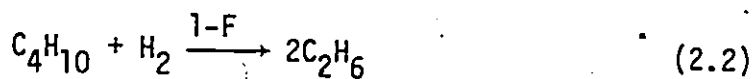
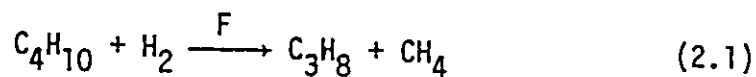
2.1 Introduction

In this chapter, the stoichiometry for the hydrogenolysis of n-butane on a silica on nickel gel catalyst is reviewed. Particular attention is paid to the concept of independence of reaction. The mass and energy balances used in previous modelling studies for a packed bed reactor are reexamined. An expression is derived to explain the time and spatial behaviour of certain linear combinations of the chemical species in the reactor. As a result of this analysis, it is apparent that it is not necessary to include all species in the material balance. The sensitivity of the temperature profile in the reactor to perturbations in inlet flow rates is also studied for the existing reactor geometry. This analysis helps explain some of the difficulties encountered in previous control studies on this reactor. Finally, a modification to the reactor to reduce this parametric sensitivity is proposed.

2.2 Reaction Scheme

The reaction studied is the hydrogenolysis of n-butane over a nickel on silica gel catalyst in a tubular packed bed reactor. The reaction mechanism has been studied by Orlickas

(1970, 1972), and Shaw (1972, 1974), and the following reactions were proposed to represent this hydrogenolysis.



The following assumptions were made:

- 1) Butane and propane are adsorbed on the catalyst surface whereon a surface catalyzed reaction occurs;
- 2) The reaction products from these reactions may desorb or react further;
- 3) Reactions converting butane and propane to methane are presumed not to occur due to the low probability of breaking two or three carbon bonds simultaneously.

Models for the net rates of production of each of the components have been built, Orlickas (1972), and are presented in Appendix 1. The kinetic parameters associated with these rate expressions were calculated from experimental data and are also reported in this appendix.

Examination of these reactions reveals that the stoichiometry of the fourth reaction may be expressed as a linear combination of the stoichiometry of the first three. (It is incorrect though to say that the fourth reaction rate is a linear combination of the first three reaction rates.) One usually then says that there are three independent reactions. However, it was assumed in studying the reaction kinetics, that the rate of reactions (2.1) and (2.2) were in a constant proportion, $F/1-F$. In general, there may be restrictions among the reactions due to kinetic constraints, reaction vessel configuration, electromagnetic effects, etc. It is desirable, therefore, to clarify our definition of independence of reaction.

2.3 Reaction Stoichiometry and Independence of Reaction

The concept of independence of reactions has been studied by a number of people, Aris and Mah (1962), Denbigh (1971). Reacting systems involving isomers have been studied by Schubert (1974). In this section we will try to distinguish between two types of reaction independence.

It is assumed that there are s chemical species, r chemical reactions and n elements. The extent of reaction i , ξ_i , is defined as, Denbigh (1971)

$$\xi_i = \frac{\Delta n_{i,1}}{\nu_{i,1}} = \frac{\Delta n_{i,2}}{\nu_{i,2}} = \dots = \frac{\Delta n_{i,s}}{\nu_{i,s}}, \quad i = 1, 2, \dots, r \quad (2.5)$$

$\Delta n_{i,j}$ is the change in moles of species j in reaction i . The stoichiometric coefficient of species j in reaction i , $v_{i,j}$ is usually taken positive for products of reaction and negative for reactants. It is only defined to a scaling factor. Using (2.5), the total change in moles of species j is therefore given by

$$\Delta n_j = \sum_{i=1}^r v_{i,j} \cdot \xi_i, \quad j = 1, 2, \dots, s \quad (2.6)$$

Define the rxn stoichiometric matrix S as

$$S = \{v_{i,j}\}, \quad i = 1, 2, \dots, r, \quad j = 1, 2, \dots, s \quad (2.7)$$

The changes in moles for all species can be written compactly as

$$\Delta \underline{n} = S' \underline{\xi} \quad (2.8)$$

For the reactions involving the hydrogenolysis of n-butane, we have

$$\Delta \underline{n} = \begin{bmatrix} \Delta C_4H_{10} \\ \Delta C_3H_8 \\ \Delta C_2H_6 \\ \Delta CH_4 \\ \Delta H_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 1 & 2 \\ -1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} \quad (2.9)$$

If (2.8) were differentiated, with respect to time, one would have an expression for the net rate of reaction for all the species in terms of the rates of reaction i , $i=1,2,\dots,r$.

The number of stoichiometrically independent reactions is the minimal number of reactions that will represent the complete stoichiometry of the reacting system, Denbigh (1971). The number of stoichiometrically independent reactions is defined as the number of linearly independent rows of S , r_s , i.e., the ranks of S , Aris and Mah (1962), Denbigh (1971). It is important to realize that no constraints due to kinetics, electroneutrality, etc. have been introduced.

The terminology 'stoichiometrically independent reactions' is somewhat of a misnomer in that it implies that there may be 'dependent reactions'. The extent of these 'dependent reactions' would then appear to be linear functions of the independent extents. The true meaning is apparent from (2.8). When the rank of S is less than the number of reactions, the mole changes are not explained by a unique set of extents. If we arbitrarily assign values to $r-r_s$ extents, the mole changes can then be explained uniquely by the remaining r_s extents. (If $r-r_s$ extents are assigned the value zero, the remaining reactions are a set of 'maximal independent reactions', Schneider and Reklaitis (1975).) The arbitrary assignment of $r-r_s$ extents to explain the change in moles of the species does not imply that these extents of reaction are dependent on the extents of the remaining r_s reactions.

Define the $n \times s$ elemental matrix E as

$$E = \{e_{\ell,j}\}, \ell = 1, 2, \dots, n, j = 1, 2, \dots, s \quad (2.10)$$

The elemental coefficient $e_{\ell,j}$ is the number of atomic units ℓ in species j . Conservation of atoms requires that

$$ES' = 0 \quad (2.11)$$

Therefore by combining (2.8) and (2.10) we have

$$E\Delta n = 0 \quad (2.12)$$

For the reactions (2.1-2.4), E is given by

$$E = \begin{bmatrix} 4 & 3 & 2 & 1 & 0 \\ 10 & 8 & 6 & 4 & 2 \end{bmatrix} \quad (2.13)$$

Using (2.12) the change in moles of methane and hydrogen can be found from the changes in moles of the remaining species, as

$$\Delta CH_4 = -4\Delta C_4H_{10} - 3\Delta C_3H_8 - 2\Delta C_2H_6 \quad (2.14)$$

$$\Delta H_2 = 3\Delta C_4H_{10} + 2\Delta C_3H_8 + \Delta C_2H_6 \quad (2.15)$$

In a closed system at chemical equilibrium, the rank of E , r_e , is known as the number of components (Brinkley (1946)), the minimum number of species which must be used to prepare the equilibrium mixture, Denbigh (1971). For a perfectly mixed, closed reaction vessel, (2.12) indicates that $n - r_e$ component mass balances must be solved to determine the composition of the reacting mixture.

Equation (2.11) can be used to give an upper bound on the number of stoichiometrically independent reactions. The columns of S' are in the null space of E , which is of dimension $n-r_e$, Noble (1969). Since the columns of S' are therefore spanned by the null space of E , the rank of S is less than or equal to $n-r_e$.

Equations (2.8), (2.11) and (2.12) apply to composition changes in a finite, or elemental volume that are due solely to chemical reaction.

It may be observed experimentally that the change in moles of reacting species can be explained by fewer reactions than predicted from the rank of the stoichiometric matrix. This indicates that some reactions do not occur, or that they occur in direct proportion to other reactions. These additional relationships arise from kinetic, electroneutrality restrictions, etc. It is convenient then to define the number of kinetically independent reactions as the minimal set of reactions that will explain the stoichiometry and kinetics of the reacting system.

The r_a kinetic restrictions on the extents may be written as

$$A\underline{\xi} = [A_1:A_2] \begin{bmatrix} \underline{\xi}^* \\ \underline{\xi}^{**} \end{bmatrix} = \underline{0} \quad (2.16)$$

where $\underline{\xi}^{**}$ denotes the set of extents among which no linear relationships exist. A is of rank r_a . Using (2.16) with (2.8),

we may write for the change in moles

$$\underline{\Delta n} = S' \begin{bmatrix} -A_1^{-1} A_2 \\ \hline I \end{bmatrix} \cdot \underline{\xi}^{**} \quad (2.17)$$

The matrix

$$S^{*'} = S' \begin{bmatrix} -A_1^{-1} A_2 \\ \hline I \end{bmatrix} \quad (2.18)$$

may be considered as our new stoichiometric matrix. The rank of the matrix S^* is the number of kinetically independent reactions. If S is of full column rank, the effect of the r_a additional constraints is to reduce the number of reactions required to represent the stoichiometry and kinetics of the reacting system to $r-r_a$. However, when S is not of full column rank, the number of kinetically independent reactions is not always less than the number of stoichiometrically independent reactions.

With reference to the n-butane hydrogenolysis reactions, we have observed that there are three stoichiometrically independent reactions. However, the kinetic model developed for this reaction assumes that of the butane that reacts, a fraction F reacts to form propane, and a fraction $1-F$ reacts to form ethane. The matrix of kinetic restrictions A is, therefore, given by

$$A = \begin{pmatrix} -(1-F) & F & 0 & 0 \end{pmatrix} \quad (2.19)$$

The stoichiometry of the reacting mixture is then given by

$$\underline{n} = \begin{bmatrix} -1/(1-F) & 0 & 0 \\ F/(1-F) & -1 & 0 \\ 2 & 1 & -1 \\ F/(1-F) & 1 & 2 \\ -1/(1-F) & -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} \quad (2.20)$$

Reactions two, three and four have been chosen as our reaction basis. The rank of S^* in (2.20) is three. Thus, the kinetic constraint has not reduced the number of reactions required to completely specify the reaction stoichiometry and kinetics.

2.4 Description of the Packed Bed Tubular Reactor

The hydrogenolysis reaction was carried out by Jutan (1976) and Tremblay (1977) in a bench scale tubular packed bed reactor. Due to the highly exothermic nature of the reactions, cooling at the tube wall is necessary to prevent large temperature excursions. This gives rise to radial temperature gradients and consequently radial concentration gradients. Axial dispersion of heat and mass have been neglected as these effects are negligible compared to the convective terms. The reactor is operated at pressures of 100-200 kPa, with a pressure drop of 20 kPa. At these pressures, the gases are assumed to obey the ideal gas law. The small catalyst particles, 1 mm, are assumed to be isothermal and at the same temperature as the

surrounding gas. The velocity profile is presumed to be flat, and not a function of axial distance. Finally, the reactor is treated as a pseudohomogenous mixture because of the large tube to particle diameter ratio, 400/1. Jutan (1977a) gives detailed justifications for these assumptions.

The mass and energy balances describing the process are (Jutan (1977a))

$$\frac{\partial C}{\partial t} = -v_c \frac{\partial C}{\partial z} + \frac{D_{er}}{eR^2} \frac{\partial}{\partial r} \left(r \frac{\partial C}{\partial r} \right) + \frac{\rho_B}{\epsilon} S' \xi \quad (2.21)$$

and

$$\frac{\partial T}{\partial t} = -v_T \frac{\partial T}{\partial z} + \frac{\lambda_{er}}{cR^2} \frac{\partial}{\partial r} \left(r \frac{\partial T}{\partial r} \right) - \sum_{i=1}^4 \frac{\Delta h_i \xi_i \rho_B}{\bar{c}} \quad (2.22)$$

The boundary conditions are

$$r = 0: \quad \frac{\partial C}{\partial r} = 0, \quad (\text{symmetry})$$

$$\frac{\partial T}{\partial r} = 0 \quad (2.23)$$

$$r = 1: \quad \frac{\partial C}{\partial r} = 0,$$

$$\frac{\partial T}{\partial r} = B_j (T_w - T) \quad (2.24)$$

$$\begin{aligned}
 z = 0: \quad T &= T_w \text{ for all } r \\
 \underline{C} &= \underline{C}(t, r, 0)
 \end{aligned}
 \tag{2.25}$$

$$\begin{aligned}
 t = 0: \quad T &= T(0, r, z) \\
 \underline{C} &= \underline{C}(0, r, z)
 \end{aligned}
 \tag{2.26}$$

There has been a slight departure from the notation of Jutan in that the mass balance has been written in vector notation. All species are included in the vector of concentrations \underline{C} . Jutan (1977a) included only three components in his mass balance. The concentrations of the remaining two components were calculated from (2.12). However, this is only correct for the steady state versions of (2.21) and (2.22). This is discussed further in the next section.

In the heat generation term, the summation includes all reactions and not only those which are 'independent'. In Jutan (1977a), the summation extended over the first, second and fourth reaction in (2.1 - 2.4). The heat of reaction for the production of ethane from propane was omitted. This probably did not affect the results of his work significantly because the extent of this reaction was very small in his work. At higher temperatures, however, the extent of this reaction increases.

2.5 Stoichiometry in Tubular Reactors

In section 2.2, we investigated independence of reaction, and reaction stoichiometry for a closed reaction vessel. The

emphasis there was to investigate these topics without the need to introduce reactor models. In this section, we will see how the particular choice of reactor model modifies our reaction stoichiometry.

To determine the reaction stoichiometry, multiply the mass balance (2.21) by the elemental matrix E . The term $S' \underline{\dot{x}}_P / \epsilon$ is the net rate of production of all the species at a particular point, or in an infinitesimal volume in the reactor. By the conservation of atoms (2.11), ES' vanishes and we obtain

$$\frac{\partial \underline{EC}}{\partial t} = -v_c \frac{\partial \underline{EC}}{\partial z} + \frac{D_{er}}{R^2} \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \underline{EC}}{\partial r} \right) \quad (2.27)$$

The boundary conditions associated with (2.27) are

$$r = 0: \quad \frac{\partial \underline{EC}}{\partial r} = 0 \quad (\text{symmetry}) \quad (2.28)$$

$$r = 1: \quad \frac{\partial \underline{EC}}{\partial r} = 0 \quad (2.29)$$

$$z = 0: \quad \underline{EC} = \underline{EC}(t, r, 0) \quad (2.30)$$

$$t = 0: \quad \underline{EC} = \underline{EC}(0, r, z) \quad (2.31)$$

The solution of the vector linear partial differential equation (2.27) defines the reaction stoichiometry for all time and spatial co-ordinates in the tubular reactor. The solution of

(2.27) is detailed in Appendix 2. We find that if the feed is distributed across the reactor entrance such that

$$\frac{\partial \underline{EC}(t,r,0)}{\partial r} = 0 \quad (2.32)$$

then the reaction stoichiometry as derived in Appendix 2 is given by

$$E(\underline{C}(t,r,z) - \underline{C}(0,r,z)) = E(\underline{C}(t-z/v_c, r, 0) - \underline{C}(0,r,0)) \quad (2.33)$$

where

- 1) $\underline{C}(0,r,z)$ is the feed composition at $t=0$. This corresponds to having the reactor initially filled with gas;

or

- 2) $\underline{C}(0,r,z)$ is the steady state solution to (2.21) and (2.22).

If the feed is distributed evenly across the reactor inlet then (2.32) is satisfied. This would be the case in most reactor systems.

The results are independent of the dispersion coefficient D_{er} . The expression z/v_c is the time required by the gas to travel the distance z . The significance of the result (2.32) is that we can compute the complete reaction stoichiometry without having to include in the mass and energy balances all s species. Instead, we need only include $s-r_e$ species. It may be no more

difficult to solve the differential equations for all species. The most time consuming step in solving the partial differential equations is the evaluation of the rate expression. However, the local rate expressions for the dependent species are linear combinations of those for the independent species. The most important drawback to the inclusion of all species in the material balance is the extra memory required to solve the additional differential equations.

To compute the response of the process to changes in the feed composition, we need only 'remember' past values of the feed composition for the reactor residence time. In most gas phase reactions, this transport lag is much less than the time required to observe a significant thermal response. In this instance, one might be inclined to use the steady state version of (2.33), that is

$$\underline{EC}(\infty, r, z) = \underline{EC}(0, r, z) \quad (2.34)$$

where $\underline{C}(0, r, z)$ is the feed concentration in the reactor at $t=0$. In the butane hydrogenolysis reactor, $v_T/v_C \sim 500/1$, and so the use of (2.33) was justified. In liquid phase systems, the temperature and concentration waves are of comparable size, and significant errors can result if the steady state version (2.34) is used instead of (2.33).

Gould (1969), Asbjornsen and Fjeld (1970), Fjeld et al (1974) and more recently Hammarstrom (1979), have studied

a similar topic known as reaction invariants. Their objectives are to find linear subspaces (reaction invariants) of the non-linear mass and energy balances describing the behaviour of continuous stirred tank reactors. If these subspaces exist, it may be possible to eliminate some variables from the ordinary differential equations describing the process. Equation (2.33) describes a true reaction invariant for a tubular reactor. The implication of reaction invariants in stability and control of stirred tank reactors is discussed by Fjeld et al (1974) and Hammarstrom (1979). However, the extension of their work to tubular reactors is beyond the scope of this project. It is obvious though that it is impossible to control all s species at arbitrary values when the changes in moles of some of the species are constrained by changes in moles of the remaining species.

The results in this section have been derived for a tubular reactor with radial dispersion of mass and energy. Since (2.33) is independent of these transport terms, this result also applies to isothermal and adiabatic tubular reactors. This result does not apply to tubular reactors with axial dispersion. In reactors with axial dispersion, one doesn't find an expression for the reaction stoichiometry that is particularly useful (see Appendix 2). It is necessary in the unsteady state, then, to include all species in the mass balance. The steady state reaction

stoichiometry is given by (2.34), with the radial dependence omitted, Appendix 2.

2.6 Temperature Sensitivity and Reactor Runaway

In a nonadiabatic, nonisothermal tubular reactor, one usually observes a 'hot spot' or maximum in the temperature profile. It is not desirable to have a 'dramatic' hot spot (Froment, (1967)) as this can lead to catalyst sintering, and increase the production of undesired side products. The operating conditions should not be such that small changes in the reactant flow rates, concentration or temperature, propagate into large changes in the reactor temperature profile. Previous experimental studies on this reactor exhibited this latter problem. To gain some insight into the causes, and possible solutions to this phenomenon, the parametric sensitivity of the reactor model is examined in this section.

Parametric sensitivity here refers to the sensitivity of the temperature profile to changes in the operating conditions and not to changes resulting from variations in the kinetic or mass transfer coefficients.

Barkeley (1959) and Van Welsenaere and Froment (1970) discussed parametric sensitivity and reactor runaway for kinetically simple reactions, and reactor models involving no radial gradients. The essence of their papers is that one observes two fundamentally different types of temperature profiles in a nonadiabatic,

nonisothermal reactor, Figure 2.1: One is bowl shaped. The other is more peaked, having an inflection point in the temperature profile prior to the hot spot. This latter profile is more sensitive to small changes in flow, composition and temperature of the feed stream. The operating conditions which did not give rise to inflection points were deemed feasible steady state operating conditions. The envelope of these desirable operating conditions is also affected by the dimensions of the reactor and activity of the catalyst.

Due to the simplicity of the reactor model and kinetics, studied in these papers, changes in the configuration of the process could be analytically investigated. To study the effect of process changes with our model and kinetics, requires the steady state solution of (2.21) and (2.22), a set of coupled, highly nonlinear differential equations in two spatial co-ordinates.

In the next section, we outline the solution of these equations by the technique of orthogonal collocation. In the section following that, the sensitivity of the reactor temperature to changes in feed rates is examined. The results of that section help to explain many of the difficulties encountered in controlling this reactor.

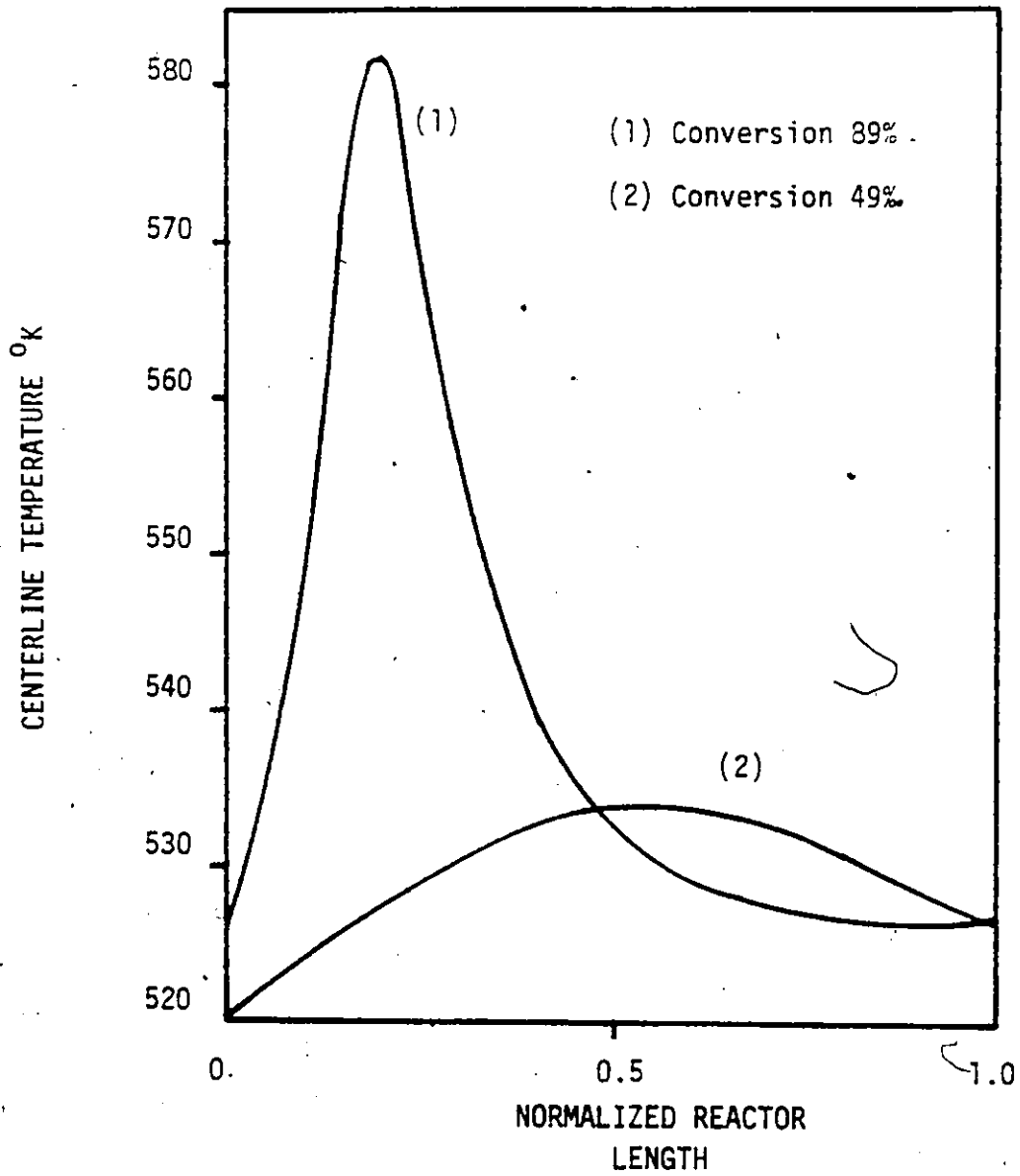


Figure 2.1: Temperature Profiles for Sensitive and Insensitive Reactors.
Jutan (1976)

2.6.1 Solution of the Partial Differential Equations by Orthogonal Collocation

The steady state versions of (2.21) and (2.22) do not have an analytic solution, and it is necessary to use a numerical technique to find the solution of these equations. In the orthogonal collocation technique, the first step in solving the partial differential equations is to approximate the radial profiles of temperature and concentration by orthogonal functions. This results in an enlarged set of ordinary differential equations which can then be solved by orthogonal collocation in the axial direction or by numerical techniques for the solution of ordinary differential equations. Jutan (1976) outlines the use of orthogonal collocation and reviews the application of this technique to chemical reactors. However, certain aspects were not well described, and no numerical results were given. In this section, the ideas of orthogonal collocation will be briefly sketched.

Experimental evidence suggests that the radial temperature profile is nearly quadratic. One might then approximate the radial profile by

$$T(r,z) = T(l,z) + a(z)(r^2-1) \quad (2.35)$$

where $a(z)$ is a coefficient yet to be determined. Substituting (2.35) into the steady state energy balance, we obtain

$$r = 1: \quad v_T \frac{\partial T(1,z)}{\partial z} = \frac{4\lambda_{er}}{\bar{c}R^2} a(z) - \sum_{i=1}^4 \frac{\Delta h_i \dot{\xi}_i \rho_B}{\bar{c}} \quad (2.36)$$

and

$$r = 0: \quad v_T \frac{\partial (T(1,z) + a(z))}{\partial z} = \frac{4\lambda_{er}}{\bar{c}R^2} a(z) - \sum_{i=1}^4 \frac{\Delta h_i \dot{\xi}_i \rho_B}{\bar{c}} \quad (2.37)$$

The partial differential equations have been transformed to simultaneous ordinary differential equations in $T(1,z)$ and $a(z)$. The radial concentration profile is nearly quartic. This is the lowest degree polynomial, except for zero, which satisfies the boundary conditions. The mass balance is also transformed to an enlarged set of ordinary differential equations by choosing a suitable approximating function.

The approximating polynomial chosen may not be very good. It may distribute the error between the true solution (if it were known) and the approximate solution very poorly. An excellent discussion of approximating polynomials is given by Villadsen and Michelson (1977). They consider examples where the differential equation has an analytic solution. Several approximating methods are compared. The maximum error and distribution of the error over the interval of interest are examined. For the particular examples chosen, orthogonal polynomials formed an excellent class

of approximating polynomials. They do not recommend using (1, $\sin x$, $\sin 2x$, . . .) since these are usually approximated by polynomials.

Consider approximating a radial temperature profile by the N th order polynomial (in r^2)

$$T^{(N)}(r,z) = T^{(N)}(1,z) + (1-r^2) \sum_{k=0}^{N-1} a_k(z) p_k^{(\alpha,\beta)}(r^2) \quad (2.38)$$

where $p_k^{(\alpha,\beta)}(r^2)$ is a suitably chosen Jacobi polynomial.

Substitution of (2.38) into the energy balance results in an enlarged set of ordinary differential equations in $T^{(N)}(1,z)$ and $a_k(z)$. The solution is inconvenient in that it is expressed in terms of the coefficients $a_k(z)$ and not in terms of temperatures in the reactor. If one restricts attention to the zeroes of the N th order Jacobi polynomial (in r^2), which are known as the interior collocation points, and $r=1$, then the derivatives of (2.38), at these points can be expressed as (Villadsen and Stewart (1967))

$$\frac{\partial T^{(N)}(r_i, z)}{\partial r} = \sum_{j=1}^{N+1} A_{ij} T^{(N)}(r_j, z), \quad i=1,2,\dots,N+1 \quad (2.39)$$

and

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial T^{(N)}(r_i, z)}{\partial r} \right) = \sum_{j=1}^{N+1} B_{ij} T^{(N)}(r_j, z), \quad i=1,2,\dots,N+1 \quad (2.40)$$

An algorithm for evaluating (A_{ij}, B_{ij}) is given in Villadsen and Stewart (1967). The derivation of (2.39) and (2.40) is not clear though.

Collocation techniques have been traditionally used to numerically evaluate integrals of the form

$$I_1 = \int_0^1 T(r,z) (r^2)^\beta (1-r^2)^\alpha r dr \quad (2.41)$$

Given a trial function for $T^{(N)}(r,z)$, the expression

$$I_2 = \int_0^1 (T(r,z) - T^{(N)}(r,z)) (r^2)^\beta (1-r^2)^\alpha r dr \quad (2.42)$$

is a measure of the accuracy of the approximation. The parameters in the approximating polynomial are chosen typically so that the integrand in (2.42) vanishes at a certain number of points in the interval $(0,1)$, or that the integrated error is distributed evenly, etc. Suppose the approximating function is chosen as the Lagrange interpolating polynomial

$$T^{(N)}(r,z) = \sum_{i=1}^{N+1} \prod_{\substack{j=1 \\ j \neq i}}^{N+1} \frac{r^2 - r_j^2}{r_i^2 - r_j^2} T^{(N)}(r_j, z) \quad (2.43)$$

If the interpolation points r_i^2 , $i=1,2,\dots,N$ are chosen as the zeroes of the N th order Jacobi polynomial $P_N^{(\alpha,\beta)}$, and $r_{N+1}^2=1$, then the integral I_2 vanishes if $T(r,z)$ is an even polynomial of degree

less than $4N$, Villadsen and Michelsen (1977). If one differentiates (2.43) and evaluates the derivatives at the collocation points, the coefficients (A_{ij}, B_{ij}) in Villadsen and Stewart (1967) are obtained.

We achieve a greater insight into the orthogonal collocation technique when it is realized that a Lagrange interpolating polynomial is being used to approximate the unknown function. The manner in which the derivative expressions (2.39 and 2.40) are obtained is now apparent. The function, in our case temperature, can easily be evaluated at other spatial co-ordinates using (2.43).

2.6.2 Application of Orthogonal Collocation to the Tubular Reactor Equations

Jacobi polynomials $P_k^{(1,0)}(r^2)$ were used by Jutan (1977b) to reduce the set of partial differential equations to ordinary differential equations. The radial profile is observed to be nearly quadratic. The appropriate Lagrange interpolating polynomial is

$$T^{(1)}(r,z) = \frac{3(1-r^2)T^{(1)}(\sqrt{3},z) - 1(1-3r^2)T^{(1)}(1,z)}{2} \quad (2.44)$$

The interpolation points are taken as $r=1$ and the zero of $P_1^{(1,0)}(r^2)$. The radial concentration profile is nearly quartic. The appropriate Lagrange interpolating polynomial is

$$\begin{aligned}
 c^{(2)}(r,z) = & \frac{(r^2 - r_B^2)(r^2 - r_C^2)}{(r_A^2 - r_B^2)(r_A^2 - r_C^2)} c^{(2)}(r_A,z) \\
 & + \frac{(r^2 - r_A^2)(r^2 - r_C^2)}{(r_B^2 - r_A^2)(r_B^2 - r_C^2)} c^{(2)}(r_B,z) \\
 & + \frac{(r^2 - r_A^2)(r^2 - r_B^2)}{(r_C^2 - r_A^2)(r_C^2 - r_B^2)} c^{(2)}(r_C,z)
 \end{aligned} \tag{2.45}$$

where $r_A = 1$, $r_B = .803$, $r_C = .393$, are the zeroes of $P_2^{(1,0)}(r^2)$. Extensive use was made of the boundary conditions (2.23), (2.24) and (2.25) to find additional linear relationships among the temperatures and concentrations, Jutan et al (1977b). The final result is that one must solve the following ordinary differential equations

$$\begin{aligned}
 r = 0: \quad v_C \frac{\partial c(0,z)}{\partial z} = & \frac{D_{er}}{R^2 \epsilon} (-8c(0,z) + 8c(1,z)) \\
 & + \frac{\rho_B S' \dot{\xi}}{\epsilon}
 \end{aligned} \tag{2.46}$$

$$r = 0: \quad v_T \frac{\partial T(0,z)}{\partial z} = \frac{4\lambda_{er} B_i}{R^2 \bar{c}(B_i+2)} (T_w - T(0,z))$$

$$+ \frac{\sum_{i=1}^4 \Delta h_i \dot{\xi}_i \rho_B}{\bar{c}} \tag{2.47}$$

$$r = 1: \quad v_c \frac{\partial C(1,z)}{\partial z} = \frac{D_{er}}{R^2 \epsilon} (8C(0,z) - 8C(1,z)) + \frac{\rho_B S' \dot{\xi}}{\epsilon} \quad (2.48)$$

$$r = 1: \quad \frac{2}{B_i + 2} T(0,z) + \frac{B_i}{B_i + 2} T_w^\circ = T(1,z) \quad (2.49)$$

The superscript on $C(r,z)$ and $T(r,z)$ has been dropped. Since numerical values were not reported by Jutan (1977b), all calculations outlined in his work were repeated.

Collocation can also be used to approximate the axial derivatives, Finlayson (1971), Jutan (1977b). One then finds that the axial temperatures and concentrations satisfy a set of nonlinear algebraic expressions. Since the solution of nonlinear algebraic equations is difficult, the nonlinear differential equations were solved using a high order Runge-Kutta method.

2.6.3 Solution to Ordinary Differential Equations

Typical temperature profiles are shown in Figure 2.1. These results were obtained by Jutan (1976) and closely matched experimental data. There is an inflection in the centreline temperature in one of the profiles indicating that this is a sensitive operating point, Barkelew (1959).

To further investigate the sensitivity of the reactor system, the ordinary differential equations (2.46-2.49) were integrated in the region where butane and hydrogen could be varied. The wall temperature was held constant. The simulation parameters are listed in Table 2.1.

Biot Number, $B_i = 43.5$
Bulk density of catalyst, $\rho_B = .72 \text{ g/cm}^3$
Catalyst activity = 4.64
Effective radial thermal conductivity: $\lambda_{er} = 3.78 \times 10^{-4} + (T_w^4 - T^4(0,z))$ $\times 1.3 \times 10^{-13} \text{ cal/(g}\cdot\text{mole}^0\text{K)}$
Reactor Radius, $R = 2.045 \text{ cm}$
Reactor Pressure, $P = 1.67 \text{ kPa}$

Table 2.1: Steady State Simulation Parameters
Reference Jutan (1976)

The maximum temperature rise above the wall temperature for $T_w = 512^0\text{K}$ is shown in Figure 2.2 and the conversion of butane is shown in Figure 2.3.

Define the selectivity of propane as

$$S_p = \frac{\text{change in moles of propane}}{\text{change in moles of butane}} \quad (2.50)$$

$1/S_p$ is plotted in Figure 2.4. We immediately notice the cliff

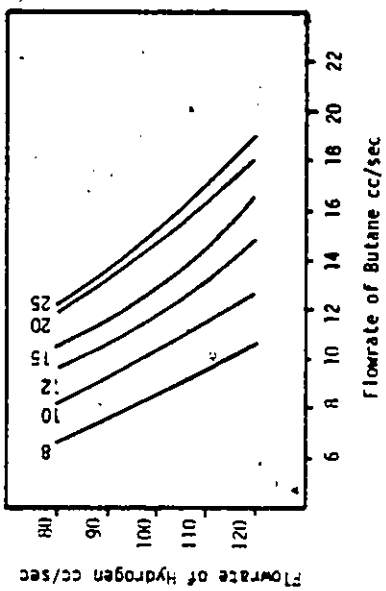


Figure 2.2: Contours of Maximum Temperature Rise

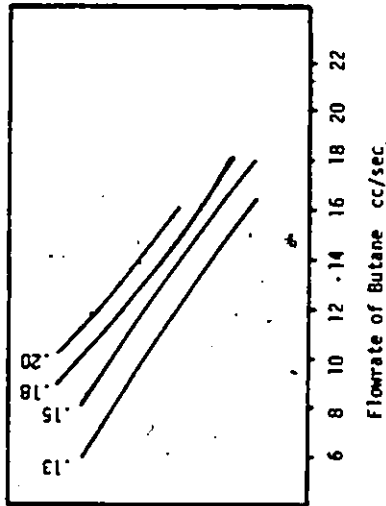


Figure 2.3: Contours of Butane Conversion

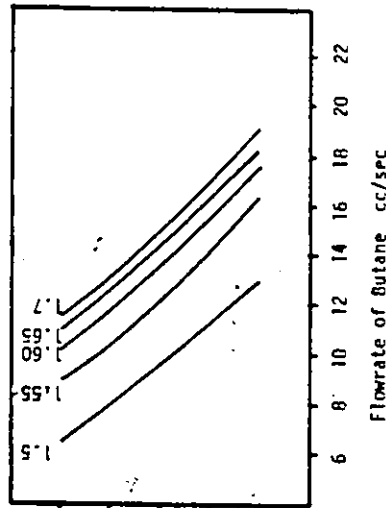


Figure 2.4: Contours of Propane Selectivity

in the upper right hand portion of Figure 2.2. It is in this region, along a contour of nearly constant propane selectivity, Figure 2.4, that the reactor has been traditionally operated. Most control strategies are based on linearized models (Chapter 3), and small changes in the flowrates will certainly change the process description from that of the linearized model. Although the selectivity is not as sensitive to changes in the flowrates, the production rate of propane is obviously affected by these changes.

The effect of increasing the length five fold and decreasing the catalyst activity by 20% is examined as a means of reducing this parameter's sensitivity. The L/D ratio is now nearly comparable to that of a commercial reactor (Hlavacek and Votruba (1977)), where it is typically in excess of thirty. Part of the motivation for such long reactors is that the profile can be moved down the length of the reactor as catalyst deactivates, preventing costly shutdowns. As well, experimental operation of the existing reactor has shown that when the hot spot is near the end of the reactor, large changes in the flow rate sometimes push the temperature profile down and out the end of the reactor. This results in large upsets in production rates and selectivities.

Typical temperature profiles for this new configuration are shown in Figure 2.5. The hot spot temperature is now much closer to the front of the reactor. The maximum rise above the

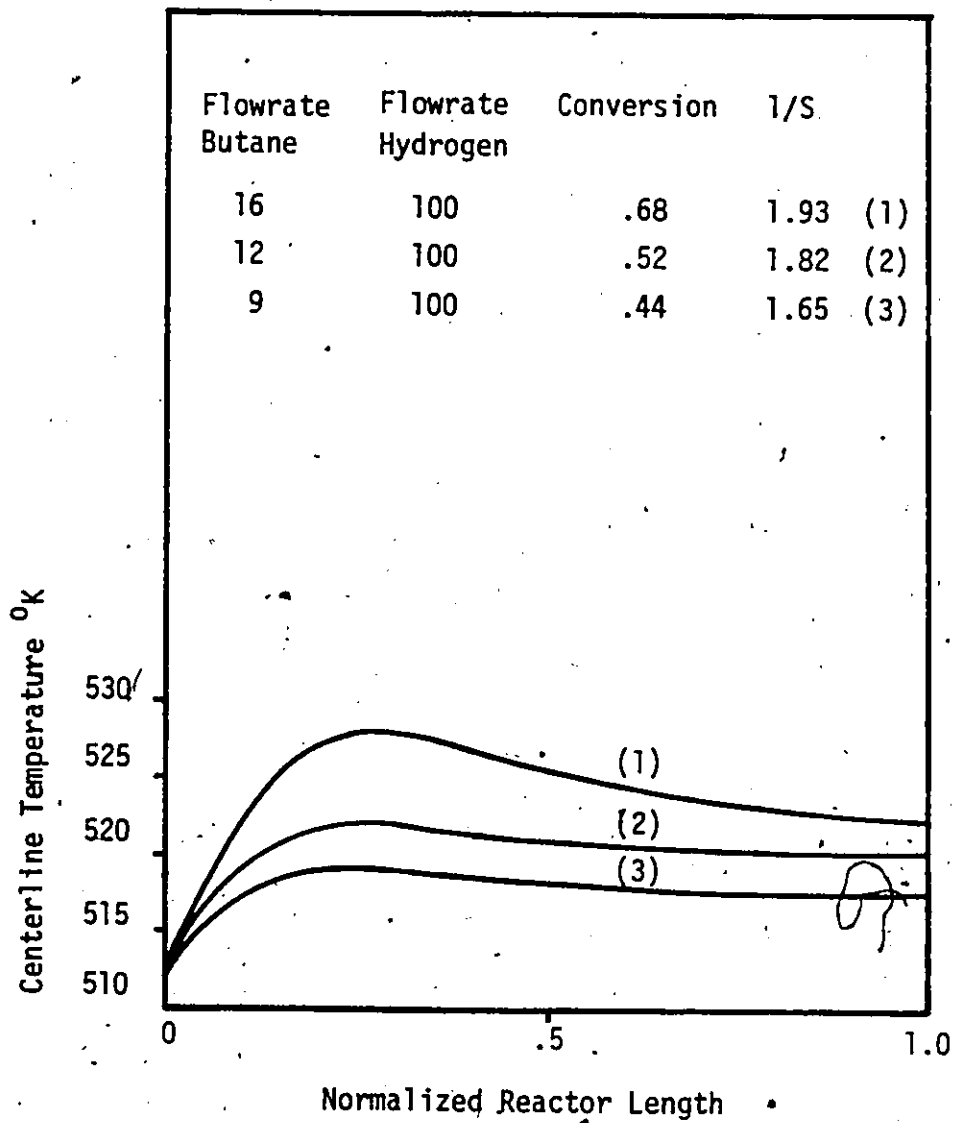


Figure 2.5: Temperature Profiles for Modified Reactor

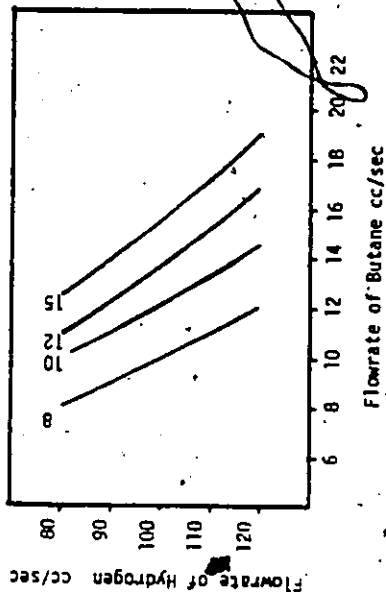


Figure 2.6: Contours of Maximum Temperature Rise: Modified Reactor

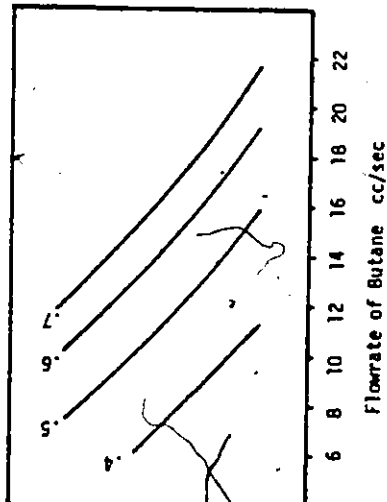


Figure 2.7: Contours of Butane Conversion: Modified Reactor

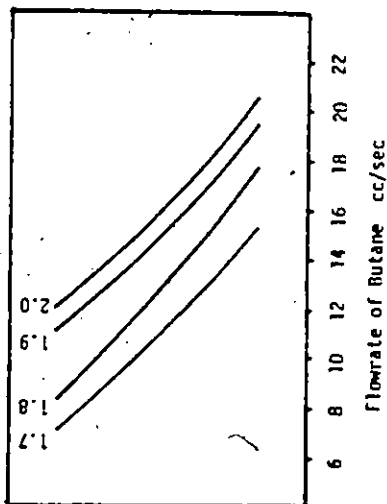


Figure 2.8: Contours of Propane Selectivity: Modified Reactor

wall temperature, conversion of butane and inverse of propane selectivity are shown in Figures 2.6, 2.7 and 2.8 for the new configuration. The cliff observed in Figure 2.2 is still present in Figure 2.6. If we now desired a propane selectivity of 1/1.7, it would not be necessary to operate the reactor as closely to the cliff as in the original configurations, Figures 2.4 and 2.8. In addition, the conversion of butane is much higher in the new configuration.

2.7 Summary

In this chapter, the reaction stoichiometry for the hydrogenolysis of n-butane has been presented. The concepts of independence of reaction were reviewed. This then formed the basis for a study of reaction invariants in tubular reactors. Finally the sensitivity of the steady state temperature profile to changes in flow rate of reactants was examined. The extreme nonlinearity of the hot spot temperature to these changes helps to explain some of the previous difficulties encountered in controlling this reactor.

CHAPTER 3

STATE SPACE MODELLING AND CONTROL OF TUBULAR REACTORS

3.1 Introduction

The mathematics used for estimation and control of linear(ized) finite dimensional processes are presented in the first section of this chapter.. Extensive use of this material will be made in this and subsequent chapters. Recently there have been a number of experimental and theoretical studies on modelling and control of tubular reactors. These papers will be briefly reviewed, and their contributions examined. Finally, a number of modifications to the reactor model studied in this work are considered.

3.2 Review of Linear Quadratic Control Theory

Tubular reactors are highly nonlinear distributed parameter processes. Although control of such processes to minimize some arbitrary objective function is theoretically possible, Sage and White (1977), Pell and Aris (1970), difficulties in computation of these optimal control policies usually precludes their use. Furthermore, these policies are usually not of a feedback nature, which is a requirement for most industrial processes. At this time, the only tractable approach appears to be that of obtaining a linearized, finite dimensional (lumped) approximation to the original process. (Methods of obtaining this will be discussed

in section 3.4.) The approximation to the original process is then described by a linear vector differential equation of the form

$$\dot{\underline{x}}(t) = A_c \underline{x}(t) + G_c \underline{u}(t) \quad (3.1)$$

The n -states $\underline{x}(t)$, are usually temperatures and concentrations at various points within the reactor. These positions are determined by the lumping procedure. There are r manipulated variables, $\underline{u}(t)$, which may be flow rates, cooling oil temperature, etc. A_c and G_c may be functions of time, although this is usually not the case.

In most processes, only a few states or variables related to these states are measured. These m measured variables, or outputs, are denoted by $\underline{y}(t)$. The measurement equation is given by

$$\underline{y}(t) = H \underline{x}(t) \quad (3.2)$$

All variables are deviations from their steady state values.

The emphasis in this work will be on the use of linear quadratic control (LQ) strategies. This is one of the few modern control theories for which a tractable, and appealing solution exists. Here one is interested in finding the control $\underline{u}(t)$, in the interval $0 \leq t \leq T$ to transfer the state from $\underline{x}(0)$ to the origin while minimizing the objective function

$$J_1(\underline{x}, \underline{u}) = \frac{1}{2} \underline{x}'(T) S \underline{x}(T) + \int_0^T (\frac{1}{2} \underline{x}'(t) Q \underline{x}(t) + \frac{1}{2} \underline{u}'(t) R \underline{u}(t)) dt \quad (3.3)$$

where Q, R, S are positive definite matrices. The solution to this problem is well known, Sage and White (1977), Kwakernaak and Sivan (1972), and is given by

$$\underline{u}(t) = -L(t)\underline{x}(t) \quad (3.4)$$

where $L(t)$ satisfies a certain system of equations. It is important to realize that the control problem is formulated as one in which there are no external disturbances affecting the process in the interval $0 < t \leq T$. However, the initial departure of the state from the origin at $t=0, \underline{x}(0)$, can be considered due to an impulse disturbance at this time. If it is desired to design a controller to guard against the effect of disturbances, such as steps, ramps, etc., the mathematical structure of these must be included in the process description. This topic is discussed at length in Chapter 8.

There are a multitude of other techniques that might be used to control tubular reactors, such as pole placement. Of the few experimental and theoretical control studies of tubular reactors, most have used LQ control, Vakil et al (1973), Sorensen (1977), Jutan et al (1977c), Wong (1977). Georgakis et al (1977a,b,c) have proposed pole placement as another technique for tubular reactor control, and this will be discussed in the next section.

Many times the model (3.1) is a linearized approximation to a nonlinear model. The use of the quadratic objective function (3.3)

reduces the influence of second order terms neglected in this linearized approximation, Athans (1971). The ability of the linearized model to represent the true nonlinear process is then better than might normally be expected.

Implementation of LQ strategies requires computer assistance. We, therefore, consider only changing the manipulated variable at equispaced time intervals, $k, k+1, k+2, \dots$. The difference in time between k and $k+1$ is the time interval ΔT , over which the manipulated variables is held constant. The process description at the sampling instants is now given by

$$\underline{x}(k+1) = A\underline{x}(k) + G\underline{u}(k) \quad (3.5)$$

and

$$\underline{y}(k) = H\underline{x}(k) \quad (3.6)$$

The double (A, G) is obtained from (A_c, G_c) by integrating (3.1) over the interval $(t, t+\Delta T)$.

The process description (3.5), (3.6) may not be exact due to disturbances affecting the process. In chemical reactors these may be due to fluctuation in wall temperature, catalyst activity, measurement uncertainty, or modelling errors. If these disturbances are of a random nature, they can be modelled as stochastic processes. The most common representation is

$$\underline{x}(k+1) = A\underline{x}(k) + B\underline{u}(k) + \underline{w}(k) \quad (3.7)$$

and

$$\underline{y}(k) = H\underline{x}(k) + \underline{v}(k) \quad (3.8)$$

where $\underline{w}(k)$ and $\underline{v}(k)$ are process and measurement noise, respectively. They are usually modelled as Gaussian white noise with mean 0 and covariances R_w and R_v . $\underline{v}(k)$ and $\underline{w}(k)$ are usually uncorrelated with each other.

The continuous representation (3.1) and (3.2) need not be considered as limiting representations of (3.7) and (3.8) as $\Delta T \rightarrow 0$. To do so may be meaningless especially when observations are obtained at discrete intervals of time.

We now wish to find the sequence of controls $\underline{u}(k)$, $k=1, 2, \dots, N-1$ to minimize the objective function

$$J = \frac{1}{N} E\{ \underline{x}'(N) Q_0 \underline{x}(N) + \sum_{k=1}^{N-1} (\underline{x}'(k) Q \underline{x}(k) + \underline{u}'(k-1) R u(k-1)) \} \quad (3.9)$$

where E denotes the mathematical expectation and Q and R are positive semidefinite matrices. The solution is readily derived using dynamic programming, Astrom (1970), Sorenson (1968) and is given by

$$\underline{u}(k) = -L(k)\underline{x}(k) \quad (3.10)$$

where the gain matrix $L(k)$ is computed from

$$L(k) = (R + G'S(k+1)G)^{-1} \cdot G'S(k+1)A \quad (3.11)$$

The matrix $S(k)$ satisfies the matrix Riccati equation

$$S(k) = A'S(k+1)(A - GL(k)) + Q \quad (3.12)$$

with initial conditions $S(N) = Q_0$.

In regulatory control, as opposed to trajectory control, one is usually interested in the solution to (3.9) as $N \rightarrow \infty$. In this instance, the control policy is given by

$$\underline{u}(k) = -L_\infty \underline{x}(k) \quad (3.13)$$

where L_∞ is the steady state solution of (3.11) and (3.12). The steady state controller gain may be precomputed as it does not depend on the data. This disjunction is known as the separation principle. A thorough discussion on the properties of the optimal controller, and convergence of the matrix Riccati equation are given by Kwakernaak and Sivan (1972). The controller (3.13) requires exact knowledge of all the process states. Since we very seldom measure all the states and those that we do are never measured perfectly, this strategy is not achievable.

The states may be reconstructed using a Kalman filter, Kalman (1960). The estimate of the state at time $k+1$ based on information up to and including k , $\hat{\underline{x}}(k+1/k)$, is computed from

$$\hat{\underline{x}}(k+1/k) = A\hat{\underline{x}}(k/k) + G\underline{u}(k) \quad (3.14)$$

and the state estimate at time k given information up to time k by the filter

$$\hat{\underline{x}}(k/k) = \hat{\underline{x}}(k/k-1) + K(k)(\underline{y}(k) - H\hat{\underline{x}}(k/k-1)) \quad (3.15)$$

The Kalman gain $K(k)$ satisfies

$$K(k) = P(k/k-1)H'(HP(k/k-1)H' + R_v)^{-1} \quad (3.16)$$

where

$$P(k+1/k) = AP(k/k-1)A' + R_w \quad (3.17)$$

and

$$P(k/k) = P(k/k-1) - K(k)HP(k/k-1) \quad (3.18)$$

$P(k/k-1)$ is the covariance of the state prediction error, that is

$$P(k/k-1) = E\{(\underline{x}(k) - \hat{\underline{x}}(k/k-1))(\underline{x}(k) - \hat{\underline{x}}(k/k-1))'\} \quad (3.19)$$

whereas $P(k/k)$ is the covariance of the filtered estimate

$$P(k/k) = E\{(\underline{x}(k) - \hat{\underline{x}}(k/k))(\underline{x}(k) - \hat{\underline{x}}(k/k))'\} \quad (3.20)$$

The covariances $P(k/k-1)$ and $P(k/k)$ directly reflect the uncertainty in the predicted and filtered state estimates.

The equations (3.14 - 3.18) can be derived from algebraic or statistical arguments, Astrom (1970): $\hat{\underline{x}}(k/k)$ is the conditional mean of $\underline{x}(k)$ given $(\underline{y}(k), \underline{y}(k-1), \dots, \underline{y}(0))$. The conditional mean minimizes the mean square error of the state estimate. It also minimizes the determinant of $P(k/k-1)$, and minimizes the joint confidence region of the state estimates since $\underline{w}(k)$ and $\underline{v}(k)$ have a Gaussian distribution.

The recursive equations are solved in time from $k=1$ onwards, assuming an initial distribution for $\underline{x}(0)$. Conditions for $P(k/k-1)$ to converge to a unique positive definite value $P^\infty(k/k-1)$ are summarized in Kwakernaak and Sivan (1972).

We now seek a control strategy to minimize our objective function (3.9) (as $N \rightarrow \infty$) but only allowing our control to be a function of the information available up to time k (i.e., $y(k), y(k-1), \dots$). The solution is given by Astrom (1970)

$$\underline{u}(k) = -L_\infty \hat{\underline{x}}(k/k) \quad (3.21)$$

where the gain satisfies (3.11) and (3.12), and the simultaneous estimate $\hat{\underline{x}}(k/k)$, is computed from (3.14) and (3.15). This remarkable result, known as the certainty equivalence property, says that the optimal control policy, using the conditional mean for the state estimate, is identical to the optimal control policy using the true value of the state. (The value of the objective function is different though.) The certainty equivalence property applies without qualification to quadratic control of linear systems subject to additive disturbances, Root (1969).

In this section, most of the control theory to be used in this and subsequent chapters has been presented. In the next section, the most pertinent literature on tubular reactor control is reviewed.

3.3 State Space Formulation of Tubular Reactors

To employ the tractable control theory presented in the previous section, one must obtain a linearized, lumped approximation to the partial differential equation describing the process.

Orthogonal collocation has been used successfully to reduce the partial differential equations to an enlarged set of ordinary differential equations in time, Finlayson (1971), Michelsen et al (1973), Sorensen (1977), Jutan et al (1977a), Georgakis et al (1977a). In the previous chapter, we saw how symmetric trial functions were used to approximate radial temperature and concentration gradients. In a similar fashion, nonsymmetric trial functions are used to approximate the axial profiles. If need be, boundary conditions can also be approximated, Villadsen and Stewart (1967), Georgakis et al (1977a). The axial derivatives are then evaluated at a number of interior collocation points. The result is that the nonlinear partial differential equations in time and distance are approximated by a set of nonlinear, ordinary differential equations of the form

$$\dot{\underline{x}} = f(\underline{x}, \underline{u}) \quad (3.22)$$

The state vector \underline{x} , is composed of temperatures and concentrations at the collocation points. The tremendous advantage of using this technique is that low order approximations, typically 5-10 interior collocation points, accurately characterize the dominant

dynamics of the process. This contrasts with the tanks in series approximation which require 1000 cells to achieve comparable accuracy, Finlayson (1971), Michelsen et al (1973); Georgakis et al (1977a).

The nonlinear equations are then linearized by a first order Taylor series expansion about a reference profile, either the steady state solutions of the original nonlinear partial differential equations, Michelsen et al (1977), or a profile constructed from operating records, Jutan et al (1977a). The result is a set of linear, ordinary differential equations of the form

$$\dot{\underline{x}} = \underline{A}\underline{x} + \underline{G}\underline{u} \quad (3.23)$$

The states are deviations from their reference values. For processes with complex kinetics, i.e., other than first or second order kinetic models and multiple reactions, the linearization step is a very tedious and time consuming task. Implicit in this technique is that all of the transport and kinetic parameters are known.

3.4 Control of Tubular Reactors

There have been very few reported applications of modern control theory on tubular reactors. This may be attributed to the problem of formulating a dynamic model of the process that is suitable for control purposes. Michelsen et al (1973), and

Vakil et al (1973) approached the modelling of tubular reactors with the control objectives in mind. The hypothetical process was a packed bed tubular reactor with a liquid phase reaction. Radial gradients and diffusion effects were neglected. In the first paper, a state space formulation of the reactor dynamics was obtained using the technique of orthogonal collocation. A number of transfer functions for this process had previously been developed, Strangeland and Foss (1970), and these were used to examine the effectiveness of the collocation technique in approximating the major dynamic response of the process. It was concluded that the dominant responses of the process were well represented by low order state space models.

In the second paper, Vakil et al (1973), the use of linear quadratic controls and Kalman filtering was evaluated. The control objective was to maintain the effluent temperature and concentration near their steady state values in the presence of inlet disturbances having a known mathematical structure. The control variable was the temperature, or composition of a sidestream, injected at some position along the reactor. The states were reconstructed using a Kalman filter. The only temperature measurement was at the sidestream injection point. It was found that as the injection point moved towards the end of the reactor, more accurate state estimates were obtained. However, unacceptably large manipulations of the control variable were required to minimize the effect of the inlet disturbance in the remaining

length of the reactor. A compromise was thus required. Although not studied, the use of additional temperature measurements would have increased the accuracy of the state estimates.

It was concluded that Kalman filtering and LQ control were an effective means of controlling tubular reactors. However, one must question the need for these techniques since it was shown that excellent control resulted from a control strategy based on a simple transfer function. The contribution of these papers was to outline an approach to the modelling and control of tubular reactors where conventional control schemes may give unsatisfactory performance.

The modelling and control of an empty tube (no catalyst particles) reactor was studied by Georgakis et al (1977a,b,c). The manipulated variable was the temperature of an oil cooling network jacketing the reactor. The effects of axial diffusion were included in the reactor model. Radial gradients were ignored, although the reactor was controlled by adjusting the temperature of the tube wall. By choosing an empty tube reactor, the thermal and concentration wave velocities are identical. However, many of the problems in controlling tubular reactors stem from the difference in thermal and wave velocities, due to the thermal capacity of the packing, Michelsen et al (1973).

In the first paper, the reactor model was described and orthogonal collocation used to obtain a lumped, linearized approximation to the partial differential equation describing

the process. The effect of the particular choice of Jacobi polynomials, and the order of approximation were examined. Again, low order approximations ($N=7$) were found to accurately describe the important dynamic characteristics of the process.

Control of the simulated reactor was studied in the second paper. By a careful choice of the model parameters, the reactor system had a number of steady states. The steady state temperature profile about which the reactor was to be controlled was unstable. Modal control was then used to stabilize this unstable process. However, only one mode (lacking physical interpretation, Michelsen et al (1973)), of the 14 used to approximate the process was unstable. The dynamic characteristics of the controller were also examined.

In the last paper, a filter was used to reconstruct the temperatures and concentrations at the collocation points from a single temperature measurement by using a reduced order observer. The measurement was noise free. As well, there were no disturbances affecting the process. The importance of measurement accuracy and placement is thus not considered.

These papers do not give nearly the insight into the problems of controlling tubular reactors, as do those of Michelsen et al (1973), and Vakil et al (1973). This is primarily due to the absence of packing, radial gradients, process and measurement uncertainties. Rather, these papers are an example of the use of modern control theory on a hypothetical process

which may not reflect conditions encountered in industry.

The dynamics and optimal controls of laboratory tubular reactors have been studied by Sorensen (1976, 1977). The process was a nearly adiabatic tubular reactor in which hydrogen (> 99 mole %) and oxygen reacted on an alumina supported platinum catalyst. The kinetic rate expression included a term to account for changes in catalyst activity due to changing temperature conditions along the reactor. The reactor model accounted for axial dispersion and radial variations in temperature, although not in composition. The radial gradients were very small, arising from imperfect insulation at the tube wall. Orthogonal collocation was used to obtain a lumped/linearized approximation to the original model. A number of parameters were estimated from operating data. The manner in which the parameters were estimated deserves some comment. The thermal and kinetic parameters were estimated separately. The transient behaviour of the reactor was studied when no chemical reaction occurred and the heat transfer parameters determined from this experiment. The kinetic parameters were then determined from operating data when the reaction occurred. The difficulties encountered by Jutan et al (1977b) in simultaneously estimating the heat and mass transfer parameters might have been alleviated by this approach.

To collect the data, the reactor was allowed to reach steady state by excluding the flow of oxygen. A constant flow of

this reactant was then added, and the process allowed to achieve a new steady state. Data collected during this transient was used to estimate a number of transport parameters. Since the process was open loop stable, a feedback controller was not required to stabilize the reactor during the data collection. Problems encountered in identifying parameters under feedback control were thus avoided.

In the second paper, Sorenson (1977), the use of Kalman filtering and LQ control was evaluated. The controller and filter were designed to compensate for inlet disturbances, modelled as the output of a linear stochastic difference equation of the form

$$\underline{N}(t) = \phi_s \underline{N}(t) + \underline{a}(t) \quad (3.24)$$

As the process had no inlet disturbances, the performance of the controller was examined by subjecting the inlet to step disturbances in temperature and composition. The ability of the Kalman filter to estimate the disturbances and minimize their effect on the temperature and composition at the process output was examined. As expected, as the elements of the shaping matrix ϕ_s in (3.24) increased, the offset in the filter and controlled variables were reduced. This occurred at the expense of a more oscillatory response. The particular problems of estimating and controlling deterministic disturbances in the presence of stochastic disturbances is discussed in Chapter 8 of this thesis.

It was concluded that the use of Kalman filtering and linear quadratic control was an excellent method of controlling their reactor. Prior to undertaking the time consuming task of modelling a tubular reactor for control purposes, one must be sure that there is a requirement for this. Comparison with a more conventional controller, i.e. a proportional integral controller, would verify this need if it existed.

The most recent experimental study using many of the techniques was by Jutan et al (1977a,b,c). The reactions and reactor have already been discussed in chapter 2. Difficulties in controlling the reactor by conventional means were well demonstrated and therefore considerable work was put into modelling its dynamics. The dynamic model of the process was obtained by again using the technique of orthogonal collocation. Several mass and heat transfer parameters, and a stochastic model of the process disturbances were identified from dynamic data. The reactor model was expressed in discrete form (at 60 second time intervals), since temperature data was only available at discrete time intervals. The final result was a model of the form

$$\underline{x}(k+1) = A\underline{x}(k) + G\underline{u}(k) \quad (3.25)$$

and

$$\underline{y}(k) = \underline{C}\underline{x}(k) + \underline{N}(k) \quad (3.26)$$

The states were temperatures at the collocation points, and the controls were the flow rates of butane and hydrogen. $\underline{N}(k)$ represents the total effect on the process outputs of all disturbances not accounted for by the reactor model. $\underline{N}(k)$ was modelled as a multi-variate autoregressive process.

$$\underline{N}(k) = \phi \underline{N}(k-1) + \underline{a}(k) \quad (3.27)$$

where $\underline{a}(k)$ is a white noise Gaussian process. The radially averaged concentrations at the outlet were obtained from

$$\underline{c}_{av}(k) = \underline{C} \underline{x}(k) + \underline{D} u(k-1) \quad (3.28)$$

There is no time derivative in this concentration equation because a pseudo steady state assumption was made, Jutan et al (1977a).

Linear quadratic control with Kalman filtering was experimentally demonstrated to be an excellent means of controlling this process when compared to more conventional techniques.

However, this improvement in control must be reconciled against the tremendous time and effort required to obtain a dynamic model of the process. The complexity of the modelling procedure was increased very much by the inclusion of radial gradients in the reactor model. By comparison, the inclusion of axial dispersion effects is not a major task. A less rigorous and time consuming approach would have been to ignore radial gradients, and to evaluate the control from this simplified model and only proceed with the more detailed model if there was a need.

This concludes the literature review on reactor control via dynamic modelling. As has been seen, orthogonal collocation has been used successfully in modelling packed bed reactors. Several experimental applications have successfully demonstrated the use of Kalman filtering and linear quadratic control. In the next section, a number of extensions to the work of Jutan (1976) are examined.

3.5 Development of the State Space Model in Ratio and Total Flow Variables

In the modelling stage, the nonlinear model is linearized about an operating temperature profile, and the average flow rates. Rather than linearize about the average flow rates, it may be desirable to linearize about some function of these, say total flow and the ratio of the flow rates. One method of doing this is to express the model in terms of total flow and ratio, and then linearize with respect to these new variables. A much easier approach is to use the chain rule for derivatives. Denoting the flow rates of butane and hydrogen by u_1 and u_2 then the total flow u_T and flow ratio u_R are given by

$$u_T = u_1 + u_2 \quad (3.29)$$

and

$$u_R = \frac{u_2}{u_1} \quad (3.30)$$

where u_1 and u_2 are the butane and hydrogen flows. Using the chain rule, we find for a function $f(u_1, u_2)$ that

$$\begin{bmatrix} \frac{\partial f}{\partial u_T} \\ \frac{\partial f}{\partial u_R} \end{bmatrix} = \begin{bmatrix} \frac{\partial u_T}{\partial u_1} & \frac{\partial u_R}{\partial u_1} \\ \frac{\partial u_T}{\partial u_2} & \frac{\partial u_R}{\partial u_2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f}{\partial u_1} \\ \frac{\partial f}{\partial u_2} \end{bmatrix} \quad (3.31)$$

This is of course only defined if $u_1 \neq 0$.

If we have linearized our model about some steady value \bar{u}_1, \bar{u}_2 , we can easily obtain the model linearized about \bar{u}_R, \bar{u}_T from (3.31). Examination of the algebra used to obtain the discrete representation (3.25 and 3.28) ~~of~~ ^{from} Tan et al (1977b), indicates that one simply replace $\underline{G}u(k)$ and $\underline{D}u(k-1)$ by $\underline{G}\underline{J}u^*(k)$ and $\underline{D}\underline{J}u^*(k-1)$. \underline{J} is the matrix of partial derivatives in (3.31) evaluated at \bar{u}_1, \bar{u}_2 , and $\underline{u}^*(k)$ is the vector of controls $(u_R - \bar{u}_R, u_T - \bar{u}_T)'$. This simple technique allows us to easily linearize our nonlinear model in variables we might not otherwise have considered.

3.6 Wall Temperature Effects

Experimental operation of the reactor has shown that the wall temperature is an important operating variable. It has been used to compensate for changes in catalyst activity resulting from shutdown of the reactor. In several experimental

studies, Jutan et al (1977c), Wong (1977), the wall temperature has also been used to test the ability of their control algorithms to compensate for major load upsets. The reactor model developed by Jutan et al (1977a) did not include explicitly the wall temperature, as it was assumed to be constant. The wall temperature effects are obtained by linearizing the nonlinear mass and energy balances at the interior collocation points. The continuous equations are then integrated over the sixty second control interval to obtain a discrete representation. The final result is that one obtains a model of the form

$$\underline{x}(k+1) = A\underline{x}(k) + G\underline{u}(k) + Fz(k) \quad (3.32)$$

and

$$\underline{c}_{av}(k) = C\underline{x}(k) + D\underline{u}(k-1) + E z(k-1) \quad (3.33)$$

$z(k)$ is the deviation of the wall temperature from its steady state value. The vectors (E,F) are presented in Appendix 3. We have assumed, as did Jutan et al (1977a), that the heating oil and tube wall are at the same temperature. Models of the form of (3.32) and (3.33) will be used in Chapter 8 where the design of controllers to compensate for load changes will be studied.

3.7 Estimation of R_w

The development of state space models represents a considerable investment in time. This effort should not be compromised.

by the use of poor statistical techniques to estimate any model parameters. To obtain accurate estimates of the transport parameters, it is also necessary to model the process disturbances. Since the effect of process disturbances can only be observed at the measured outputs, it is appropriate then to use time series models to represent these disturbances. We may, however, obtain more insight into the nature of the disturbances affecting the process if we can identify two contributions, measurement and state noise. We seek then a state representation of the form (3.7) and (3.8) where $(\underline{v}(k), \underline{w}(k))$ are Gaussian white noise processes with covariances R_V and R_W .

Several techniques have been proposed to estimate R_V and R_W , Mehra (1970), Carew and Belanger (1973), Sinha and Tom (1978). The essence of these techniques is that R_V and R_W can be expressed as linear functions of the autocovariance matrices

$$\Gamma_k \equiv E((\underline{y}(k) - H\hat{\underline{x}}^S(k-1))(\underline{y}(k) - H\hat{\underline{x}}^S(k/k-1))) \quad (3.34)$$

where $\hat{\underline{x}}^S(k/k-1)$ is suboptimal estimate of the state vector based on erroneous estimates for R_V and R_W . Considerable manipulation of the autocovariance matrices and state equations is required to obtain this linear dependence. In the original paper, Mehra (1970) claims that R_V and R_W can be uniquely estimated from the observations of \underline{y} alone. However, unique values for R_V and R_W cannot be obtained from an analysis of \underline{y} alone. Rather, a unique value of

the steady state Kalman gain can be estimated, Alspach (1972). These techniques could still be used to estimate R_w if R_v were known. A number of shortcomings of these methods have been outlined by Neethling and Young (1974). As they point out, since the sampled autocovariance matrices are themselves autocorrelated, observed autocorrelations in the sampled estimates cannot be attributed solely to an improper choice for R_v or R_w .

In addition to these drawbacks, there are computation objections to this method. For a seven dimensional state process, where all the states are measured, the method of Mehra requires computation of the generalized inverse of a matrix of dimension 49×7 . Using a modification suggested by Neethling and Young (1974), this matrix would be replaced by a matrix of dimension 98×7 .

For these reasons, an alternative approach was taken. Instead of assuming a general structure for R_w , in which case 28 parameters would have to be identified, the following two parameter structure was proposed.

$$R_w = R_w(\sigma^2, \rho) = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \dots \\ & 1 & \rho & \dots & \dots \\ & & 1 & \dots & \dots \\ \text{sym} & & & & \\ & & & & 1 \end{bmatrix} \quad (3.35)$$

This structure will allow for a simple correlation structure among the process disturbances. If the states were of mixed units

the common variance term would have to be replaced by individual variances in order to retain the same correlation structure. It was known from experimental observations that the temperature measurements are uncorrelated, with common variance $\sigma^2 = 4 \cdot C^{02}$, Jutan (1976).

The parameters (σ^2, ρ) were estimated by minimizing the generalized prediction error variance (or the volume of the joint confidence region of the prediction errors).

$$J = \left| \sum_{k=1}^{\text{NOB}} (\underline{y}(k) - H\underline{\hat{x}}(k/k-1))(\underline{y}(k) - H\underline{\hat{x}}(k/k-1))' \right| \quad (3.36)$$

In (3.36), $\underline{\hat{x}}(k/k-1)$ is the Kalman filter estimate for $\underline{x}(k)$ evaluated with $R_w = R_w(\hat{\sigma}^2, \hat{\rho})$. The parameter estimates $(\hat{\sigma}, \hat{\rho})$ that minimized (3.35) were found to be $\hat{\sigma}^2 = 3.0$ and $\hat{\rho} = .1$.

The low value of the correlation coefficient implies that the disturbances in the reactor are of a local nature.

3.8 Summary

In this chapter, most of the control theory that will be used in subsequent chapters has been presented. The more recent applications of modern control theory to tubular reactors have been reviewed. Orthogonal collocation has been used successfully to obtain a lumped approximation to partial differential equations which describe the mass and energy balances. Once a linearized, finite dimensional model is obtained in this manner,

many control theories can be applied. The estimation of mass and heat transfer parameters from operating data is also straightforward when the process model is expressed in discrete form. In the last part of this chapter, a number of extensions to Jutan's work were considered. These results will be used in subsequent chapters.

CHAPTER 4

OPTIMAL SENSOR LOCATION IN TUBULAR REACTORS

4.1 Introduction

In most processes it is not feasible to measure all the process variables of interest. This may be due to economic constraints, lack of on-line detectors, or the hostile environments in which the sensors are located. Therefore, only a certain subset of these variables, or variables related to these process or state variables are measured. If all the states are to be inferred from such measurements, the precision of the state estimates will depend on which subset of the states, or auxiliary variables are measured, and upon the precision of the measurements. Assuming that one has chosen the type of measurements and the measurements to be used, thereby fixing their precision, then it would be desirable to select the location of these sensors in some optimal manner.

If the purpose is to obtain the best estimate of the state vector, then this problem is just the dual of that which arises in the optimal design of regression experiments. In that problem one is interested in choosing the settings of the independent variables in a number of experiments to maximize the amount of information that will be obtained on the parameters in the regression equation. The statistical literature contains an abundance of theory on this subject and this provides a basis for

the sensor location problem. Some of the theoretical aspects of these techniques and their potential applications to sensor location problems have been examined by Mehra (1976).

Previous studies on sensor locations in tubular reactors, Colantuoni and Padmanabhan (1977), and Kumar and Seinfeld (1978b), looked at the state estimation problem, using a chosen optimality criterion, by retaining the continuous time models for simple reaction systems. In this study, we review a number of different design criteria, and apply some of them to the discrete time model of a highly exothermic packed bed tubular reactor carrying out the multiple reactions of the hydrogenolysis of n-butane.

Consider a process which can be described by the discrete linear (linearized) state variable equation

$$\underline{x}(k+1) = A\underline{x}(k) + G\underline{u}(k) + \underline{w}(k) \quad (4.1)$$

where $\underline{x}(k)$ is an $(n \times 1)$ vector of state variables, $\underline{u}(k)$ is an $(r \times 1)$ vector of manipulated inputs, and $\underline{w}(k)$ is a vector of Gaussian white noise disturbances with covariance matrix, R_w .

The process variables that are measured, $\underline{y}(k)$, an $(m \times 1)$ vector, are related to the state vector $\underline{x}(k)$ by the measurement equation

$$\underline{y}(k) = H\underline{x}(k) + \underline{v}(k) \quad (4.2)$$

where $\underline{v}(k)$ is an $(m \times 1)$ vector of measurement noise with covariance matrix, R_v . Provided that the system (4.1) and (4.2) is observable then the entire state vector $\underline{x}(k)$ can be estimated from

the measurements of $\underline{y}(k)$ by the use of a filter equation. Obviously, the precision of the state will depend upon the precision of the measurements, R_V , as well as upon the location of these measurements, that is, upon H . Assuming that one has chosen the number and type of measurements to be used (R_V is therefore fixed), then it may be desirable to select, in some optimal manner which variables should be measured (that is, to specify H).

4.2 Theory on the Optimal Location of Sensors

The placement of sensors must obviously depend upon one's optimality criterion, and there have been a number of different approaches in the literature.

4.2.1 Optimizing an Observability Index

One of the earliest approaches to the sensor location problem was suggested by Johnson (1969) for the nonstochastic state reconstruction problem ($\underline{w}(k) = \underline{0}$, $\underline{y}(k) = \underline{0}$). He suggested that as a measure of the observability of the system, one use the determinant of the generalized observability matrix,

$$W = \sum_{j=0}^{n-1} (HA^j)' (HA^j) \quad (4.3)$$

Johnson (1969) and Muller and Weber (1972) then suggested maximizing $|W|$ through a choice of the sensor locations in order to improve the degree of complete observability.

4.2.2 Obtaining "Best" State Estimates

When stochastic disturbances are present, unmeasured states can be estimated with the Kalman Filter.

$$\hat{\underline{x}}(k+1/k) = A\underline{\hat{x}}(k/k) + G\underline{u}(k) \quad (4.4)$$

$$\hat{\underline{x}}(k/k) = \hat{\underline{x}}(k/k-1) + K(k)(\underline{y}(k) - H\underline{\hat{x}}(k/k-1)) \quad (4.5)$$

$$K(k) = P(k/k-1)H'(HP(k/k-1)H' + R_V)^{-1} \quad (4.6)$$

$$P(k+1/k) = AP(k/k)A' + R_W \quad (4.7)$$

$$P(k/k) = P(k/k-1) - K(k)HP(k/k-1) \quad (4.8)$$

with $P(0/0) \equiv P_0$.

$P(k/k-1)$ is the covariance matrix of the state prediction error ($\underline{x}(k) - \hat{\underline{x}}(k/k-1)$) and $P(k/k)$ is the covariance matrix of the filter error ($\underline{x}(k) - \hat{\underline{x}}(k/k)$). These covariance matrices provide a direct measure of the precision of the state estimates ($\hat{\underline{x}}(k/k-1)$, $\hat{\underline{x}}(k/k)$) and their dependence upon the sensor location matrix H is given explicitly in equations (4.6), (4.7) and (4.8). Since these matrices are independent of the observations, ($\underline{y}(k)$ $k = 1, 2, \dots, N$) they can be evaluated for $k = 1, 2, \dots, N$ and examined for any given choice of the sensor locations. However, the number of elements in these covariance matrices makes it difficult to assess the precision of the state estimates in each case, and therefore a single scalar function of one of these is usually chosen.

One such overall criterion is to minimize, by the choice of H , the function

$$v_1 = \frac{1}{N} \sum_{k=1}^N \text{trace } P(k/k-1) \quad (4.9)$$

This is an extension of the usual A-optimality criterion, in the statistical literature in which the average variance of the state estimates over the first N observation periods is minimized. This criterion was suggested by Mier (1967). Athans (1972) used this criterion to determine the selection of one measurement from a number of possible measurements. Herring and Melsa (1974) extended this to the optimal selection of a set of measurements. Both methods require finding switching functions to determine the best time at which to change to a different sensor or set of sensors. One drawback of this criterion is that it is not invariant with respect to linear transformations and its minimum will therefore depend on the units in which the states are expressed. This becomes a problem when the states are of mixed units, say temperatures and concentrations.

Another overall criterion is to minimize, by choice of H, the quantity

$$v_2 = \frac{1}{N} \sum_{k=1}^N |P(k/k-1)| \quad (4.10)$$

This is an extension of the commonly used D-optimality criterion in the statistical literature. It can be shown that, by minimizing this, one is minimizing the average volume of the joint confidence region of the states over the first N steps. Since the determinant

is invariant for linear transformations, the minimum of this criterion is independent of the units employed for the various states.

In many instances, one is interested in the steady-state solutions to the Kalman filter equations (4.7) and (4.8). These equations will converge to unique positive definite solutions $P^\infty(k/k-1)$ and $P^\infty(k/k)$ if the pair (A,H) is observable (A,R_w) is controllable and $|P| > 0$, Kushner (1971). The A-optimality criterion for sensor location then becomes that of minimizing

$$v_3 = \text{trace } P^\infty(k/k+1) \quad (4.11)$$

and D-optimality involves minimizing

$$v_4 = |P^\infty(k/k-1)| \quad (4.12)$$

Although we have used $P(k/k-1)$ throughout in these criteria one could use the covariance matrix $P(k/k)$ of the filtered estimate if that was of most interest. Furthermore, if only a subset of the state vector (say the first n_1 elements) is of interest, then the criteria (4.11) and (4.12) can be applied to the upper left $(n_1 \times n_1)$ submatrix of $P^\infty(k/k-1)$.

4.2.3 Detection of Load Disturbances

Attention has been focussed on the placement of sensors to detect load disturbances acting on the system. Jorgensen and Clement (1977) considered the placement of temperature and concentration sensors in a tubular reactor to best detect inlet flow,

temperature and concentration disturbances as well as estimating the states of the original system. These inlet disturbances were assumed to be modelled by the autoregressive model

$$\underline{x}^*(k+1) = \phi \underline{x}^*(k) + \underline{a}(k) \quad (4.13)$$

These authors used a simulation approach to evaluate the reactor concentration and temperature profile responses to such load disturbances by considering one disturbance at a time. A qualitative analysis of these results led them to their choice of sensor locations. When many load disturbances are simultaneously operating on the system such an intuitive approach may be difficult. However, the optimality criteria (4.11) and (4.12) can easily be applied to this situation. Defining the augmented state model

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{x}^*(k+1) \end{bmatrix} = \begin{bmatrix} A & F \\ 0 & \phi \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{x}^*(k) \end{bmatrix} + \begin{bmatrix} G \\ 0 \end{bmatrix} \cdot \underline{u}(k) + \begin{bmatrix} \underline{w}(k) \\ \underline{a}(k) \end{bmatrix} \quad (4.14)$$

the D-optimal design would be to choose the sensor locations to minimize the determinant of the augmented covariance matrix

$$\begin{bmatrix} P_{11}^{\infty}(k/k-1) & P_{12}^{\infty}(k/k-1) \\ P_{21}^{\infty}(k/k-1) & P_{22}^{\infty}(k/k-1) \end{bmatrix}$$

If only the detection of load disturbances was of interest one would minimize $|P_{22}^{\infty}(k/k-1)|$.

4.2.4 Location for Optimal Control

Since control of a process is often the objective behind the installation of sensors, it is logical that one try to improve the control as much as possible by selecting the location of the sensors. Mier (1967), and Mellefont and Sargent (1977, 1978) considered such an approach for the case of linear quadratic control. The control strategy minimizing the objective function

$$v_5 = E\{\underline{x}'(N)Q_1\underline{x}(N) + \sum_{k=1}^{N-1} \underline{x}'(k)Q_1\underline{x}(k) + \underline{u}'(k-1)Q_2\underline{u}(k-1)\} \quad (4.15)$$

is given by

$$\underline{u}(k) = -L(k)\hat{\underline{x}}(k/k) \quad (4.16)$$

where $L(k)$ satisfies the equation

$$L(k) = (Q_2 + G'S(k+1)G)^{-1} G'S(k+1)A \quad (4.17)$$

$$S(k) = A'S(k+1)A + Q_1 - A'S(k+1)L(k) \quad (4.18)$$

with initial condition $S(1) = Q_1$, Astrom (1970). If this control law is substituted into the objective function, one obtains

$$v_5 = \underline{m}'S(1)\underline{m} + \text{trace}(S(1)R_0) + \sum_{k=1}^{N-1} \text{trace}(S(k+1)R_w) \\ + \sum_{k=1}^{N-1} \text{trace}(P(k/k-1)L'(k)(B'S(k+1)B + Q_2)L(k)) \quad (4.19)$$

where R_0 is the initial covariance of the state, and \underline{m} is the initial state estimate. The only contribution to the objective function v_5 that arises from the state estimator is through the last term in (4.19). Mellefont and Sargent (1977) found switching functions which determined how sensor location should be changed to minimize (4.19). However, such an analysis assumes that the matrices Q_1 and Q_2 in the quadratic cost functions (4.14) would be the same for all sensor arrangements and are known in advance. This usually is not the case. In fact, they may be considered as design parameters which are to be varied until the variances of \underline{x} and \underline{u} are jointly acceptable.

In regulatory control where one is usually interested in the steady-state solution L of (4.17), the following approach could be used to evaluate the effect of sensor locations on L.Q. control. The covariance matrices of the state and control variable vectors under optimal steady-state L.Q. control are given by, MacGregor (1973)

$$\begin{aligned} \text{Var}(\underline{x}) = V_x = & (A - GL_\infty)V_x (A - GL_\infty)' + (A - GL_\infty)P^\infty(k/k)L_\infty'G' \\ & + GL_\infty P^\infty(k/k)(A - GL_\infty)' + GL_\infty P^\infty(k/k)L_\infty'G' + R_w \end{aligned} \quad (4.20)$$

$$\text{Var}(\underline{u}) = V_u = L_\infty(V_x - P^\infty(k/k))L_\infty' \quad (4.21)$$

For a given choice of Q_1 in the objective function and any choice of the sensor location, one can evaluate the trace or the

determinant of the state covariance matrix (4.20). Q_2 is then used as a design matrix to find the control which yields acceptable variances for the manipulated variables, u .

Thus, for any choice of the sensor locations, one can evaluate the performance of the L.Q. controller as indicated by the elements of the covariance matrix of the state vector, x . This provides a means of either selecting sensor locations directly or of evaluating the performance of the controller for sensor locations chosen by other means. This is applied to the catalytic reactor in the present work.

Note that since the optimal control (4.16) is a linear function of the state estimate, $\hat{x}(k/k)$, the D-optimal design procedure discussed earlier should provide good control since it selects sensor locations which provide the "best" estimate of the entire state vector. However, since only a reduced m -dimensional subspace of the state estimates ($L(k)\hat{x}(k/k)$) is used in control, one could use the D-optimality criterion directly on this subspace, that is minimize, by choice of the sensor locations, the determinant

$$v_6 = |L_{\infty} P^{\infty}(k/k) L_{\infty}'| \quad (4.22)$$

This would directly provide the best sensor locations for evaluating the L.Q. control action. However, again since this subspace determined by L_{∞} depends on the choice of Q_1 and Q_2 in the

performance index, this probably should involve some iteration on Q_2 (as discussed above) in order to make comparisons under conditions for the same $\text{Var}(\underline{u})$.

4.3 Distributed Parameter Systems and Tubular Reactors

Sensor location for linear distributed parameter systems has been studied by Yu and Seinfeld (1973), Kumar and Seinfeld (1978a), Aidarous et al (1975), and Omatu et al (1978). No attention is directed towards the use of the estimated states for control purposes. Due to the nonlinear nature of the distributed parameter Riccati equation a function of the upper bound on the error covariance matrix is minimized. Most of these methods require knowledge of the Green's function for the system and approximate the solution of the differential equation by some orthogonal function. Tubular reactors are usually described by coupled non-linear partial differential equations in several spatial co-ordinates if significant temperature and concentration gradients exist. The above methods are difficult to apply to such processes.

Sensor allocation in tubular reactor systems has been studied by Kumar and Seinfeld (1978b), Jorgensen and Clement (1977), and Colantuoni and Padmanabhan (1977) for a single reaction. In the first two cases, orthogonal collocation was used to obtain a lumped approximation to the original distributed parameter system. The process models were not discretized. Measurements of concentrations and temperatures were assumed to be available

continuously. Kumar and Seinfeld (1978b) consider placement of concentration and/or temperature sensors by minimizing the continuous equivalent of criterion v_1 using an algorithm suggested by Athans (1972). Their conclusion as to the optimal location of the temperature sensors using this criterion are seen to correspond to that obtained in this chapter for the infinite time estimation problem. However, their comment that for non-optimally located sensors, the trace $P(k/k)$ is in general not a monotonically decreasing function of time deserves some qualifications. If the system (A,H) is observable and (A,R_w) controllable, then the trace $P(k/k)$ indeed may not converge monotonically to its steady state value. Using an expression developed in Bucy and Joseph (1968), page 77, it can be shown that $P(k/k)$ is monotone non-increasing if $|P_0 - P^\infty(k/k)| > 0$. When P_0 is chosen as cI , this condition is satisfied if c is greater than the largest eigenvalue of $P^\infty(k/k)$. Depending on the initial choice of the covariance matrix P_0 it could be possible for $P(k/k)$ to converge in a non-monotonic fashion to its steady state value for any choice of sensor locations.

Colantuoni and Padmanabhan (1977) retained the partial differential equations (linearized) in time and axial distance, and minimized the integral trace of the prediction error covariance matrix over the spatial domain and over the time elapsing between discrete measurements. However, their procedure is very

complicated, and it is difficult to imagine employing it for the case of multiple reactions or when radial gradients exist.

The present work considers the placement of temperature sensors in a catalytic packed bed tubular reactor involving the multiple reactions of the hydrogenolysis of n-butane. The criterion used here for sensor location is that of D-optimality on the steady-state filter equations (4.12). A simple method is used to check the sensitivity of the criterion to the order of the lumped model approximation. The effect of the sensor locations on the performance of LQ control strategies is examined.

4.4 Description of the Tubular Reactor System and Model

As we recall, the temperature dynamics were represented by the state equations:

$$\underline{x}(k+1) = A\underline{x}(k) + G\underline{u}(k) \quad (4.23)$$

The states, $\underline{x}(k)$, represent the temperatures at the interior collocation points used in obtaining the lumped approximation to the partial differential equations. Jutan et al (1977a) found that seven interior collocation points adequately described the axial temperature profile. The manipulated input $\underline{u}(k)$ comprises the hydrogen and butane feedrate to the reactor. An experimental operating profile is shown in Figure 4.1. The concentrations at the exit are given by an algebraic relationship (resulting from the quasi-steady state assumption)

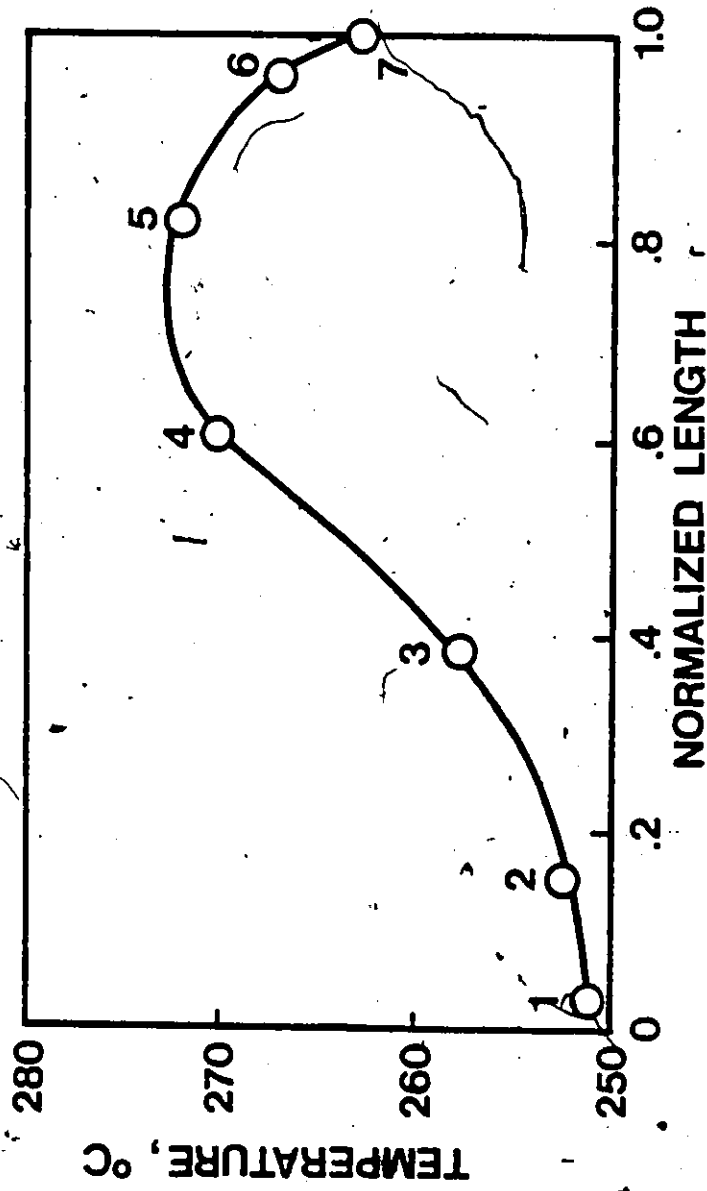


Figure 4.1: Typical Temperature Profile
Circles denote collocation points
by number Reference Jutan (1976)

$$\underline{c}(k) = \underline{C}\underline{x}(k) + \underline{D}u(k-1) \quad (4.24)$$

The model (4.23) does not exactly describe the dynamic behaviour of the reactor. Disturbances due in part to fluctuations in flow rates, catalyst activity and packing, wall temperature and local hot spots, were modelled as additive Gaussian white noise with mean zero and covariance,

$$\underline{x}(k+1) = \underline{A}\underline{x}(k) + \underline{B}u(k) + \underline{w}(k) \quad (4.25)$$

Rather than try to identify a full matrix R_w from operating data, the following two parameter covariance structure was used.

$$R_w(\sigma^2, \rho) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \dots & \dots \\ & 1 & \rho & \rho^2 & \dots & \dots & \dots \\ & & 1 & \rho & \dots & \dots & \dots \\ & & & 1 & \rho & \dots & \dots \\ & & & & 1 & \dots & \dots \\ & & & & & 1 & \dots \\ & & & & & & 1 \end{bmatrix} \quad (4.26)$$

Sym

The values $\sigma^2 = 3$ and $\rho = 0.1$ gave the best fit.

Temperatures were measured by thermocouples at nine points along the reactor central axis. Since these positions did not exactly coincide with the collocation points, quadratic interpolation was used to estimate the temperature at the seven collocation points. The measurement equation was

$$\underline{y}(k) = H\underline{x}(k) + \underline{v}(k) \quad (4.27)$$

where $H = I$ implies that the full state vector was measured. The measurement error vector was a mean zero Gaussian process with covariance matrix $R_V = \sigma^2 I$, $\sigma = 2^\circ\text{C}$. The variance σ^2 was independent of the frequency at which the measurements were taken.

4.5 The Optimal Location of Reactor Thermocouples

Although our ultimate objective for reactor control studies was to maintain the exit concentrations at specified levels some practical problems precluded this. The effluent chromatographic analysis was available only infrequently (every 6 minutes) which is inadequate for control of this reactor. Therefore, control decisions were based upon temperatures measured along the axis of the reactor. The interesting problems in looking at optimal thermocouple locations were: (i) to see which locations provided the greatest amount of information on the temperature profile, and (ii) to see how much the optimal stochastic control scheme studied by Jutan et al (1977b) could be expected to degrade if only one or two thermocouples had been used instead of the full set as in equation (4.27).

Therefore, consider the measurement equation to be of the form

$$\underline{y}(k) = H\underline{x}(k) + \underline{v}(k) \quad (4.28)$$

where each of the m rows of H consists only of one non-zero element

whose value is 1.0. To determine the sensor locations which best estimate the temperatures given by equation (4.23), we use the D-optimality criterion, and minimize with respect to the choice of the thermocouple locations, the determinant, v_4 . When there are a finite number of time invariant sensor locations, the minimization of v_4 is straightforward. One method is to simply evaluate v_4 for all seven possible combinations of locating m sensors. Alternatively a less time consuming approach is to proceed sequentially and to choose the second sensor to minimize the objective function conditional upon the location of the first sensor and so on. This latter approach does not insure that the sensors' locations will be optimal as in the simultaneous approach, although this approximation may often be true in practice. When the number of possible locations is large, one should resort to using any of a number of efficient algorithms that exist in the statistical literature for adding and dropping locations in order to arrive at a globally D-optimum design. (See for example Wynn (1970) and Mitchell (1974)).

Excluding, for the moment, the possibility of having multiple, independent measurements at the same position, the locations of sensors were picked sequentially using the D-optimality criterion (4.12) for three ranges of the covariance matrices (R_w, R_v). The results of this optimal selection are shown in Table 4.1. For the first thermocouple we see that it appears in location 5 for all cases. The second thermocouple,

Sequential Thermocouple Location Conditional on Preceding Additions Selection

Number of Sensors	$R_V = 10I$	$R_V = 4I$	$R_V = I$
	$R_W = I$	$R_W = \sigma^2(\rho)$	$R_W = 10I$
	(Equation 4.26)		
1	5	5	5
2	4	4	4
3	6	6	6
4	3	3	3
5	2	2	2
6	1	7	7
7		1	1

Table 4.1: Sequential Sensor Location (Refer to Figure 4.1 for Thermocouple Positions)

Sequential Thermocouple Location Conditional on Preceding Additions Selection

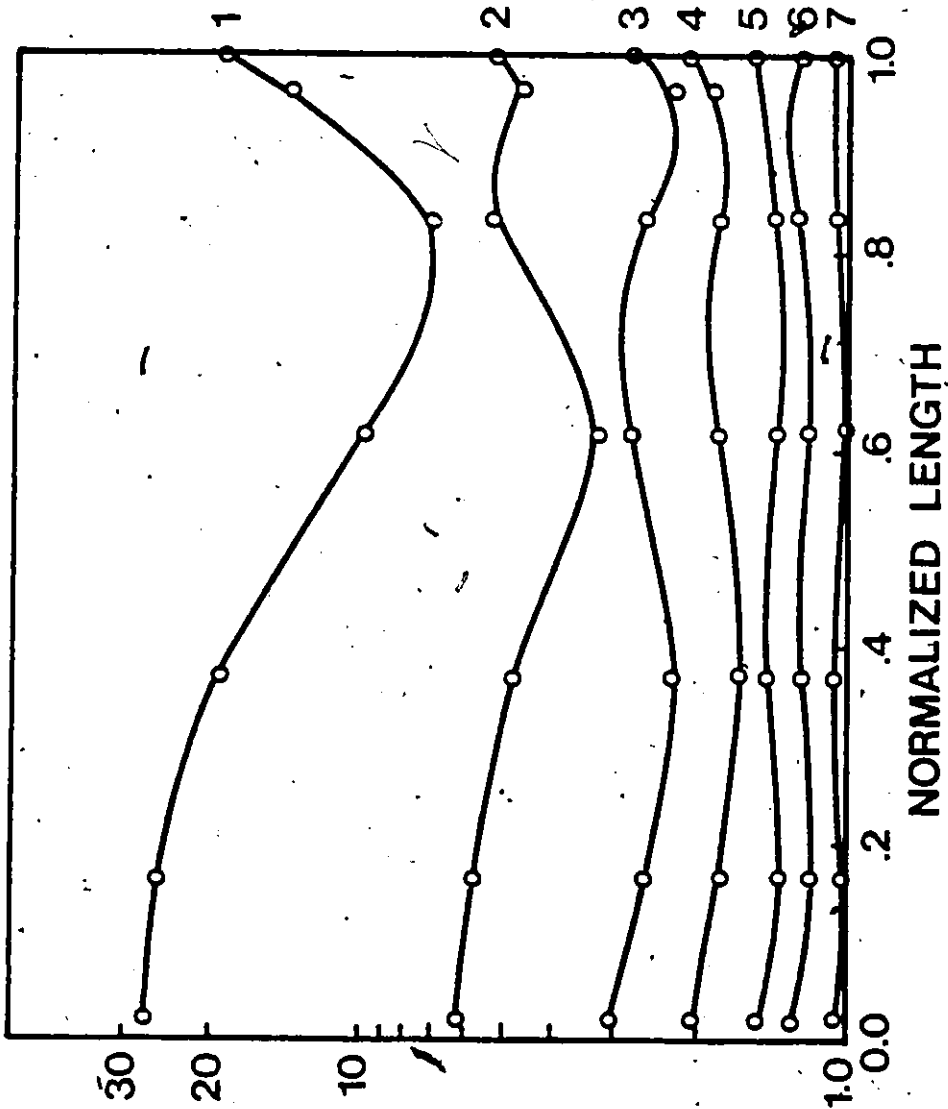
Number of Sensors	$R_V = 10I$	$R_V = 4I$	$R_V = I$
	$R_W = I$	$R_W = \sigma^2(\rho)$	$R_W = 10I$
	(Equation 4.26).		
1	5	5	5
2	4	4	4
3	5	6	6
4	4	3	3
5	5	5	2
6	6	2	7
7	5	4	1

Table 4.2: Sequential Sensor Location Multiple Sensors at the same position permitted. (Refer to Figure 4.1 for Thermocouple Positions)

given the first, is located at position 4 and so on. The preferential selection of sensors at location 5, 4 and 6 which bracket the hot spot location of the profile was not surprising. This is what any chemical engineer would have guessed a priori. However, it does show that the sensor location procedures do lead to sensible results. In other situations where the optimal locations are not at all obvious, the procedures will provide a logical basis for their selection. The insensitivity of the choice of (R_w, R_v) indicates that the reactor dynamics, rather than the stochastic disturbances, influence the optimal location of the sensors.

Figure 4.2 shows the effect that sequentially adding sensors in the optimal manner has on the determinant $P^\infty(k/k-1)$ for the case with the identified reactor stochastics $R_v = 4I$ and $R_w = \sigma^2(\rho)$. For successive sensors, the dotted lines indicate the value of the determinant across the reactor. The minimum on each curve gives the optimum sensor location (as shown in Table 4.1). As we increase the number of thermocouples the determinant, $|P^\infty(k/k-1)|$, will decrease continually. However, one can see that most of the information about the temperatures is obtained after the first few thermocouples have been chosen.

The simultaneous location of 3 sensors was investigated. The thermocouple locations were identical to those placed in the sequential fashion for the covariance matrices considered.



$|P^{\infty}(k/k-1)|$ for sequential additions
 $\min |P^{\infty}(k/k-1)|$ for 7 sensors

Figure 4.2: Objective function v_4 evaluated at different sensor locations.
 $R_V = 4I, R_W = \sigma^2(p)$

The reactor model (4.25) is discrete and expresses the temperatures at the collocation points only. Having placed the first thermocouple near the hot spot, we cannot decide how close the next thermocouple should be placed since we have restricted our thermocouple locations to the collocation points. A laborious approach would be to increase the number of collocation points and examine the sensor location for this new model. An easier approach is to allow multiple independent sensors at the same collocation point. Although this may not be appealing from an operating viewpoint, it does allow us to examine the sensitivity of sensor locations to the number of collocation points.

Table 4.2 shows the results allowing for multiple sensors at the same collocation point. Only when the signal to noise ratio is low do we find clustering of sensors.

To assess the extent to which the optimal stochastic control scheme of Jutan et al (1977b) might be expected to degrade if fewer thermocouples were used, the procedure outline in Section 2.4 was applied. Using the optimal one sensor, two sensor, and three sensor locations obtained by minimizing v_4 for $m = 1, 2,$ and 3 . The trace of the state variable covariance matrix ($\text{trace Var}(\underline{x})$) was evaluated for each of these cases under the L.Q. control using $Q_1 = I$ in the performance index (4.15) and using a Q_2 which gave essentially the same $\text{Var}(\underline{u})$ in each case. The results are shown in Table 4.3 together with those obtained for one non-optimal

Thermocouples Used	trace $\text{Var}(x)$, ($\text{Var}(u)$ constant)
None-(open loop)	59.6
5	44.0
4,5	42.9
4,5,6	42.8
all	42.6
1, 4 (non-optimal)	51.9*

*when thermocouples 1, 4 are used it was impossible to reduce the trace $\text{Var}(x)$ below this value no matter what $\text{Var}(u)$ was used.

Table 4.3: Effect of Thermocouple position on controller performance $R_v = 4$, $R_w = \sigma^2(\rho)$ (Equation 4.26), $Q_1 = I$ (For thermocouple locations refer to Figure 4.1)

placement of the thermocouples, for complete measurement (all 7 positions), and for the predicted open-loop process. (The open-loop variance predicted is only entered here for comparative purposes and it is not accurate since the extreme nonlinearities in the reaction system lead to reactor instability in this situation. The predicted results under the closed-loop conditions, however, agree well with experimental results.) One can conclude from these results that by using only one or at most two optimally located thermocouples one should be able to achieve essentially the same quality of control as with the entire temperature profile (all seven measurements). However, the same conclusion cannot be said for the situation where the thermocouples are poorly located. For this reactor we know from operating experience that a variance as large as 51.9 appears to preclude stable operation.

If one is interested in concentrations rather than temperatures one can locate thermocouples to obtain good estimates of the effluent concentration. In this case we minimize, by the choice of H , the objective function

$$v_7 = |GP^{\circ}(k/k-1)C'| \quad (4.29)$$

For the covariances of R_w and R_v previously considered, the first optimally located sensor was always at position 5. Additional sensors located in a sequential fashion did not coincide with those positions found by minimizing v_4 . However, most of the

information about the effluent concentrations was obtained from the first thermocouple and additional thermocouples had only a small effect on the value of the objective function.

4.6 Summary

Theory for the optimal location of sensors closely parallels theory on the optimal design of regression experiments in the statistical literature. This theory was applied to the optimal location of thermocouples along a packed bed tubular reactor carrying out a highly exothermic series-parallel reaction. The results of this analysis revealed that very efficient state estimation can be accomplished using only one or two thermocouples located in the region of the hot spot. The effect of sensor location on the performance of a linear quadratic controller was also studied. The variation in the temperature profile using linear quadratic control based on only one or two optimally located thermocouples was shown to be almost identical to that obtained using complete state measurements. This was not true for non-optimally located thermocouples.

CHAPTER 5

FORECASTING AND DIMENSIONALITY REDUCTION OF MULTIPLE TIME SERIES

5.1 Introduction

In the previous sections, we investigated some topics in reactor control and modelling. The reactor model was obtained from an analysis and simplification of the mass and energy balances. The use of such a mechanistic model gives tremendous understanding into the manner in which the process operates. This insight must be reconciled against the considerable effort required to formulate such a dynamic model and simplify it into a form suitable for control. Dynamic models of a process can also be obtained by fitting empirical models to input/output data. For multivariable processes, these models usually have a large number of parameters to be estimated. The estimation and diagnostic checking for parsimonious multivariate time series models are not trivial extensions of the techniques outlined in Box and Jenkins (1970) for univariate time series.

In this chapter, we will investigate the use of a multivariate autoregressive structure as a means of obtaining a dynamic/stochastic model of the process. As well, the application of some multivariate statistical techniques to reduce the dimensionality of the process output will be explored.

5.2 Time Series Models

A linear dynamic stochastic model of a process can be represented as

$$\underline{Y}(k) = V(z^{-1})\underline{U}(k) + \underline{N}(k) \quad (5.1)$$

z^{-1} is the backward shift operator such that $z^{-b}\underline{Y}(k) = \underline{Y}(k-b)$. $V(z^{-1})$ is an $m \times r$ transfer function matrix. Each element of $V(z^{-1})$, $v_{ij}(z^{-1})$ is the transfer function relating the i th output to the j th input. $v_{ij}(z^{-1})$ is of the form

$$v_{ij}(z^{-1}) = \frac{\omega_0 - \omega_1 z^{-1} - \dots - \omega_s z^{-s}}{1 - \delta_1 z^{-1} - \dots - \delta_r z^{-r}} z^{-b} \quad (5.2)$$

The number of whole periods of delay b is

$$b = 1 + \text{modulus} \left(\frac{\tau_{ij}^d}{T} \right) \quad (5.3)$$

where τ_{ij}^d is the transport delay between the i th input and the j th output. Although not indicated in the notation, the orders of the transfer function (r, s) are dependent on the indices i and j .

The stochastic disturbances may be represented by an ARIMA model of the form

$$(I - \phi_1 z^{-1} - \dots - \phi_p z^{-p})(1 - z^{-1})^d \underline{IN}(k) = (I - \theta_1 z^{-1} - \dots - \theta_q z^{-q}) \underline{a}(k) \quad (5.4)$$

$\underline{a}(k)$ is a multivariate white noise sequence with mean $\underline{0}$ and covariance matrix $\underline{\Sigma}_a$.

The structure (r,s,b) , (p,d,q) and parameters of (5.2) and (5.4) must be determined from input/output data. In multivariate processes, the spectrum of $\underline{N}(k)$ does not have a unique ARIMA representation, Whittle (1963). A canonical form for $\underline{N}(k)$ must therefore be specified.

A suitable canonical structure (r,s,b) , (p,d,q) may be tentatively specified using cross correlation techniques, Wilson (1970). The parameters may then be estimated by minimizing the determinant criterion

$$J_1 = \left| \sum_{j=1}^N \underline{a}(j)\underline{a}'(j) \right| \quad (5.5)$$

This is equivalent to maximizing the approximate likelihood function, Wilson (1970). Wong (1977) employed this technique to fit the effluent concentrations (three outputs) to the flowrates of butane and hydrogen, for data collected from the butane-hydrogenolysis reactor. The entire procedure is far from trivial. Complications arise if the stochastic disturbances affecting the process are not of full dimension. In this case, J_1 is theoretically zero for all values of the parameters. When the data is collected with a feedback controller in use, structural identification is more involved.

To circumvent these difficulties, Akaike (1971) has promoted the use of pure autoregressive models as a means of representing dynamic/stochastic systems. The autoregressive models of unspecified order p represents one possible choice of a canonical form for representing multivariate time series. Although this form could usually not be parsimonious in its use of parameters, the parameters do enter linearly thereby admitting linear least squares estimation. This latter advantage has often argued to override any objections to the former disadvantage.

Consider the process model where the input-output behaviour is given by

$$\begin{aligned} \underline{Y}(k) = & \Pi_1 \underline{Y}(k-1) + \Pi_2 \underline{Y}(k-2) + \dots + \Pi_\ell \underline{Y}(k-\ell) \\ & + \Gamma_0 \underline{U}(k-1) + \dots + \Gamma_m \underline{U}(k-1-m) + \underline{a}_1(k) \end{aligned} \quad (5.6)$$

or equivalently

$$\underline{Y}(k) = \Pi(z^{-1}) \underline{Y}(k-1) + \Gamma(z^{-1}) \underline{U}(k-1) + \underline{a}_1(k) \quad (5.7)$$

The parameters (Π_i, Γ_i) include the effect of the process dynamics and autoregressive terms from the stochastic model. This representation is certainly not parsimonious. Due to the large number of terms, and confounding of terms, it is difficult to give a physical interpretation to the parameters. If the objective in model building is to design a stochastic controller

this procedure has some merits. The design of multivariable stochastic controllers is most readily accomplished when the input/output behaviour is expressed in state space form.

There are a number of ways to obtain a state space representation for a dynamic stochastic model of the form (5.1). One method, MacGregor (1973), is to first write (5.1) in the form (5.7). A state representation can then be written down by inspection. Since, in this method, it is necessary to obtain an input/output model of the form (5.7), it is not unreasonable then to directly estimate the parameters of this model.

The structure of (5.7) is such that the parameters can be efficiently estimated by least squares. The determination of the orders (l, m) is the only uncertain aspect of this model building procedure. Akaike (1974, 1978a, 1978b) has investigated order determination for autoregressive models of the form

$$\underline{z}(k) = \phi_1 \underline{z}(k-1) + \dots + \phi_p \underline{z}(k-p) + \underline{a}(k) \quad (5.8)$$

The order of the autoregression is taken as that value of p which minimizes the statistic

$$AIC = N \ln |\hat{\hat{t}}_a| + 2k^* \quad (5.9)$$

N is the length of the data record, and k^* is the number of free parameters estimated. $\hat{\hat{t}}_a$ is the maximum likelihood estimate of \hat{t}_a . As more terms are added to the model $|\hat{\hat{t}}_a|$ decreases. The AIC statistic represents a compromise between a model that fits the data better and the increased number

of parameters required to obtain this fit. The AIC criterion has a theoretical justification, and this is discussed in Akaike (1974).

To use this technique for dynamic models whose input/output behaviour is described by (5.6), $\underline{z}(k)$ is defined as $\underline{z}'(k) = (\underline{Y}(k), \underline{U}(k))'$. If it is suspected that the disturbances are nonstationary in their mean, the input and output should be differenced, Box and Jenkins (1970). By including the manipulated variable in the autoregressive model (5.8), it is obvious that the model order p which minimizes (5.9) is a compromise between that required to adequately represent the process model (5.7) and that required to fit the input sequence.

The input sequence can be generated in a number of ways. If no feedback controller is employed during the period of data collection, the $U(k)$ will be of the form

$$\phi(z^{-1})\underline{U}(k) = \underline{a}_2(k) \quad (5.10)$$

where $\phi(z^{-1})$ is a matrix polynomial in the backward shift operator. In this instance, (5.7) and (5.10) are in forms that can be represented by an autoregressive model.

In many instances, due to safety or economic considerations, the data must be collected while the output is under feedback control. During the period of data collection, let us suppose that the feedback controller is of the form

$$\underline{U}(k) = K_1(z^{-1})\underline{Y}_1(k) + K_2(z^{-1})\underline{U}(k-1) + \underline{a}(k) \quad (5.11)$$

$K_1(z^{-1})$ and $K_2(z^{-1})$ are matrix polynomials in the backward shift operator. $\underline{a}(k)$ is a mean zero process, uncorrelated in time. We notice that the manipulated variable $\underline{U}(k)$ is a function of the current value of $\underline{Y}(k)$. However, the autoregressive structure, (5.8) does not admit contemporaneous relationships among the variables $\underline{z}(k)$. Substituting (5.7) into (5.11) we obtain

$$\begin{aligned} \underline{U}(k) = & K_1(z^{-1})\Pi(z^{-1})Y(k-1) + \{K_1(z^{-1})\Gamma(z^{-1}) + K_2(z^{-1})\}\underline{U}(k-1) \\ & + K_1(z^{-1})\underline{a}_1(k) + \underline{a}(k) \end{aligned} \quad (5.12)$$

or

$$\underline{U}(k) = \phi_{11}(z^{-1})\underline{Y}(k-1) + \phi_{12}(z^{-1})\underline{U}(k-1) + \underline{a}_3(k) \quad (5.13)$$

We note that $\underline{a}(k)$ in (5.13) will be autocorrelated unless the feedback controller is proportional to $\underline{Y}(k)$, i.e., $K_1(z^{-1}) = K_1$. Equations (5.7) and (5.13) are a joint autoregressive model of the form (5.8). Depending on the order of the polynomials in the various equations, the joint autoregressive model (5.7) and (5.13) might be of higher order than required to separately fit the process model (5.7) and the controller equation (5.11).

A lower order autoregression might be obtained if a model of the form

$$\phi_0^* \underline{z}(k) = \phi_1^* \underline{z}(k-1) + \dots + \phi_p^* \underline{z}(k-p) + \underline{a}^*(k) \quad (5.14)$$

were fit to the data. $\underline{a}^*(k)$ is a multivariate mean zero white noise process with variance \underline{I} . The matrix ϕ_0^* must be at least

lower triangular, Quenouille (1957). The usual procedure for estimating ϕ_i^* is to estimate ψ_i , in (5.8), factor \hat{f}_a as

$$\hat{f}_a^{-1} = \phi_0^{*'} \phi_0^* \quad (5.15)$$

and then multiply (5.8) by ϕ_0^* to obtain (5.14). Since using the model structure (5.14) may lead to a lower order model in the presence of feedback, a better approach would be to simultaneously identify the ϕ_i^* 's directly, although this was not investigated.

In spite of some difficulties, the AR representations provide a convenient means of identifying models when the data is collected under feedback operation, Akaike (1978b). The feedback, however, cannot be noise free. If there is nearly perfect feedback, then the parameters will be highly correlated. They will also be sensitive to the accumulation of round off errors in their calculation since a nearly singular matrix must be inverted. To avoid these problems, there must either be noise in the feedback loop, a nonlinear feedback controller in use during the period of data collection (or switches between several linear ones), or a 'dither' signal, uncorrelated with the process output, artificially introduced into the feedback loop, Akaike (1967), Box and MacGregor (1973), Gustavsson et al (1977). The actual value of the dither signal need not be recorded, as it is not used explicitly to determine the structure of the input/output model.

The autoregressive model fitting proposed by Akaike eliminates many of the difficulties in fitting multivariate time

series model to input/output data. The parameters of the model are very easily estimated but one must avoid the pitfall of not carefully examining the data. Autoregressive model fitting is well suited for the case where we wish to quickly obtain a state space representation of the process, Akaike (1978a).

5.2.1 Application of Autoregressive Fitting to Reactor Data

Let us consider an application of these ideas to the butane hydrogenolysis reactor. In a previous experimental study (Wong (1978)) the flowrates of hydrogen and butane were perturbed, and temperatures along the central axis of the reactor measured. The flow rate of hydrogen was a stochastic process. To stabilize the reactor, the hot spot (the maximum temperature) was controlled by manipulating the butane flow rate. A 'dither' signal uncorrelated with the hot spot temperature or hydrogen flow rate, was superimposed on the butane flow. This had the effect of reducing the correlation between the hotspot temperature and butane flow rate that resulted from the presence of the feedback controller. Three-hundred and sixty-nine sets of temperature versus flowrates were obtained. The average temperature profile for this data collection experiment is shown in Figure 5.1.

Let us define the vector $\underline{Y}(k)$ as the set of mean corrected temperatures along the reactor. $\underline{U}(k)$ is the mean corrected vector of hydrogen and butane flowrates. When plotted,

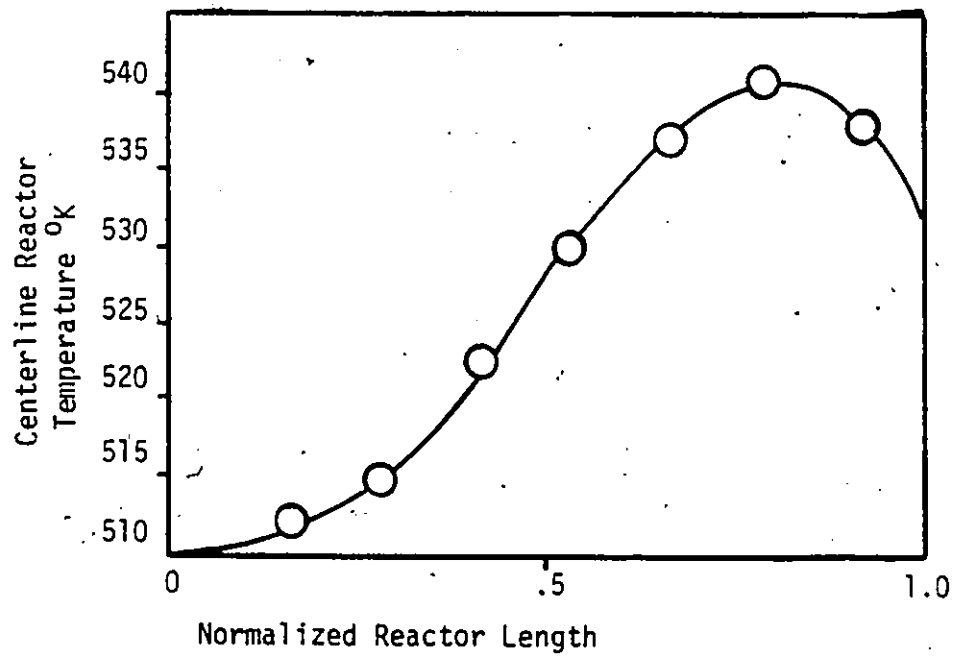


Figure 5.1: Average Temperature Profile
Circles denote measurement positions. Reference, Wong (1977)

the temperatures appeared to be stationary. The individual autocorrelation functions of the temperatures damped out quickly also indicating stationarity. The variable $\underline{z}'(k) = (\underline{Y}(k)\underline{U}(k))'$ was fit as an autoregression using the sequential estimation procedure of Whittle (1963b). The AIC statistic versus model order is shown in Table 5.1. Using the AIC criterion, a second order model was selected as representing the data adequately.

AIC	Order of AutoRegression
4883	0
3788	1
3670	2
3694	3
4651	4

Table 5.1: AIC Statistic for Reactor Data

The variance-covariance matrix of the estimated parameters is cumbersome to evaluate. Let us define A as

$$A = \begin{bmatrix} \sum_{k=1}^N \underline{z}(k)\underline{z}'(k) & \sum_{k=2}^N \underline{z}(k)\underline{z}'(k-1) & \sum_{k=3}^N \underline{z}(k)\underline{z}'(k-2) \\ \text{sym} & \sum_{k=1}^N \underline{z}(k)\underline{z}'(k) & \sum_{k=2}^N \underline{z}(k)\underline{z}'(k-1) \\ & & \sum_{k=1}^N \underline{z}(k)\underline{z}'(k) \end{bmatrix} \quad (5.16)$$

(In general, A will be $n \times n$ where n is the dimension of $\underline{z}(k)$.) The variance-covariance matrix of the parameters is given by, Anderson (1957)

$$\hat{\underline{t}}_a \otimes A^{-1} \quad (5.17)$$

where \otimes denotes the kronecker product. The determinant of this matrix is a measure of the correlation among the parameters. It can be shown that, Anderson (1957)

$$|\hat{\underline{t}}_a \otimes A^{-1}| = |\hat{\underline{t}}_a|^n \cdot |A^{-1}|^{np} \quad (5.18)$$

If there is perfect feedback between the elements of $\underline{y}(k)$ and $\underline{u}(k)$, then A will be nearly singular, and the parameters will be highly correlated.

A state model is conveniently obtained from the input/output model

$$\underline{x}(k+1) = \begin{bmatrix} \phi_{11}^{(1)} & I_7 \\ \phi_{11}^{(2)} & 0 \end{bmatrix} \cdot \underline{x}(k) + \begin{bmatrix} \phi_{12}^{(1)} \\ \phi_{12}^{(2)} \end{bmatrix} \underline{u}(k) \quad (5.19)$$

$$\underline{y}(k) = \underline{y}(k) = (I_7 \quad 0) \underline{x}(k) + \underline{y}(k) \quad (5.20)$$

I_7 denotes the 7 x 7 identity matrix. $\phi_{11}^{(j)}, \phi_{12}^{(j)}$ are obtained from.

$$\phi^{(j)} = \begin{bmatrix} \phi_{11}^{(j)} & \phi_{12}^{(j)} \\ \phi_{21}^{(j)} & \phi_{22}^{(j)} \end{bmatrix} \quad (5.21)$$

The covariance of $\underline{y}(k)$ is given by the 7×7 upper left hand submatrix of the (9×9) covariance matrix $\hat{\Sigma}_a$.

The state model (5.19) has fourteen states. However, there may be models of small order, which can give the same input/output behaviour. The state model, of lowest dimension, is both controllable and observable, and is known as a minimal realization. The realization (5.19) and (5.20) can be used as a starting point to find such a representation, Rissanen (1976).

The temperatures along the reactor must satisfy the non-linear mass and energy balances. We might suspect, therefore, that not all the temperatures can move freely, and that there may only be several degrees of freedom in the movement of the temperature profile. For example, the temperature profile may be constrained to move up and down or shift along the reactor. The model fit using the autoregressive approach of Akaike is a linear approximation to the process dynamics. Using this model, we should be able to establish the dimensionality or degrees of freedom of the linearized process.

The model can also be used for forecasting and control. We might anticipate that not all the temperatures can be equally well predicted from observations of past history. As seen in the previous chapter, control of a stochastic variable is heavily dependent on the predictability of this variable. It is of interest, therefore, to establish which variables are the most predictable. These topics will be explored in the next section.

5.3 Dimensionality Reduction of the Output Space

In the previous section, we used a simple technique to build an input/output model for the butane hydrogenolysis reactor. In this section, some multivariable statistical techniques are used to further analyze the results.

Linear processes can be represented by models of the form

$$\underline{Y}(k) = \hat{\underline{Y}}(k/k-1) + \underline{a}(k) \quad (5.22)$$

The prediction of the process output, given information at time $k-1$, $\hat{\underline{Y}}(k/k-1)$, includes past values of the input and output, and exogenous variables. $\underline{a}(k)$ represents that part of the process output at time k which is unpredictable and uncontrollable at time $k-1$. The predictor is the conditional mean of $\underline{Y}(k)$ given information to time $k-1$. Since $\underline{a}(k)$ is uncorrelated with $\hat{\underline{Y}}(k/k-1)$, the variance-covariance matrix of $\underline{Y}(k)$, $\hat{\Sigma}_Y$, can be written as

$$\hat{\Sigma}_Y = \hat{\Sigma}_{\hat{Y}} + \hat{\Sigma}_a \quad (5.23)$$

Analysis of the covariance matrices will indicate fundamental relationships among the variables. The results of this analysis can be used to establish the dimensionality of the outputs, and the most predictable outputs. The sample covariance matrices $\hat{\Sigma}_{\hat{Y}}$ and $\hat{\Sigma}_a$ for reactor temperature data are given in Tables 5.2 and 5.3. Extensive use will be made of these matrices in the next sections.

.3583							
.3133	.4810						
.1416	.3030	1.1930					
.1531	.4519	1.8813	3.8047		sym		
-.1116	.2337	2.1670	5.0437	7.7915			
-.4511	-.1206	2.0307	5.1198	8.2424	9.3221		
-.5159	-.3228	1.3029	3.8846	6.6935	7.8855	6.9124	

Table 5.2: Covariance Matrix $\hat{\Sigma}_y$ for AR(2) Model

3.7041							
1.5962	3.1623						
.1019	.2913	4.1573					
1.7738	1.2901	1.0605	3.7780		sym		
1.7614	1.4153	.7511	1.1047	3.5579			
-.0396	-.1084	2.5410	.7431	1.7714	4.6917		
-.2489	-.2797	2.6046	.5493	1.0727	4.2201	5.1210	

Table 5.3: Covariance Matrix $\hat{\Sigma}_a$ for AR(2) Model

5.3.1 Principal Component Analysis

Principal component analysis is a convenient means of establishing the existence of contemporaneous linear relationships among a set of variables. This analysis can then indicate the dimensionality of a set of variates. Anderson (1957) gives the 'classical' interpretation of this method. A different interpretation is offered by Box et al (1973). To summarize the latter briefly, let us suppose that we have a mean corrected set of variables \underline{Y} of dimension m , that is

$$\underline{Y} = \underline{Y}^* - \underline{Y}^m \quad (5.23a)$$

where \underline{Y}^m is the mean of the variables \underline{Y}^* . The covariance matrix of \underline{Y} is taken as \underline{t}_Y .

The trace of \underline{t}_Y can be written as

$$\text{tr} \underline{t}_Y = E\{(\underline{Y}^* - \underline{Y}^m)'(\underline{Y}^* - \underline{Y}^m)\} = \lambda_1 + \lambda_2 + \dots + \lambda_n \quad (5.24)$$

where the λ_i 's are the eigenvalues of \underline{t}_Y arranged in decreasing order of magnitude. For each eigenvalue there is a corresponding eigenvector $\underline{\alpha}_i$. Since \underline{t}_Y is a symmetric positive definite matrix, it is always possible to construct an orthogonal set of eigenvectors, even when there are eigenvalues of multiplicity, Noble (1969). If we scale the eigenvalues so that $\underline{\alpha}_i' \underline{\alpha}_i = 1$, (5.24) can be written as

$$\text{tr} \underline{t}_Y = \underline{\alpha}_1' \underline{t}_Y \underline{\alpha}_1 + \underline{\alpha}_2' \underline{t}_Y \underline{\alpha}_2 + \dots + \underline{\alpha}_n' \underline{t}_Y \underline{\alpha}_n \quad (5.25)$$

The magnitude of the eigenvalues of $\hat{\Sigma}_Y$ can be partitioned roughly into three sizes, Box et al (1970). These groupings correspond to three different kinds of relationships among the variables.

Suppose there are r_a restrictions on the variables \underline{Y}^* of the form

$$A_1 \underline{Y}^* = \underline{C}_1 \quad (5.26)$$

where \underline{C}_1 is a vector of constants. The matrix A_1 is $r_a \times n$ and is of rank r_a . When these restrictions are present, then r_a eigenvalues of $\hat{\Sigma}_Y$ will be zero. Since one usually only has a sampled estimate of this covariance matrix, these eigenvalues will be zero except for roundoff error. This type of restriction can arise in a number of ways. For example, an analytical instrument may not give independent measurements on all components, i.e., a chromatograph. These dependencies are also found when some of the variables are computed as linear combinations of the other variables, i.e., from the reaction stoichiometry.

Let us suppose that there are r_b linear combinations of the \underline{Y} 's unaffected by its past history, or by the settings of the manipulated variables. With reference to (5.22), we see that the covariance of $\hat{\Sigma}_Y$ for these variables will be of the order of magnitude of the corresponding elements of $\hat{\Sigma}_a$. As shown in Box et al (1973), this type of behaviour occurs when there are restrictions on the expected values of the \underline{Y}^* 's of the form

$$E\{A_2 Y^* Y^*^T\} = C_2 \quad (5.27)$$

Restrictions of this form reflect fundamental linear relationships among the variables. If there were no residual or measurement error in the variables, r_b eigenvalues of $\hat{\Sigma}_Y$ would be zero. However, the process or measurement noise obscures these relationships. The eigenvalues of $\hat{\Sigma}_Y$ will reflect these relationships. If there are r_b restrictions of the form (5.27), r_b eigenvalues of $\hat{\Sigma}_Y$ will be of the order of, Box et al (1973);

$$E(\lambda_i) = \alpha_i^T \hat{\Sigma}_a \alpha_i \quad (5.28)$$

when $\hat{\Sigma}_a$ is unknown, a rough approximation can be obtained by substituting for it the approximate variance-covariance matrix of the measurement errors.

The remaining $n-r_a-r_b$ eigenvalues of $\hat{\Sigma}_Y$ will be larger as they include the effect of the measurement noise $\hat{\Sigma}_a$, and the variation caused by the perturbation of the manipulated variables, $\hat{\Sigma}_U$. The variation in the outputs as measured by the trace of $\hat{\Sigma}_Y$, due to the process variation, and not only the measurement or residual process noise, can be attributed to $n-r_a-r_b$ eigenvalues of $\hat{\Sigma}_Y$, or equivalently, $n-r_a-r_b$ linear combinations of the Y 's. Thus the dimensionality of the process outputs is $n-r_a-r_b$.

From an analysis of the eigenvectors, it is usually not possible to calculate the matrices A_1 and A_2 , Box et al (1973). It is possible though to decide from which group an arbitrary

linear combination of \underline{Y} , of the form $\underline{\alpha}'\underline{Y}$ belongs. To do so we must briefly review the classical interpretation of principal components.

Suppose we were to find the linear combination of \underline{Y} , $\underline{\alpha}'\underline{Y}$, that has the maximum variance among all possible linear combinations subject to the scaling constraint that $\underline{\alpha}'\underline{\alpha} = 1$. This is equivalent to finding the unconstrained maximum of

$$J_1 = \frac{\underline{\alpha}'\underline{\Sigma}\underline{\alpha}}{\underline{\alpha}'\underline{\alpha}} \quad (5.29)$$

It can be shown, Anderson (1957), that $\underline{\alpha}_1$ is the eigenvector of $\underline{\Sigma}$ corresponding to the largest eigenvalue of $\underline{\Sigma}$ which we have defined to be λ_1 . $\underline{\alpha}_2, \underline{\alpha}_3, \dots, \underline{\alpha}_n$ are those linear combinations of \underline{Y} having the maximum variance subject to the additional restrictions that $\underline{\alpha}_i$ and $\underline{\alpha}_j$ be orthogonal for $i \neq j$. The variances of the linear combinations are $\lambda_1, \lambda_2, \dots, \lambda_n$.

Can any physical interpretation be given to the transformed variables $\underline{\alpha}_i$? If any interpretation is possible, it can only be assigned to the linear combination corresponding to the largest eigenvalue (or the smallest, if we first sought the linear combination having the smallest variance). The reason for this is that orthogonality constraints have been imposed on all but the first linear combination. For example, $\underline{\alpha}_2$ is the linear combination of \underline{Y} , orthogonal to $\underline{\alpha}_1$, having maximum variance. These orthogonality constraints will mask physically meaningful relationships among the variables.

Even though the transformed variables may not be physically meaningful, one can still make use of the eigenvectors. Let us suppose that there are no zero eigenvalues. There will then be r_b linear combinations of \underline{Y} whose expectations are constant, and $n-r_b$ linear combinations of \underline{Y} which exhibit true process-variation. The eigenvectors of \underline{Y} form an orthonormal basis (in the algebraic sense) for the n -dimensional space of \underline{Y} . In other words, any linear combination of \underline{Y} can be described by a weighted sum of these eigenvectors. The r_b eigenvectors, corresponding to the appropriate eigenvalues, are a basis for an r_b dimensional subspace of the \underline{Y} 's which have no process variation. We shall call this subspace the null space of \underline{Y} . The remaining $n-r_b$ eigenvectors form a basis for those linear combinations of \underline{Y} for which there is process variation. It is possible then, to test whether a certain linear combination of variables lies in the null space.

To illustrate this, let \underline{Y} be of dimension three. We wish to test whether the vector $\underline{a}' = (a_1, a_2, a_3)$ lies in the plane spanned by two of the normalized eigenvectors of \underline{Y} , $\underline{\alpha}_1$ and $\underline{\alpha}_2$. One means of testing this is to compute the angle between the vector \underline{a} and the plane (call this the π -plane) spanned by $\underline{\alpha}_1$ and $\underline{\alpha}_2$, see Figure 5.2. Let the vector \underline{n} be orthogonal to $\underline{\alpha}_1$ and $\underline{\alpha}_2$. \underline{n} is also orthogonal to the π -plane. The parametric equation of a line orthogonal to the plane spanned by $\underline{\alpha}_1$ and $\underline{\alpha}_2$ and passing through the point \underline{a} is

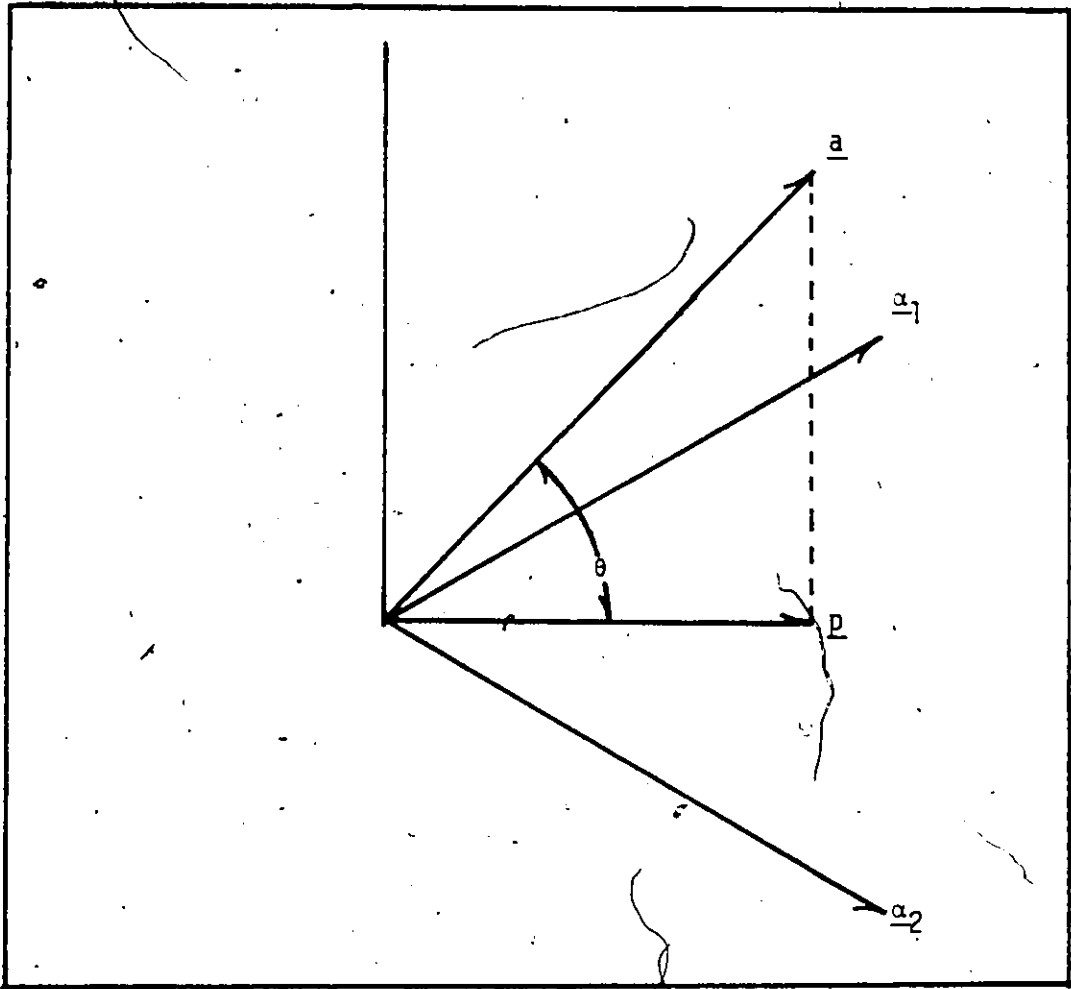


Figure 5.2: Geometry of Principal Components

$$\underline{r} = \underline{a} + s\underline{n} \quad (5.30)$$

Any point in the π -plane can be represented as

$$\underline{p} = \beta_1 \underline{\alpha}_1 + \beta_2 \underline{\alpha}_2 \quad (5.31)$$

The point of intersection of the π -plane and the line (5.30) is given by (5.31) with

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \alpha_1' \alpha_1 & \alpha_1' \alpha_2 \\ \alpha_2' \alpha_1 & \alpha_2' \alpha_2 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1' \underline{a} \\ \alpha_2' \underline{a} \end{bmatrix} \quad (5.32)$$

Since $\underline{\alpha}_1$ and $\underline{\alpha}_2$ are orthogonal (5.32) reduces to

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \alpha_1' \\ \alpha_2' \end{bmatrix} \underline{a} \quad (5.33)$$

The cosine of the angle between the π -plane and the vector \underline{a} is therefore given by

$$\cos \theta = \frac{\underline{a}' \cdot \underline{p}}{|\underline{a}| |\underline{p}|} \quad (5.34)$$

The values of β_1 and β_2 in (5.32) and (5.33) are identical to those obtained by fitting a linear regression model of the form

$$z = \underline{X}' \underline{B} + \epsilon \quad (5.35)$$

where $z = \underline{a}'\underline{Y}$, $\underline{X}' = (\underline{\alpha}_1'\underline{Y}, \underline{\alpha}_2'\underline{Y})$ and ϵ is a white noise sequence. We recognize $\cos\theta$ as the square root of the ratio of the sum of squares due to regression to the sum of squares of z . z can be defined as $z = \underline{a}'(\underline{Y} - \underline{Y}_c)$ with no loss of generality, where \underline{Y}_c is a vector of constants. A level term must be included in the regression equation (5.35) in this instance.

5.3.2 Principal Components Analysis of Reactor Temperature Data

The eigenvalues and eigenvectors of $\hat{\underline{\Sigma}}_Y$ are shown in Table 5.4 for the axial temperature data analyzed in section (5.2.1). There are no eigenvalues that are nearly zero. This is not unexpected since the temperatures were measured independently. Substituting $\sigma^2\mathbf{I}$, with $\sigma=2$, for the approximate variance-covariance matrix of $\hat{\underline{\Sigma}}_a$, we would expect that those eigenvalues arising from linear relationships of the form (5.27) would have an expected value of 4.0. On this basis we would conclude that the dimensionality of the temperatures along the reactor axis is two. Using the model built in section 5.2.1, a more precise estimate of $\hat{\underline{\Sigma}}_a$ is available. Using this matrix, the expected values for the eigenvalues are shown in Table 5.5. Again we would conclude that the dimensionality of the process is two, or perhaps three. For the reactor temperatures, this result is perhaps not surprising. If there were no internal disturbances in the reactor, we would not expect the temperatures to vary independently, as they must satisfy the energy balance which is

Eigenvalues	Eigenvectors						
37.51	.025	.032	.220	.317	.482	.592	.516
9.25	.523	.467	.036	.467	.307	-.238	-.370
4.26	-.055	.049	.841	.238	-.478	-.058	.008
2.56	-.310	-.477	-.262	.754	-.098	-.040	-.169
2.12	.662	-.147	-.268	.124	-.496	.023	.454
1.80	-.418	.723	-.307	.196	-.320	.024	.259
0.53	.117	.082	-.091	-.052	-.294	.766	-.544

Table 5.4: Eigenvalues and Eigenvectors of $\hat{\Sigma}_Y$

Eigenvalues	37.51	9.25	4.26	2.56	2.12	1.80	.53
Expected Values	11.1	6.25	3.28	2.33	2.19	1.68	.48

Table 5.5: Expected Values of the Eigenvalues of $\hat{\Sigma}_Y$

a nonlinear partial differential equation. There would probably be nonlinear relationships among the temperatures. The principal component analysis can only uncover linear dependencies in the data.

The dimensionality of the temperature profile appears to be two. The transformed variables $\alpha_1 Y$ and $\alpha_2 Y$ may not have physical significance. Two important operating variables are the area between the axial temperature and wall temperature, and the hotspot temperature. The former represents the net energy release in the reactor, or it can be interpreted as the average temperature rise above the wall temperature. It is of interest to decide whether these two variables lie in the null space of the temperature profile. The average temperature rise above the wall temperature can be approximated by

$$z = \underline{a}'(\underline{Y} + \underline{\Delta}) \quad (5.36)$$

where $\underline{\Delta}$ is the difference between the average temperature profile and the wall temperature, see Figure 5.1. The vector \underline{a} is taken as

$$\underline{a} = \frac{\underline{\Delta}}{|\underline{\Delta}|} \quad (5.37)$$

In the experiment of Wong (1975) the average wall temperature was 512⁰K. \underline{a} is calculated as

$$\underline{a}' = (.04, .075, .254, .34, .48, .56, .511)$$

The vector \underline{a} nearly coincides with the eigenvector α_1 . The line passing through $\{\underline{a}'(\underline{Y} + \underline{\Delta}), \underline{a}'\underline{\Delta}\}$ subtends an angle of 9.2° with the line passing through $\{\alpha_1'\underline{Y}, 0\}$.

The maximum temperature rise above the wall temperature is also given by (5.36) with \underline{a} taken as

$$\underline{a} = (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0)$$

The line passing through $\{\underline{a}'(\underline{Y} + \underline{\Delta}), \underline{a}'\underline{\Delta}\}$ subtends an angle of 9.6° with the plane spanned by the vectors $\alpha_1'\underline{Y}$, and $\alpha_2'\underline{Y}$. By contrast the line passing through $\{\underline{a}'(\underline{Y} + \underline{\Delta}), \underline{a}'\underline{\Delta}\}$, subtends an angle of 37.5° with this plane when \underline{a} is taken as

$$\underline{a}' = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$$

We would conclude that the hotspot temperature and net energy release in the reactor are variables with real process variation.

A more rigorous analysis is to test the hypothesis that the regression model (5.35) adequately fits the data. The test of this hypothesis is a standard F-test. Using $\underline{a}'\hat{\underline{t}}_a$ as an estimate of the residual-process variation (with the appropriate value of \underline{a}), the hotspot temperature and net energy release were found to statistically lie in the vector space spanned by $\alpha_1'\underline{Y}$ and $\alpha_2'\underline{Y}$ at a significance level of .05. The hypothesis that $(1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)'\underline{Y}$ was spanned by this vector space was rejected at this significance level.

If one were to build a parsimonious dynamic/stochastic model of the reactor, the variables $\underline{\alpha}_1^i Y$ and $\underline{\alpha}_2^i Y$ would be a sensible selection of variables to consider as process outputs. Since these variables are contemporaneously independent, correlation among the parameters of the model fit to input/output data is reduced.

5.3.3 Other Approaches to Dimensionality Reduction

The use of principal components in the frequency domain to reduce the dimensionality of process outputs has been studied by Priestley et al (1974a, 1974b), Haggan and Priestly (1975), and Haggan (1975). In addition to several limitations noted by these authors, the frequency domain analysis obscures what is actually being studied in the time domain. Their work is briefly described below.

$E\{\underline{Y}'\underline{Y}\}$ can be decomposed into its frequency components as

$$E\{\underline{Y}'\underline{Y}\} = \frac{1}{\pi} \int_0^{\pi} \text{trace} (F_{\underline{Y}\underline{Y}}(\omega)) d\omega \quad (5.38)$$

$F_{\underline{Y}\underline{Y}}(\omega)$ is the spectral density of the output at frequency ω .

Expressing the trace $F_{\underline{Y}\underline{Y}}(\omega)$ in terms of its eigenvalues, $E\{\underline{Y}'\underline{Y}\}$ can be written as

$$E\{\underline{Y}'\underline{Y}\} = \frac{1}{\pi} \int_0^{\pi} \sum_{i=1}^n (\alpha_i^*(\omega) F_{\underline{Y}\underline{Y}}(\omega) \alpha_i^*(\omega)) d\omega \quad (5.39)$$

The eigenvalues $\alpha_i^*(\omega)$ are the solutions to

$$F_{YY}(\omega)\underline{\alpha}_i^*(\omega) = \lambda_i(\omega)\underline{\alpha}_i^*(\omega) \quad (5.40)$$

At each frequency, a decision is made as to whether $\lambda_i(\omega)$ is statistically zero. If this is found to be true, $\underline{\alpha}_i^*(\omega)$ is taken as the null vector. After the eigenvectors $\underline{\alpha}_i^*(\omega)$ have been computed for $0 \leq \omega \leq \pi$, the time domain coefficients are calculated. The essential result is that we can obtain a transformed variable that has the largest variance. However, this variable is of the form

$$z_i(k) = \sum_{j=-\infty}^{\infty} \alpha_i'(j)Y(k-j) \quad (5.41)$$

$\alpha_i'(j)$ is the time domain vector corresponding to $\underline{\alpha}_i^*(\omega)$. The transformed variable involves past, present and future values of the process output. As such, it is physically unrealizable. If $\alpha_i'(j)$ were zero for $j < 0$, then we would not have this problem. However, one has no assurance that these values will be non-zero when they are computed from their corresponding spectral estimates, Priestley et al (1974b).

Parzen (1969) has pointed out that spectral methods cannot be used to construct predictors whose memory depends only on present and past information. Causal relations can only be fitted through "innovations" which are obtained by fitting models in the time domain.

We should not interpret the linear combinations of variables having the largest variability (or any variables spanned

by the corresponding eigenvectors of $\hat{\mathbf{t}}_y$) as those variables most in need of control. If any of the responses are under feedback control, the variability in these variables will have been reduced. The linear combinations having the most variability will change if the feedback controller is altered. Collecting the data in an open loop experiment where the inputs are artificially perturbed introduces variation in the process not encountered in normal operation. As a final point to note, the principal components are scale or unit dependent. Thus a large eigenvalue may arise from the units in which the responses are measured. This is of particular importance when the outputs have mixed units, say temperatures and concentrations.

Rao (1975) has used the autoregressive model technique of Akaike to examine the dimensionality of $\underline{a}(k)$. The trace of $\hat{\mathbf{t}}_a$ is decomposed, as per equation (5.25). The variables $\alpha_1^1 Y(k)$, $\alpha_2^1 Y(k), \dots$ are then assumed to be variables in most need of control. If the first few eigenvalues account for most of the variation in the trace of $\hat{\mathbf{t}}_a$, control of the variables $\alpha_1^1 Y(k)$, $\alpha_2^1 Y(k)$ will control most of the variation in the process. This is not true, as the variation in the outputs arises from the predictor and the driving force $\underline{a}(k)$.

Principal component analysis is readily applied to $\hat{\mathbf{t}}_y$ as a dynamic/stochastic model of the process is not required to estimate this matrix. The eigenvalues of the matrix can be grouped roughly with knowledge of the measurement variability.

By analyzing \underline{y} alone, it is impossible to distinguish between the variability due to the unpredictable component $\underline{a}(k)$, and the variability due to the predictable part of the process. In the next section, we will look at the decomposition of \underline{y} into its predictable and unpredictable components.

5.4 Canonical Variate Analysis of Time Series

In the previous section, a variance decomposition of the covariance matrix of the outputs was studied. In this section, a correlation analysis of the time series is examined. Let us consider a time series model of the form (5.22). We seek a linear combination of \underline{y} , $\underline{\alpha}'_1 \underline{y}(k)$, and a linear combination of the predictors, $\underline{\alpha}'_1 \hat{\underline{y}}(k/k-1)$, to maximize the ratio

$$J = \frac{\underline{\alpha}'_1 \hat{\underline{y}} \underline{\alpha}}{\underline{\alpha}'_1 \underline{y} \underline{\alpha}} \quad (5.42)$$

The objective function can be interpreted as finding the linear combination of the outputs, whose variance is most predictable (or forecastable). Box and Tiao (1977) call this canonical variate analysis of time series. Minimization of (5.42) requires solution of a generalized eigenvalue problem. Subsequent linear combinations of $\underline{y}(k)$ are found that are the next most forecastable subject to the constraint that they be orthogonal to the previous linear combinations. If the denominator in (5.42) is changed to $\underline{\alpha}'_1 \underline{a} \underline{\alpha}$, then the same results will be obtained, but the objective function is interpreted as

finding the linear combination of the predictors that maximize the signal to noise ratio. The results of the most forecastable variate analysis are not dependent on the units in which the responses are measured.

This analysis is not intended to reduce the dimensionality of the output space. Rather, it is to indicate which linear combinations of the outputs $\underline{Y}(k)$ can be best predicted given information to time $k-1$. Those linear combinations that are poorly predicted, are essentially white noise, Box and Tiao (1977). The space defined by these white noise vectors $\underline{a}_i \underline{Y}(k)$ is unforecastable and therefore can be dropped in further forecasting analysis. Not being forecastable they are also probably not controllable.

Jutan et al (1977c) used this technique to analyze the correlation structure of the 7-dimensional model

$$\underline{N}(k) = \phi \underline{N}(k-1) + \underline{a}(k) \quad (5.43)$$

This model was identified as adequately characterizing the disturbances affecting the reactor. Their analysis indicated that two or three of the disturbances were forecastable, and the remainder were white noise. One can look upon this technique as a dimensionality reduction of the ϕ matrix in (5.43).

The results of the canonical variate analysis and principal component analysis will change if any part of the process configuration is altered. If a different feedback controller

is implemented, the most forecastable variates can be expected to change. Of what value are these techniques? For data collected under normal operating conditions, these techniques can be used to explain the source of variation in the process outputs, i.e., is it mostly white noise, and therefore uncontrollable. The results of the most forecastable variate analysis indicate which variables are the most effective predictors of the process outputs.

In some instances, the inputs are artificially perturbed. The intention may be to build a model of the process and perhaps design a more effective controller. In a designed experiment, the manipulated variables should have sufficient variability to elicit a process response. Inferences about the dimensionality of the output space should not be altered if the structure of the input sequence is changed. The results of the canonical variate analysis would appear to be more susceptible to changes in the structure of the feedback controller.

5.4.1 Canonical Variates and Canonical Correlation Analysis

Anderson (1957) discusses a multivariate statistical technique known as canonical correlation analysis. When $\underline{a}(k)$ comes from a normal distribution, the joint distribution of $\underline{Y}(k)$ and $\hat{Y}(k/k-1)$ is normal, with mean 0 and covariance matrix

$$\underline{\Sigma} = \begin{bmatrix} \hat{\Sigma}_Y + \Sigma_a & \hat{\Sigma}_Y \\ \hat{\Sigma}_Y & \hat{\Sigma}_Y \end{bmatrix} \quad (5.44)$$

Canonical correlation analysis of the variables $\underline{Y}(k)$ and $\underline{Y}(k/k-1)$ leads to the same analysis and results as the canonical variate analysis of Box and Tiao (1977). The interpretation of the results is somewhat different though.

5.4.2 Canonical Variate Analysis of the Reactor Data

Canonical variate analysis was applied to the reactor data. Maximization of (5.42) requires solution of the determinantal equation

$$|\hat{\gamma} - \lambda \hat{\gamma}| = 0 \quad (5.45)$$

For each eigenvalue λ_i , there is an associated eigenvector $\underline{\alpha}_i$. If the eigenvalues are arranged in decreasing order of magnitude $\lambda_1 > \lambda_2 > \dots > \lambda_n$, then the variable most predictable is $\underline{\alpha}_1' \underline{Y}(k)$. The best predictor of this variable is $\underline{\alpha}_1' \hat{\underline{Y}}(k/k-1)$. The next most predictable variable, orthogonal to $\underline{\alpha}_1' \underline{Y}(k)$ and $\underline{\alpha}_1' \hat{\underline{Y}}(k/k-1)$ is $\underline{\alpha}_2' \underline{Y}(k)$, and so on. It is easy to show that

$$\lambda_i = \frac{\underline{\alpha}_i' \hat{\gamma} \underline{\alpha}_i}{\underline{\alpha}_i' \hat{\gamma} \underline{\alpha}_i} \quad (5.46)$$

that is, the ratio of the predictable variance to the total variance of the new canonical variates. Thus λ_i is always in the range $0 \leq \lambda_i \leq 1$. The eigenvalues of (5.45) for the reactor data are shown in Table 5.6. As we can see only two linear combinations can be forecasted. Even so, fifty-nine percent of the variation in the second variable $\underline{\alpha}_2' \underline{Y}$ is white

Eigenvalue	Eigenvectors						
.820	-.121	-.058	-.075	.122	.182	.040	.043
.405	.075	-.003	-.321	-.163	-.088	-.065	.378
.095	.134	.512	-.185	.035	-.183	.115	.097
.065	.150	-.110	-.033	-.270	.508	-.375	.084
.046	.498	-.363	.038	.143	-.127	-.111	.185
.017	-.210	-.112	-.289	.395	.107	-.378	.136
-.008*	.211	-.064	-.259	-.056	-.223	.905	-.717

Table 5.6: Eigenvalues and Eigenvectors from
Canonical Analysis of Reactor
Data

(* eigenvalue negative due to
numerical procedure)

noise. The variables $\alpha_1'Y$ and $\alpha_2'Y$ are taken as a basis for variables that can be forecasted. As in section 5.3, it is possible to test whether an arbitrary linear combination of the outputs is spanned by these vectors. Alternatively, the ratio (5.46) with \underline{a} replacing α_1 is a measure of the predictability of the linear combination $\underline{a}'Y$. The predictability of the average temperature rise above the wall temperature is .68, and the predictability of the hotspot temperature is .66. However, these variables are not the most predictable variables. The linear combination which has the most predictable variation $\alpha_1'Y$, appears to have no physical significance.

In this section, the relative predictability of a time series from its past behaviour has been examined. The connection between this technique and the method of canonical correlation analysis has been noted. In the next section, the use of variate analysis to match up manipulated variables and output variables will be explored.

5.5 Input/Output Pairing

In some instances, it is not known which manipulated variables can be most effectively used to control which output variables. A technique, proposed by Bristol (1966), can be used to select input/output pairs. This analysis, however, selects the input/output pair that are decoupled best in the steady state. In this section, we tentatively look at the possibility of using canonical correlation analysis on transient response

data taken from a system in order to accomplish a decoupling on a dynamic basis.

A reasonable criterion is to search for a linear combination of the output variables $\alpha_j' Y(k)$ that is most highly correlated with a linear combination of the manipulated variance $\beta_j' U(k)$, over some period of time. We therefore seek to maximize by choice of α_j and β_j the objective function

$$J = \alpha_j' \sum_{j=1}^L E(Y(k)U(k-j)) \beta_j \quad (5.47)$$

subject to the scaling restrictions

$$\alpha_j' Y \alpha_j = 1 \quad (5.48)$$

and

$$\beta_j' U \beta_j = 1 \quad (5.49)$$

A low value of L , indicates that we wish to find those linear combinations of variables that are most correlated over the fairly recent past. Having found the first pairing of variables, we look for a second pairing that have maximum correlation, but are orthogonal to $\alpha_j' Y(k)$ and $\beta_j' U(k)$ over the interval $j=1, 2, \dots, L$. This leads to a standard canonical correlation, or canonical variate analysis. The logical result then is that we should control $\alpha_j' Y(k)$ with $\beta_j' U(k)$, $\alpha_2' Y(k)$ with $\beta_2' U(k)$ and so on. Furthermore, $\alpha_j' Y(k)$ and $\beta_j' U(k)$ are uncorrelated with

$\alpha_j' Y(k)$ and $\beta_j' U(k)$, $i \neq j$, over the interval $k=1, 2, \dots, L$. This decoupling would allow the use of multiple loop simple control algorithms, such as proportional integral controllers.

Let us suppose that the process output can be represented as

$$Y(k) = \sum_{j=0}^{\infty} v_j U(k-j) + N(k) \quad (5.50)$$

v_j is the impulse response of the process. Let us assume that $U(k)$ is a white noise sequence uncorrelated with $N(k)$. Then minimization of (5.47) subject to (5.48) and (5.49) is equivalent to maximizing, by choice of α_1 , the ratio

$$J^* = \frac{\alpha_1^T (v_0 + v_1 + \dots + v_N) \dagger_U (v_0 + v_1 + \dots + v_N)' \alpha_1}{\alpha_1^T \left(\sum_{j=0}^{\infty} v_j \dagger_U \sum_{j=0}^{\infty} v_j' + \dagger_N \right) \alpha_1} \quad (5.51)$$

We can view the numerator in (5.51) as the predictable variation in the outputs due to the transfer function alone and not due to the predictors, which include contributions from $N(k)$. We notice that the maximizing linear combinations will depend on the control effort, \dagger_U , we are willing to use.

When the transfer function is not known, the covariances and variances in (5.48-5.49) are replaced by their sampled estimates. When there is pure multivariate feedback, the estimated cross covariances between the input and output do not contain information about the forward transfer function,

which is what we want. Rather, these covariances contain information only about the feedback controller, Box and MacGregor (1973). In the case of an added dither signal, (or imperfect feedback) information about the forward transfer function is obtained by cross correlating the added dither signal and the outputs. As previously noted, Akaike's method does not give us information about the forward transfer function alone and we cannot use the results of the autoregressive model building to estimate the transfer function.

When the open-loop data is available, or the added dither signal has been recorded, the sampled estimates of the covariance matrices can be substituted into (5.47-5.49). It is important to note that when the input sequence is not white noise, the estimates of the cross correlations are themselves autocorrelated, Box and Jenkins (1970).

To get a feel for the technique, the process was assumed to be represented by

$$\begin{bmatrix} Y_1(k) \\ Y_2(k) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} U_1(k) \\ U_2(k) \end{bmatrix} + \underline{a}(k) \quad (5.52)$$

where $\underline{I}_a = I$. In the first example, the inputs were a white noise sequence with covariance $\underline{I}_U = I$. The objective function was taken as

$$J = \underline{\alpha}' \left\{ \sum_{j=1}^2 E\{\underline{Y}(k)\underline{U}'(k-j)\} \right\} \underline{\beta} \quad (5.53)$$

(5.52) was minimized subject to the constraints (5.48) and (5.49) by evaluating the theoretical autocorrelations. The results are summarized in Table 5.7.

$\alpha_1^1 Y(k)$	=	.447 (-Y ₁ (k) + 1.62Y ₂ (k))
$\alpha_2^1 Y(k)$	=	.455 (Y ₁ (k) + 0.58Y ₂ (k))
$\beta_1^1 U(k)$	=	.89 (-U ₁ (k) + 0.52U ₂ (k))
$\beta_2^1 U(k)$	=	.54 (U ₁ (k) + 1.57U ₂ (k))

Table 5.7: Canonical Variate Analysis for White Noise Input

In the second example, U₁(k) was a white noise sequence. However, U₂(k) was a positively correlated input sequence modeled by

$$U_2(k) = .8U_2(k-1) + a_2(k) \quad (5.54)$$

The variance of a(k) was adjusted so that the variance of the input was the same as the previous example. The covariance of the $\underline{a}(k)$ remained unchanged. The results for the minimization of (5.51) are shown in Table 5.8.

$\alpha_1^1 Y(k)$	=	.55 (-Y ₁ (k) + 1.13Y ₂ (k))
$\alpha_2^1 Y(k)$	=	.318 (Y ₁ (k) + 1.42Y ₂ (k))
$\beta_1^1 U(k)$	=	.974 (-U ₁ (k) + .23U ₂ (k))
$\beta_2^1 U(k)$	=	.224 (U ₁ (k) + 4.35U ₂ (k))

Table 5.8: Canonical Variate Analysis for Autocorrelated Input

The results of this simple example are difficult to interpret. The structure of the input sequence has a large influence on the results. This suggests that if the structure of the input sequence changes, the variables $\alpha_i^j Y(k)$, $\beta_i^j U(k)$ will no longer have the desired orthogonality properties.

The matching of input and output, using a criterion such as (5.47) would be of value when there are more variables to be controlled than manipulated inputs. In this section, some preliminary ideas have been presented and a few potential problems noted.

5.6 Summary

The use of an autoregressive structure to fit dynamic/stochastic models has been briefly discussed in this chapter. A state space model of the process is readily obtained using this method. However, the transfer function of the process is not directly available from this analysis.

The result of the autoregressive model fitting can be used to investigate the dimensionality of the process outputs, and the relative forecastability of the outputs. Additional insight into the process operation is obtained if it is possible to find a physical explanation for the results of the principal component and canonical variate analysis. If one intends to build a parsimonious dynamic/stochastic model of the process, the results of the autoregressive model fitting can be used for preliminary analysis of the data. The use of these techniques was

illustrated from operating data collected from the butane hydrogenolysis reactor.

In the next chapter, some properties of univariate stochastic controllers are investigated. Experimental application of adaptive versions of these regulators to control the hot spot temperature in a pilot scale reactor are studied in a subsequent chapter.

CHAPTER 6

FREQUENCY RESPONSE OF STOCHASTIC CONTROLLERS

6.1 Introduction

The synthesis of digital filters has been studied extensively in the communications literature. These filters are used to enhance voice and pictorial data transmitted by electronic means. An excellent presentation of these techniques is given by Oppenheim and Schaffer (1975). There are basically two approaches to the design of digital filters. With a suitable selection of the sampling interval, one may design digital filters to imitate the spectral characteristics of an analog filter. Alternatively, one may take advantage of the special features of digital filters, the high degree of accuracy in filter realization, the unique structures possible, and the ease with which time-varying coefficients are incorporated, to obtain spectral characteristics not obtainable by analog devices, Rader and Gold (1967).

In the control literature, rarely does one find design techniques for digital control algorithms that take advantage of these features. Tou (1967) has examined the addition of real zeroes and poles in digital algorithms to shape the spectral characteristics of processes with no transport delay. The spectral characteristics of continuous proportional, derivative and integral controllers are well known. Many digital control

algorithms are designed to imitate continuous controllers. Tuning charts for continuous controllers are employed, with a pseudo delay of one-half the control interval, used to account for the effect of the sample and hold, Smith (1972).

The spectral characteristics of some discrete stochastic controllers will be examined in this chapter. As observed in Chapters 3 and 4, the synthesis of these controllers is not based explicitly on frequency domain considerations. It is of interest then to study their spectral properties. The manner in which these controllers compensate for transport delay will be examined and compared to more classical techniques of controlling deadtime.

6.2 Sampling of a Continuous Signal

When one samples a continuous signal at equispaced intervals of time, it is impossible to determine frequency components in the sampled signal whose period is less than twice the sampling interval. If the sampling interval is T seconds, the highest frequency we can detect in the sampled signal is $\omega_s/2$, where ω_s is $2\pi/T$ rad/sec. $\omega_s/2$ is commonly referred to as the Nyquist frequency. (The slowest frequency we can detect has a period of one-half the length of the sampled data record.) If the continuous signal has no frequencies greater than the Nyquist frequency, the spectrum of the continuous signal in the range $0 \leq \omega \leq \omega_s/2$ can be theoretically recovered from the spectrum of

the sampled signal. In practice, one's ability to reconstruct the spectrum of the input signal is affected by the resolution of the sampling device, and quantization of the computations, Rader and Gold (1967). When the input signal contains frequencies greater than the Nyquist frequency, one observes a "fold over" effect in the spectrum of the sampled signal. The spectrums of the sampled and continuous signal may differ appreciably. To obtain the spectrum of the input signal in the range $0 \leq \omega \leq \omega_s/2$ from the sampled signal, it is necessary to filter the continuous signal with an analog filter to remove the frequency components greater than the Nyquist frequency.

Time series models have been used by Box and Jenkins (1970) and Astrom (1970), to model process disturbances. These stochastic models can be viewed as a parametric representation of the spectral density function of the disturbances. If we have a "fold over" effect in our sampled signal (which can never be determined from the sampled signal alone) the stochastic model will not be a true representation of the process disturbances. As will be shown in the next section, the spectrum of a disturbance is easily computed if we have a model for it. The presence of a significant frequency component near the Nyquist frequency suggests that the continuous signal was not adequately filtered.

6.3 Frequency Response of Digital Filters

The response of a digital filter to an input $U(k)$ can frequently be described by linear difference equations of the form

$$Y(k) = \delta_1 Y(k-1) + \dots + \delta_r Y(k-r) + \omega_0 U(k) - \dots - \omega_s U(k-s) \quad (6.1)$$

If we define the backward shift operator z^{-1} , such that $z^{-b}Y(k) = Y(k-b)$, then (6.1) can be written more compactly as

$$Y(k) = H(z^{-1})U(k) = \frac{\omega_0 z^{-1} - \dots - \omega_s z^{-s}}{1 - \delta_1 z^{-1} - \dots - \delta_r z^{-r}} U(k) \quad (6.2)$$

Alternatively, we may write (6.2) as the convolution sum

$$Y(k) = \left(\sum_{n=0}^{\infty} h(n) z^{-n} \right) U(k) = \sum_{n=0}^{\infty} h(n) U(k-n) \quad (6.3)$$

if $H(z^{-1})$ can be expanded in a convergent series in z^{-1} . The particular representation we use is a matter of convenience.

To determine the frequency response of this filter, let the input $U(k)$ be the sampled cosine wave shown in Figure 6.0. At any of the equispaced points $k, k-1$, the input is given by

$$U(k) = \frac{A}{2} (e^{j\omega kT} + e^{-j\omega kT}) \quad (6.4)$$

A is the amplitude and ω the radian frequency of this periodic input. Substituting this into (6.3), we obtain

$$Y(k) = \frac{A}{2} (e^{-j\omega kT} \sum_{n=0}^{\infty} h(n) e^{j\omega nT} + e^{+j\omega kT} \sum_{n=0}^{\infty} h(n) e^{-j\omega nT}) \quad (6.5)$$

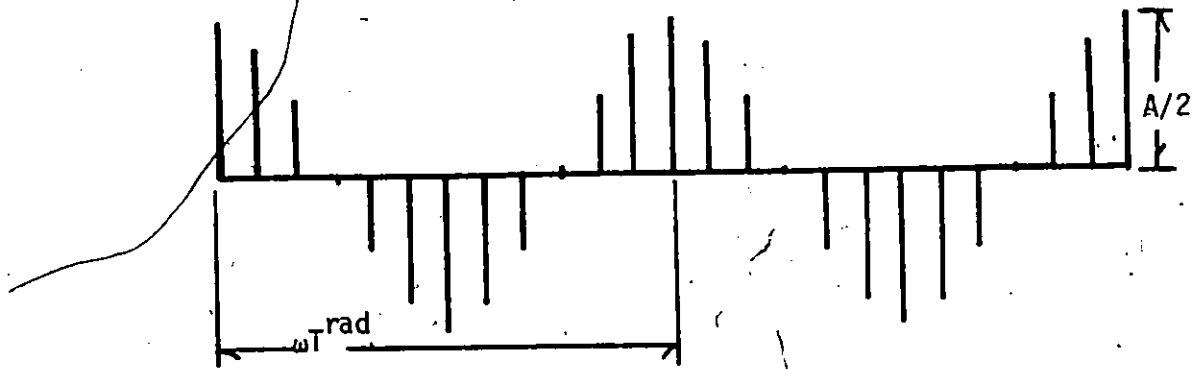


Figure 6.0: Sampled Cosine Wave

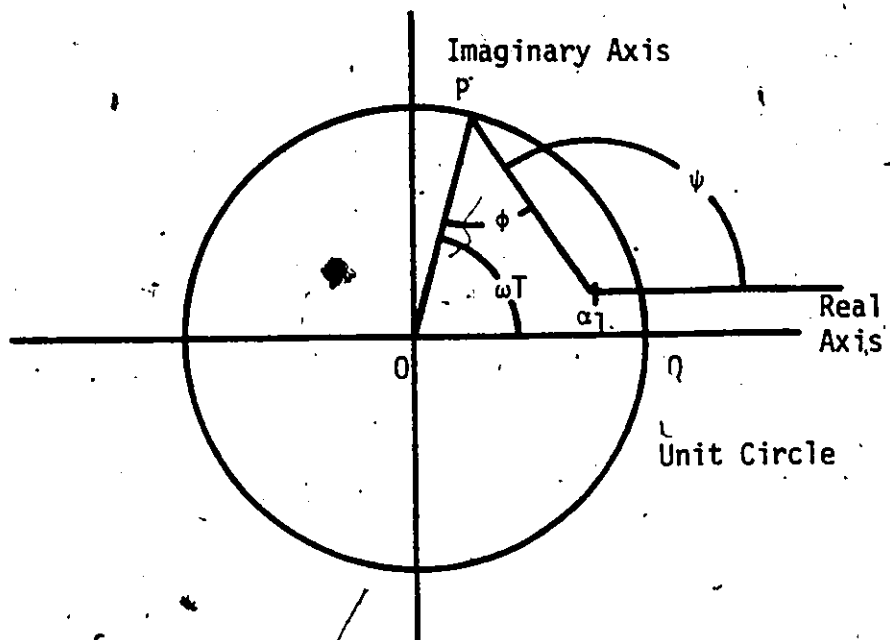


Figure 6.1: Frequency Response of the Filter
 $H(z) = 1 - \alpha_1 z^{-1}$

We recognize $Y(k)$ in (6.5) as the particular solution to the linear difference equation (6.1), that is, the solution after the transients due to the initial conditions of Y have died out. We might suspect that if the input is a periodic function, then $Y(k)$ is also a periodic function. This is stated in a number of references, Raggazini and Franklin (1958), Kuo (1967), Oppenheim and Schaffer (1975). To establish this, $Y(k)$, $k=1, 2, \dots$, is decomposed into its frequency components using the discrete Fourier transform. One finds that $Y(k)$ is a periodic function with the same radian frequency as the input. Thus $Y(k)$ can be represented as

$$Y(k) = Ar(e^{j(\omega k + \phi)T} + e^{-j(\omega k + \phi)T}) \quad (6.6)$$

where Ar is the magnitude and ϕ the phase shift of the output. Comparing (6.6) to (6.5) we identify

$$r = \left| \sum_{n=0}^{\infty} h(n)e^{-j\omega nT} \right| \quad (6.7)$$

and

$$\phi = -\arg \left(\sum_{n=0}^{\infty} h(n)e^{-j\omega nT} \right) \quad (6.8)$$

$|q|$ denotes the magnitude and $\arg(q)$ the phase angle of the complex number q . As we can see, the phase and gain characteristics can

be obtained by substituting $z^{-1} = e^{-j\omega T}$ in (6.2) or (6.3). Substitution of $z^{-1} = e^{-j\omega T}$ to determine the frequency response is only meaningful if the particular solution to (6.1) does not grow in time. A digital filter of the form

$$H(z^{-1}) = \prod_{i=1}^N \frac{1 - \alpha_i z^{-1}}{1 + \beta_i z^{-1}} \quad (6.9)$$

will be stable if all poles are inside or on the unit circle in z . This means that $|\beta_i| \leq 1$, $i=1, 2, \dots, N$.

Initially we introduced z^{-1} as a convenience to facilitate writing our filter equations. To evaluate the frequency response though, we let z^{-1} take on the complex value $e^{-j\omega T}$. It is usually assumed, Raggazini and Franklin (1958), Kuo (1967), that z^{-1} is the complex number e^{-ST} , where S is the complex argument associated with the Laplace transform. It must be emphasized that in these latter developments, it is assumed that we are sampling, at equispaced time intervals, a continuous function, which has a Laplace transform, Raggazini and Franklin (1958). In this case, it is assumed that the discrete difference equation (6.1) represents the sampled version of this underlying continuous process. If one wants to find the discrete equivalent of an analog filter, or the discrete equivalent of a continuous process, such an approach is appropriate. However, the representation (6.1) is justified in its own right as a discrete

filter, or process transfer function, without assuming the existence of an underlying continuous process. In this instance, z^{-1} does not represent a mapping of the complex variable S from the continuous domain, and it is incorrect to think of it as such.

6.3.1. Geometric Interpretation of Frequency Response

Consider the digital filter

$$H(z^{-1}) = 1 - \alpha_1 z^{-1} \quad (6.10)$$

We will allow α_1 to be complex valued. The frequency characteristics of this expression are obtained by evaluating (6.7) or (6.8). The geometric interpretation of (6.7) or (6.8) is shown in Figure 6.1. This is obtained by letting $z^{-1} = e^{-j\omega T}$ and using the de Moivre expansion for $e^{-j\omega T}$. The phase angle is $\psi - \omega T$, and the amplitude ratio is the distance PQ . The angle subtended by OPQ is $-(\psi - \omega T)$. For zeroes lying inside the unit circle, $|\psi - \omega T| \leq 90^\circ$. A resonance effect is observed as the zeroes approach the unit circle.

Digital filters may have zeroes lying outside the unit circle. In this case, the phase angle is not longer restricted to $\pm 90^\circ$. Care must be taken when the phase characteristics are determined mathematically as the phase angle is not confined to its principal values. In such cases, it is more convenient to factor the filter into two components. One term has all its zeroes inside

the unit circle, a minimum phase function, and the other term is known as an all pass filter. The latter has the following properties, Oppenheim and Schafer (1975),

- i) amplitude ratio is 1.0
- ii) phase angle < 0 , $0 \leq \omega T \leq \pi$

Consider the digital element $H(z^{-1})$ having a pair of complex conjugate zeroes lying outside the unit circle at $z = 1/q$ and $z = 1/q^*$ (the complex conjugate of $1/q$). By simple algebraic manipulation, $H(z^{-1})$ may be written as

$$\begin{aligned} H(z^{-1}) &= (1-qz^{-1})(1-q^*z^{-1}) \cdot \frac{(z-1/q)(z-1/q^*)}{(z-q)(z-q^*)} \\ &= H_{\min}(z^{-1}) \cdot H_{\text{ap}}(z^{-1}) \end{aligned} \quad (6.11)$$

where $H_{\min}(z^{-1}) = (1-qz^{-1})(1-q^*z^{-1})$. The all pass filter constructed from a pair of complex conjugate zeroes lying outside the unit circle is linear in the phase angle, Oppenheim and Schafer (1975). Thus, $H(z^{-1})$ has the same spectral representation as the filter

$$H_s(z^{-1}) = H_{\min}(z^{-1}) \cdot z^{-2} \quad (6.12)$$

The subscript s has been used to emphasize that we are considering equivalent spectral characteristics, as the transient responses of the filter $H(z^{-1})$ and $H_s(z^{-1})$ are much different.

If $H(z^{-1})$ is a polynomial of order N , and all N zeroes lie outside the unit circle, then $H(z^{-1})$ is referred to as a maximum phase function, Oppenheim and Schaffer (1975). $H(z^{-1})$ then has the equivalent spectral representation

$$H_s(z^{-1}) = z^{-N} \cdot H_{\min}(z^{-1}) \quad (6.13)$$

$H_{\min}(z^{-1})$ is formed by reflecting the zeroes of $H(z^{-1})$ through the unit circles. It will be seen that some stochastic controllers involve maximum phase function.

Consider the digital filter

$$H(z^{-1}) = (1 - (-.5 + .68j)z^{-1})(1 - (-.5 - .68j)z^{-1}) \quad (6.14)$$

The frequency response of this filter is shown in Figure 6.2. The resonance effect is due to the pair of complex conjugate zeroes, and this resonance becomes more pronounced as these zeroes approach the unit circle. Now consider the case where the zeroes are at their reflected values, $z = 1/(-.5 + .68j)$ and $z = 1/(-.5 - .68j)$. $H(z^{-1})$ has the same spectral characteristics as

$$H_s(z^{-1}) = z^{-2}(1 - (-.5 + .68j)z^{-1})(1 - (-.5 - .68j)z^{-1}) \quad (6.15)$$

The spectral characteristics of this filter are obtained by adding $-2\pi\omega T$ radians to the phase angle in Figure 6.2.

6.4 Process Dynamic and Disturbance Models

Many processes can be described by linear difference equations of the form

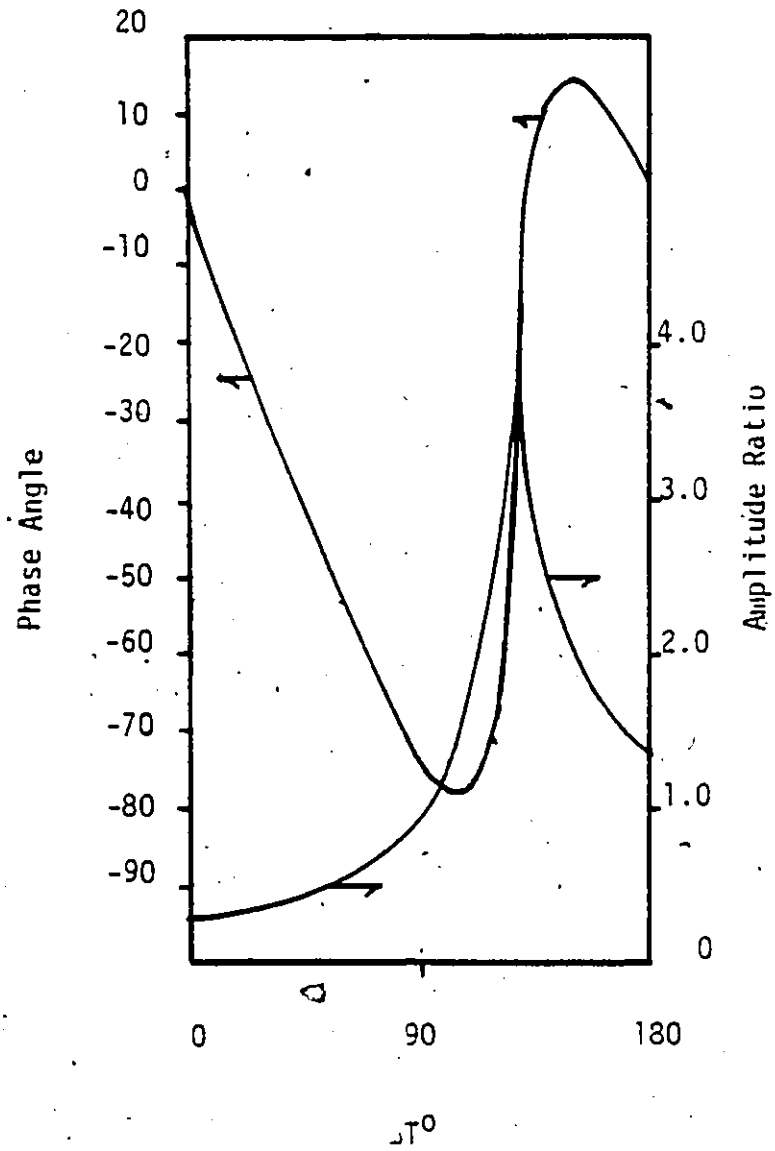


Figure 6.2: Frequency Response
of the Filter $H(z^{-1})$
 $= \frac{(1 - (-0.5 + 0.68j)z^{-1})}{(1 - (-0.5 - 0.68j)z^{-1})}$

$$Y(k) = \frac{\omega_0 - \omega_1 z^{-1} - \dots - \omega_s z^{-s}}{1 - \delta_1 z^{-1} - \dots - \delta_r z^{-r}} U(k-b) + N(k) \quad (6.16)$$

$Y(k)$ and $U(k)$ are deviations from their steady state values. $N(k)$ represents the total effect on the process output of all unobserved disturbances acting within the system, which in the absence of some compensating action would cause the process output to drift away from its target value. When these disturbances are of a random nature, they can be modelled by autoregressive moving average time series of the form

$$\begin{aligned} N(k) &= \frac{(1 - \theta_1 z^{-1} - \dots - \theta_q z^{-q}) a(k)}{\nabla^d (1 - \phi_1 z^{-1} - \dots - \phi_p z^{-p})} \\ &= \frac{\theta(z^{-1}) a(k)}{\nabla^d \phi(z^{-1})} \end{aligned} \quad (6.17)$$

The $a(k)$'s are a normally distributed sequence of independent random variables with mean 0 and variance σ_a^2 . The moving average term $\theta(z^{-1})$, and the autoregressive term $\phi(z^{-1})$ have all their roots inside the unit circle, in z , ∇ is a shorthand notation for $1 - z^{-1}$. Non-stationary disturbances are modelled with $d > 0$. The number of whole periods of delay is denoted by b and it has the value

$$b = 1 + \text{modulus} \left(\frac{\tau_d}{T} \right) \quad (6.18)$$

where τ_d is the process deadtime. Techniques for estimating the orders (r,s,b) , (p,d,q) and the parameters of the difference equations are discussed in Box and Jenkins (1970) and Astrom (1967, 1970).

6.5 Spectral Characteristics of Stochastic Controllers

The spectral characteristics of controllers designed to minimize the objective function

$$\begin{aligned}
 V &= \lim_{N \rightarrow \infty} \frac{1}{N} E \sum_{k=1}^N \{Y^2(k) + \lambda(\nabla^d U(k))^2\} \\
 &= \text{var } Y + \lambda \text{var } \nabla^d U
 \end{aligned} \tag{6.19}$$

will be examined in this section. The process model can be represented in state space form, MacGregor (1973), and dynamic programming used to find the optimal controller. Alternatively, the optimal controller can be found by spectral factorization of the covariance generating function, Wilson (1970). This latter approach gives far more insight into the manner in which these stochastic controllers regulate processes with deadtime. The frequency response is also more easily evaluated when the controller is expressed as an explicit function of the dynamic and stochastic model parameters. It is not intended to present an exhaustive analysis of these controllers with respect to stability, sensitivity and parameter variation. The conclusion drawn from such work tends to be very example oriented.

6.5.1 Unconstrained Controllers: $\lambda = 0$:

For processes where no additional constraints on the control action are required, the controller that minimizes (6.19) is given by Box and Jenkins (1970), Astrom (1970),

$$U(k) = \frac{-\delta(z^{-1})}{\omega(z^{-1})} \frac{T(z^{-1})}{\psi(z^{-1})} \cdot \frac{1}{\phi(z^{-1})\nabla^d} \cdot Y(k)$$

$$= -C(z^{-1})Y(k) \quad (6.20)$$

$T(z^{-1})$ and $\psi(z^{-1})$ are determined by factoring the disturbance transfer function as

$$\frac{\theta(z^{-1})}{\phi(z^{-1})\nabla^d} = \psi(z^{-1}) + \frac{T(z^{-1})z^{-b}}{\phi(z^{-1})\nabla^d} \quad (6.21)$$

Since $\psi(z^{-1})$ and $T(z^{-1})$ are functions only of z^{-i} , $i=0,1,2,\dots$, and not functions of z , they are uniquely determined. The expansion (6.21) has a physical interpretation. The expression

$$\hat{N}(k+b/k) = \frac{T(z^{-1})a(k)}{\phi(z^{-1})\nabla^d} \quad (6.22)$$

is a prediction for $N(k+b)$ given only information to time k . One could form many predictions for $N(k+b)$, but this one has the smallest prediction error among all possible predictions, Box and Jenkins (1970). The structure of this predictor can be determined from algebraic arguments, Box and Jenkins (1970),

Astrom (1970). $\hat{N}(k+b/k)$ in (6.22) is also the conditional mean for $N(k+b)$ given only information available to time k . Since the $a(k)$'s are normally distributed, $\hat{N}(k+b/k)$ is also the most probable estimate for $N(k+b)$. $\hat{N}(k+b/k)$ also corresponds to the complimentary solution of the difference equation (6.17) at origin k . The particular solution is given by $\psi(z^{-1})a(k+b)$ and is the error in predicting $N(k+b)$ given only information to time k . As we can see, there are a number of interpretations for the expansion in (6.21). These are discussed further in Chapter 8.

Stochastic disturbances affecting industrial processes are modelled often by low order ARIMA (p,d,q) models, Box and Jenkins (1970). Two common forms are the ARIMA $(0,1,1)$ and ARIMA $(1,1,0)$ model. For the former, it can be shown that

$$\frac{T(z^{-1})}{\psi(z^{-1})\nabla} = \frac{1-\theta}{1-(1-\theta)z^{-b}} \quad (6.23)$$

The expression $1-(1-\theta)z^{-b}$ has b roots z_k , $k = 0, 1, \dots, b-1$, located at

$$z_k = \sqrt[b]{1-\theta} e^{(2\pi j k/b)}, \quad k = 0, 1, \dots, b-1 \quad (6.24)$$

When $\theta < 0$, all roots are located outside the unit circle and $\psi(z^{-1})\nabla$ has the same spectral characteristics as

$$(\psi(z^{-1})\nabla)_s = e^{j\omega T b} (1 - \sqrt[b]{1-\theta} e^{-j\omega T b}), \quad 0 \leq j\omega \leq 2\pi \quad (6.25)$$

A negative value of θ indicates that the spectrum of $N(k)$ has more low frequency components than white noise. A positive value of θ gives rise to more frequency components near the Nyquist frequency.

When the ARIMA (1,1,0) model is expanded via (6.21), we cannot tell by inspection where the roots of $\psi(z^{-1})$ are located. However, bounds on these roots may be found. For convenience express $\psi(z^{-1})$ as

$$\psi(z^{-1}) = 1 + \psi_1 z^{-1} + \dots + \psi_{b-1} z^{-(b-1)} \quad (6.26)$$

If

$$\psi_{b-1} > \psi_{b-2} > \dots > \psi_1 > 1 \quad (6.27)$$

then all the roots of (6.26) lie outside the unit circle, Berezin and Zhidkow (1966). For ARIMA (1,1,0) disturbances with $0 < \phi < 1$, we can use this result to show that all the zeroes of $\psi(z^{-1})$ lie outside the unit circle. A positive value of θ indicates that the spectrum $N(k)$ has predominantly low frequency components. The autocorrelation function of this disturbance is positively correlated at low lags. We anticipate that for stochastic models of these structures, minimum variance controllers will have significant phase lead.

The forward transfer function is defined as

$$H(z^{-1}) = \frac{\omega^*(z^{-1})}{\delta^*(z^{-1})} z^{-b} \cdot C(z^{-1}) \quad (6.28)$$

The superscript * indicates that the transfer function used in the controller design may be slightly different than the true process transfer function due to modelling inaccuracies or changes in the process dynamics. Substituting the controller equation (6.20) into (6.28) we obtain

$$H(z^{-1}) = \frac{\omega^*(z^{-1})}{\delta^*(z^{-1})} \cdot \frac{\delta(z^{-1})}{\omega(z^{-1})} \cdot \frac{T(z^{-1})}{\psi(z^{-1})} \frac{z^{-b}}{\phi(z^{-1})^d} \quad (6.29)$$

The effect of the inverse of the process dynamics in the controller is to cancel the phase contribution of the process dynamics. If $\omega^*(z^{-1})$ and $\delta^*(z^{-1})$ have all their roots inside the unit circle in z , then small changes between the assumed and true process transfer function are not of concern. If, however, $\omega^*(z^{-1})$, or $\delta^*(z^{-1})$ have roots outside the unit circle, then the minimum variance strategy is sensitive to discrepancies between the assumed and true process transfer function, Astrom (1970), Kwakernaak and Sivan (1972). Strategies which are less sensitive to parameter variations can be obtained by using the objective function (6.19) with a non-zero value of λ . These controllers will be examined in the next section. Let us assume that $\omega^*(z^{-1}) = \omega(z^{-1})$, and $\delta^*(z^{-1}) \neq \delta(z^{-1})$. Equation (6.29) reduces to

$$H(z^{-1}) = \frac{z^{-b}T(z^{-1})}{\psi(z^{-1})\phi(z^{-1})\nabla^d} \quad (6.30)$$

The minimum variance controller results in a closed loop system which has all its poles inside the unit circle. One might suspect that the term $T(z^{-1})/(\psi(z^{-1})\phi(z^{-1})\nabla^d)$ in (6.30) contributes phase lead to cancel the lag introduced by the process deadtime.

To see this, let us consider an example. Let $N(k)$ be an ARIMA (0,1,1) process with $-1 < \theta < 0$. Using (6.23), the forward transfer function is given by

$$H(z^{-1}) = \frac{(1-\theta)z^{-b}}{1 - (1-\theta)z^{-b}} \quad (6.31)$$

For $-1 < \theta < 0$, $\psi(z^{-1})$ is unstable, i.e., has poles lying outside the unit circle in z ; although the closed loop is stable. The frequency response of $H(z^{-1})$ and the controller $C(z^{-1})$ is not strictly defined since the complementary solution of the associated difference equation grows in time. However, if we were to only consider the term $1 - (1-\theta)z^{-b}$, in (6.31), we would find that its frequency response is equivalent to

$$(1 - (1-\theta)z^{-b})_s = z^b(1 - \sqrt{1-\theta} z^{-b}) \quad (6.32)$$

The right hand side of (6.32) has tremendous phased advance in the region $0 \leq \omega T < \pi$, due to the term z^b . Thus, in a heuristic manner, we can see how $\psi(z^{-1})$ introduces phase lead to cancel the phase lag arising from the transport delay.

Not all stochastic models result in a factorization for $\psi(z^{-1})$ that has all its roots outside the unit circle. In these instances, the phase lead contribution of $\psi(z^{-1})$ is smaller than when the zeroes lie outside the unit circle. It must be emphasized though that the stochastic controllers are designed to compensate for the inherent process disturbances modelled by $N(k)$. If the disturbance were a sampled cosine wave, then the resulting controller would give adequate phase lead. The two disturbance models considered in the examples are typical of those encountered on industrial process, Box and Jenkins (1970). A particular choice of the parameter values could result in a slowly drifting disturbance, not unlike a periodic input. It is perhaps not surprising then that the minimum variance controllers give good phase lead for these disturbances.

6.5.2 Constrained Controllers $\lambda \neq 0$:

It may happen that the variance of the input $\nabla^d U(k)$ for the unconstrained controller is too large. In this case, one seeks to find a controller for which the variances of $Y(k)$ and $\nabla^d U(k)$ are jointly acceptable. (Although the controller $C(z^{-1})$ is unstable, the variance of $\nabla^d U(k)$ is finite when $\omega(z^{-1})$ has no zeroes lying outside the unit circle. To calculate the variance of $\nabla^d U(k)$, the controller must be expressed in terms of the $a(k)$'s and not the Y 's. When this is done, the term $\psi(z^{-1})$ does not appear in controller transfer function.) One solution is to

cast the dynamic and stochastic model into state space form and find the control policy by solving the appropriate Riccati equations. In this formulation, the control law is expressed as a linear combination of states, which may have no physical interpretation. However, more insight into the structure of the controllers is obtained when the solution is found by spectral factorization of the covariance generating function. The manipulated variable is now expressed as a linear combination of past inputs and outputs.

The spectral factorization solution to (6.19) is outlined by Wilson (1970). It is based on the work of Weiner (1949) and Whittle (1963). The following operations are required.

(a) express

$$\frac{\theta(z^{-1})}{\nabla^d \phi(z^{-1})} = \psi(z^{-1}) + \frac{z^{-b} \cdot \Gamma(z^{-1})}{\nabla^d \phi(z^{-1})} \quad (6.33)$$

to obtain $\psi(z^{-1})$ and $\Gamma(z^{-1})$.

(b) form $\gamma(z^{-1})$ from

$$\begin{aligned} \gamma(z)\gamma(z^{-1}) &= \omega(z^{-1})\omega(z) \\ &+ \lambda(1-z^{-1})^d \delta(z^{-1})(1-z)^d \delta(z) \end{aligned} \quad (6.34)$$

(c) express

$$\frac{\omega T(z^{-1})}{\nabla^d \phi(z^{-1}) \gamma(z)} = \frac{Q_1(z^{-1})}{\nabla^d \phi(z^{-1})} + \frac{Q_2(z)}{\gamma(z)} \quad (6.35)$$

to obtain $Q_1(z^{-1})$.

The control that minimizes the objection function (6.19) is then given by

$$\nabla^d U(k) = \frac{\delta(z^{-1}) \nabla^d Q_1(z^{-1})}{\omega(z^{-1}) Q_1(z^{-1}) z^{-b} + \gamma(z^{-1}) \theta(z^{-1})} Y(k) \quad (6.36)$$

The denominator in (6.36) has ∇^d as a factor. Certain precautions must be taken to insure that the factorizations are unique, and these are outlined by Wilson (1970).

To examine the spectral characteristics of this controller, we suppose that the process is described by

$$Y(k) = \frac{U(k-3)}{1-.92z^{-1}} + \frac{1-.4z^{-1}}{\nabla} a(k) \quad (6.37)$$

The variance of Y and ∇U are shown in Table 6.1 as a function of λ . There is a significant decrease in the variance of ∇U with little increase in the variance of Y . However, the spectral characteristics of the controller will change significantly as λ is increased. This can be explained with reference to equation (6.36). When $\lambda = 0$, the denominator contains $\psi(z^{-1})$ as a factor. This maximum phase function contributes tremendous phase lead.

λ	var Y	var U
0	4.92	3.55
.05	4.94	2.52
.10	4.98	2.00
.15	5.02	1.70
.20	5.06	1.45
.25	5.10	1.32
.40	5.19	1.01

Table 6.1: The Effect of Constraining on the Variance of Y and ∇U .

As the variance of $\nabla^d U(k)$ is constrained, we eventually reach a point where the denominator in (6.36) is no longer a maximum phase function. At this point, the controller will no longer give significant phase advance.

The large reduction in the variance of ∇U , with correspondingly little increase in the variance of y can be explained by examining the spectral density function. If the control equation is of the form

$$\nabla U(k) = \frac{L_1(z^{-1})}{L_2(z^{-1})} a(k) \quad (6.38)$$

then the variance of $\nabla U(k)$ is given by, Astrom (1970)

$$\text{var} \{ \nabla U(k) \} = \frac{1}{\pi} \int_0^{\pi} \phi_{uu}(\omega T) d\omega T \quad (6.39)$$

where the spectral density function $\phi_{uu}(\omega T)$ is defined as

$$\phi_{uu}(\omega T) = \left| \frac{L_1(e^{-j\omega T})}{L_2(e^{-j\omega T})} \right|^2 \quad (6.40)$$

The spectral density function of the controller (6.36) for the process (6.37) is shown in Figure 6.3 for $\lambda = 0$ and $\lambda = .2$. The largest contribution to the variance of ∇U , for $\lambda = 0$, is from the high frequency components. This occurs because the minimum variance controller tries to cancel the forecast of the disturbance in three control intervals.

Even if the unconstrained variance of ∇U were acceptable, there is still incentive to examine the spectral density function of ∇U for various levels of constraining. In the example just considered, the effect of constraining the variance of ∇U is to reduce the influence of high frequency components in the controller. This is important as the controller is sensitive to any high frequency noise that might enter the loop.

By contrast, the variance is dominated by low frequency components, Figure 6.4. The effect of constraining the variance ∇U , has little effect on changing the spectral density function at these low frequencies.

The dynamic and stochastic models are seldom exact. By preventing large changes in the manipulated variable, we reduce the effect of process nonlinearities that might be encountered when the operating variables are far removed from their mean value. From (6.36) we can see that there is less reliance on

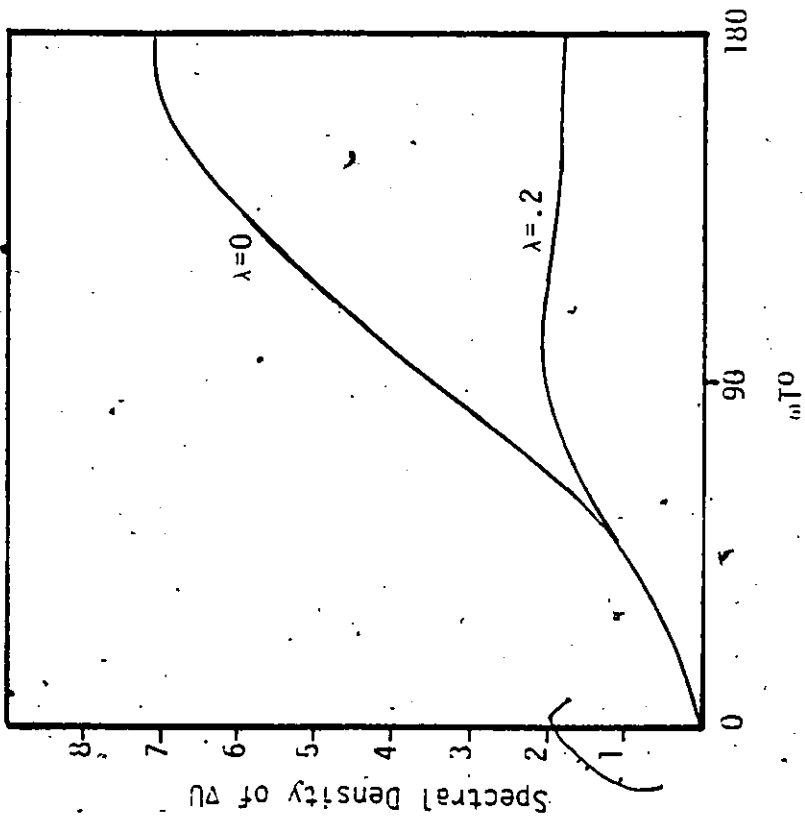


Figure 6.3: Spectral Density Function of VU for Various Levels of Constraining

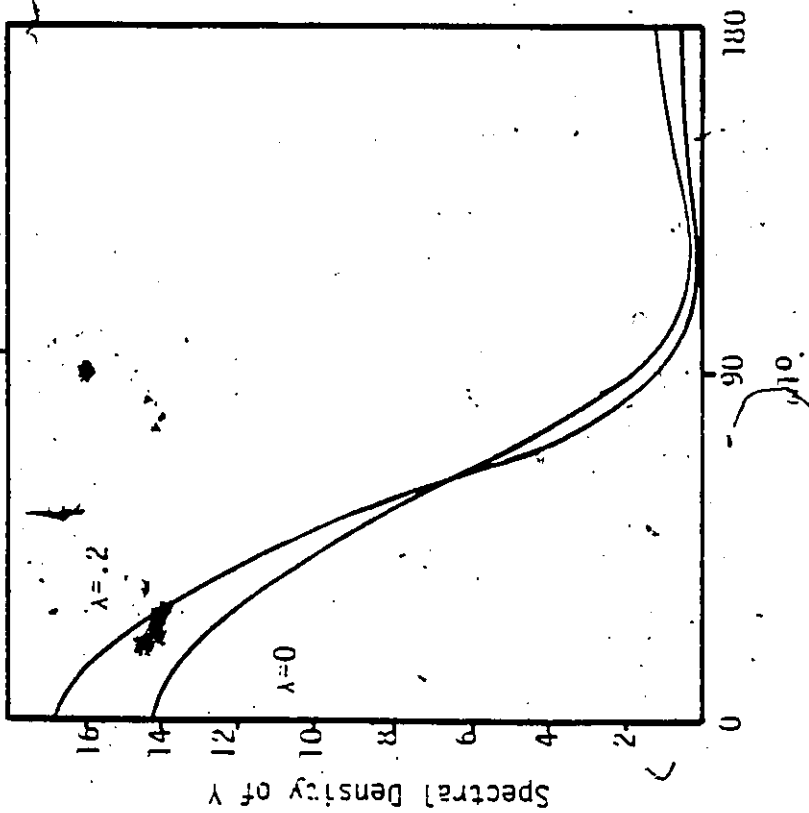


Figure 6.4: Spectral Density Function of Y for Various Levels of Constraining

the inverse of the process dynamics. Constraining the manipulated variable essentially desensitizes the control scheme to inaccuracies in the process model.

6.5.3 Other Design Techniques

There are numerous filter design techniques proposed in the communications literature, Oppenheim and Schaffer (1975), to minimize expressions of the form

$$v_3 = \frac{1}{\pi} \int_0^{\pi} W(j\omega) (|H(e^{-j\omega T})| - |H_d(e^{-j\omega T})|) P_d(j\omega T) \quad (6.41)$$

$H_d(e^{-j\omega T})$ is the desired spectral shape, and $W(j\omega)$ is a weighting function specified by the designer. These techniques could be used to design controllers to give the closed loop specific spectral characteristics. To use these techniques, one must have a model for the process dynamics and disturbances.

6.6 Stochastic Controllers and Discrete Otto Smith Regulators

A technique proposed to compensate for process deadtimes is the Smith regulator, Smith (1957). The frequency domain properties of continuous Smith regulators have been analyzed by Astrom (1977). The difficulty in implementing a continuous Smith compensator is the requirement to simulate a pure delay. This is easily accomplished though, with a digital algorithm.

Consider the block diagram shown in Figure 6.5. The control element $C(z^{-1})$ is designed to control the process as if

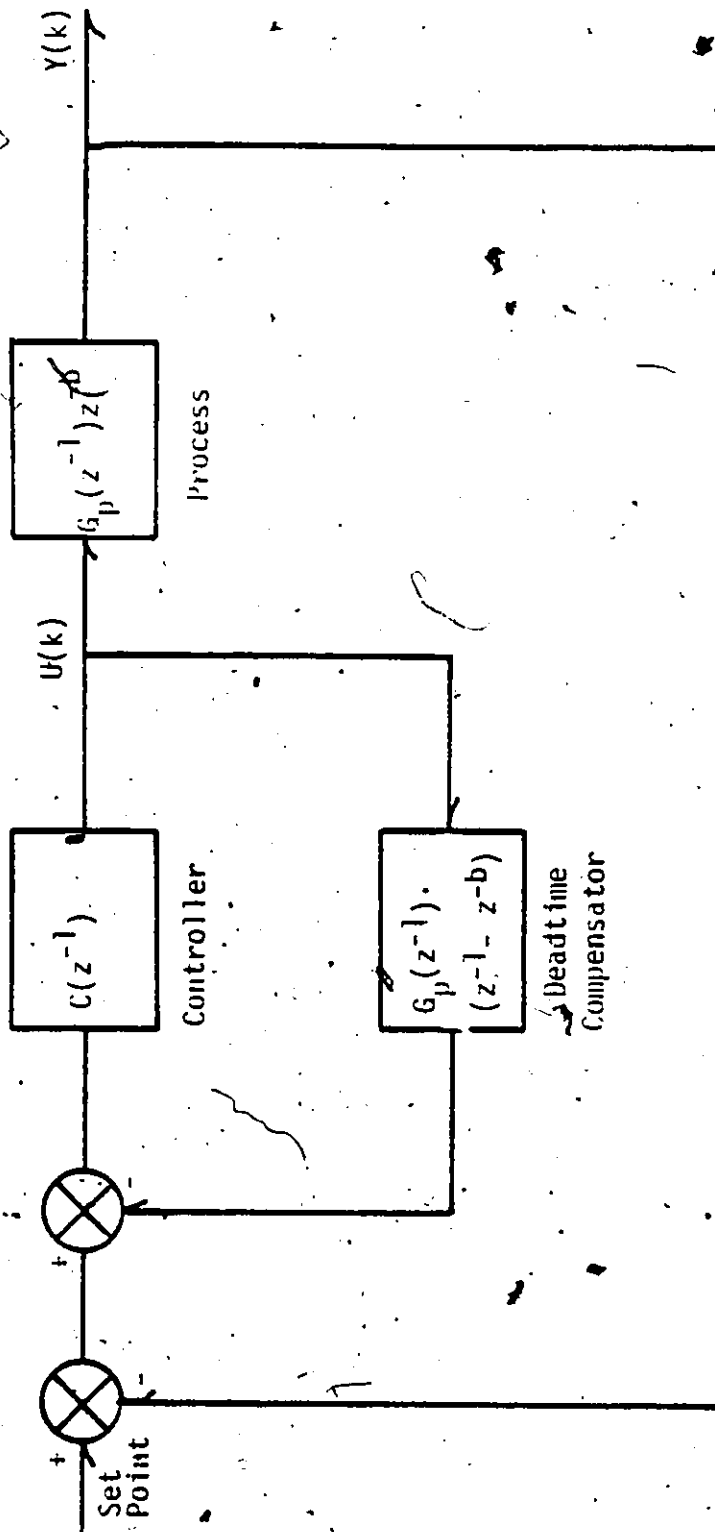


Figure 6.5: Switch Deadtime Compensator

there were no transport delay. We remember that in a digital system there is at least one period of delay. If $C(z^{-1})$ were chosen as

$$G_d(z^{-1}) = \frac{\omega(z^{-1})}{\delta(z^{-1})} \cdot (z^{-1} - z^{-b}) \quad (6.42)$$

the closed loop transfer function from the set point to the output would be

$$G_c(z^{-1}) = \frac{C(z^{-1})G_p(z^{-1})z^{-b}}{1 + G_p(z^{-1})C(z^{-1})z^{-1}} \quad (6.43)$$

where $G_p(z^{-1}) = \omega(z^{-1})/\delta(z^{-1})$. The effect of the process deadtime has apparently been removed from the characteristic equation and should not cause stability problems. The controller in Figure 6.5, may be written as

$$U(k) = - \frac{C(z^{-1})}{1 + \frac{\omega(z^{-1})C(z^{-1})(z^{-1} - z^{-b})}{\delta(z^{-1})}} Y(k) \quad (6.44)$$

Are there any stochastic controllers which have this structure? Consider an IMA (1,1) disturbance. For this model an unconstrained minimum variance controller for $b=1$ is

$$U(k) = -C(z^{-1})Y(k) = \frac{-\omega(z^{-1})}{\delta(z^{-1})} \frac{(1-\theta)}{\gamma} Y(k) \quad (6.45)$$

Substituting this expression for $C(z^{-1})$ in (6.44) we obtain

$$\nabla U(k) = \frac{-\delta(z^{-1})}{\omega(z^{-1})} \frac{(1-\theta)}{1 + (1-\theta)z^{-1} + \dots + (1-\theta)z^{-b}} Y(k) \quad (6.46)$$

We recognize (6.46) as the minimum variance controller for an IMA(1,1) disturbance and a process dynamic model with b whole periods of delay. A Smith compensator is an unconstrained minimum variance controller for an IMA(1,1) disturbance. The special property of this class of disturbances is that the 1 step ahead forecast and $\hat{N}(k+b/k)$ are identical. This is the only class of ARIMA(p,d,q) models for which this is true.

The manner in which stochastic controllers and Smith compensators handle deadtime is very similar. Both involve making a forecast of the future behaviour of the process output. The difference between the methods is the manner in which the forecasts are made. General ARIMA disturbances have more structure to the forecast and therefore lead to slightly different deadtime compensation. The Smith compensator, although removing the delay from the characteristic equation, is not optimal in the sense of minimizing the variance of Y .

The phase characteristics of a discrete Smith compensator were examined in section 6.4.1. If $-1 < \theta < 0$, the controller has the inverse of a maximum phase function which contributes a significant phase lead up to the Nyquist frequency. The phase characteristics of continuous Smith compensators have been

analyzed by Astrom (1977). For the examples considered in that paper, the controller also gave significant phase advance. However, the controllers were stable, unlike the example considered in section 6.4.1.

6.7 Summary

In this chapter, the mathematics required to analyze the frequency response of digital control algorithms was presented. A particular emphasis was placed on the computation of the frequency response of maximum phase functions. The spectral characteristics of some stochastic controllers were analyzed. The structure and parameter values of the stochastic models primarily determines the phase characteristics of these controllers. For a particular class of stochastic models, it was shown that the stochastic controller has significant phase advance. The use of frequency response techniques complements the time domain analyses of these controllers and should give greater understanding as to how they control processes with deadtime. Finally, a brief analysis of the Smith regulator showed that it is a minimum variance controller for a particular choice of stochastic disturbance model. As will be seen in Chapter 8, this stochastic model can be thought of as a randomly occurring deterministic step type disturbance.

CHAPTER 7

SELF-TUNING AND ADAPTIVE CONTROLLERS

7.1 Introduction

Of many modern control theories, one which has had a great impact on the process industries is that of optimal stochastic control. The basic theory is outlined in books by Box and Jenkins (1970) and Astrom (1970). A number of steps are usually involved in the development of such control schemes:

- (i) plant experimentation to collect data on the system,
- (ii) identification of a suitable dynamic-stochastic model and
- (iii) design of the controller using this model.

Often one of the main benefits from performing such an exercise comes from improved understanding gained about the behaviour of the process and the nature of its disturbances. However, as a procedure for developing a controller, the above steps can be time consuming and demand a level of expertise sometimes not available in industry. Furthermore, if the process changes with time (for example, due to production rate changes, catalyst activity changes in a reactor, etc) then the process model parameters must be periodically reestimated and the controller parameters readjusted.

The attempt to overcome these latter difficulties led to the idea of self-tuning controllers, that is control schemes as depicted in Figure 7.1 whereby the controller is coupled with an

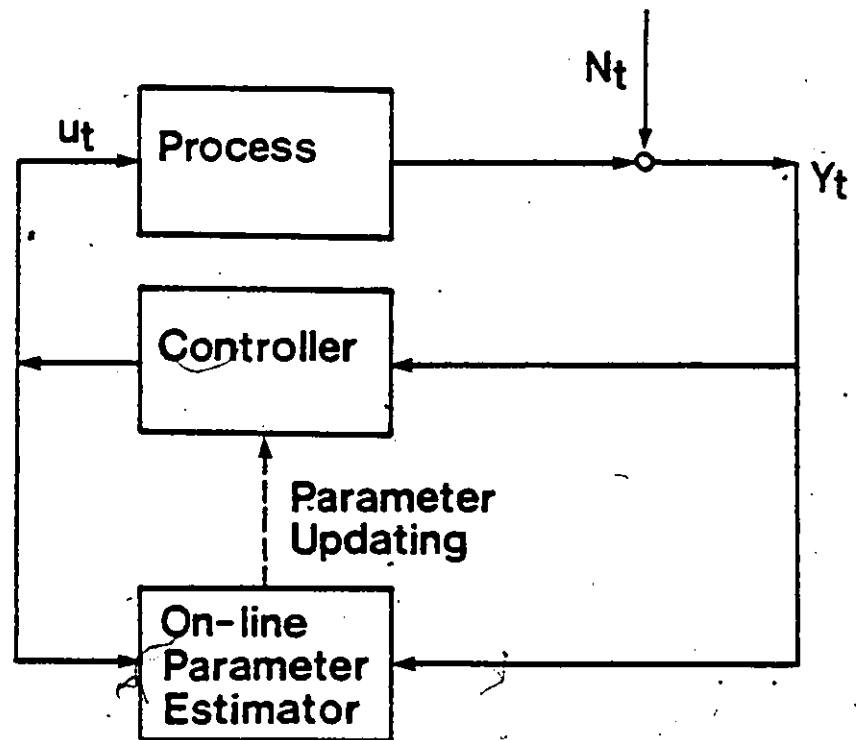


Figure 7.1: Adaptive Control Scheme

on-line estimator which is capable of using the input-output information on the process to tune the controller directly. Much of the theory on these regulators was developed by Astrom and Wittenmark (1973) and their coworkers, Wittenmark (1973), Astrom (1974), Astrom et al (1977), and Ljung and Wittenmark (1974a, 1974b). A number of industrial applications have been reported; on paper machines, Cegrell and Hedqvist (1975), Wittenmark (1974), an ore crusher, Borisson and Syding (1976), and T_iO_2 kiln Dumont and Belanger (1978). Its rapid acceptance has no doubt been due to its simplicity and robustness, and to the ease with which it can be implemented.

In this chapter, a discussion unifying much of the theory relevant to industrial applications is presented, and it is applied in various forms to the control of a highly exothermic packed bed tubular reactor carrying out the hydrogenolysis of n-butane.

7.2 Self-Tuning Control

Assume that the dynamics of the process to be controlled can be described by a linear transfer function model and that the effect of the stochastic disturbances on the output can be represented by an autoregressive-integrated-moving-average (ARIMA) model

$$Y(k+b) = \frac{\omega(z^{-1})}{\delta(z^{-1})} U(k) + N(k+b) \quad (7.1)$$

$$N(k+b) = \frac{\theta(z^{-1})}{\phi(z^{-1})\nabla^d} a(k+b) \quad (7.2)$$

$Y(k)$ is the output deviation from its target value at time k and $U(k)$ is the input deviation from a corresponding steady-state value. The terms $\omega(z^{-1})$ and $\delta(z^{-1})$ in the discrete transfer function represent polynomials of order s and r , respectively, in the backward shift operator, z^{-1} , and b (≥ 1) represents the number of whole periods of time delay in the system. The stochastic disturbance $N(k)$ represents the total effect at the output of all unobserved disturbances acting within the system, and in the absence of some compensating action would lead to deviations of $Y(k)$ from its target value. In the ARIMA model ($a(k)$) is a white noise sequence, the moving-average numerator $\theta(z^{-1})$ is a polynomial with all roots lying inside the unit circle in the z -plane, and the autoregressive denominator $\phi(z^{-1})\nabla^d$ is a polynomial of order $(p+d)$ with p roots inside the unit circle and d roots equal to unity. ($\nabla = (1-z^{-1})$ is the backwards difference operator.) The allowance for d roots equal to unity (usually $d = 0$ or 1) enables one to model the type of non-stationary (stochastic and deterministic) disturbances often found to occur in practice. As will be seen, a value of $d > 0$ always leads naturally to integral action in the controller.

7.2.1 Minimum Variance Control with Recursive Least Squares Estimation

The disturbance model (7.2) can be reexpressed as

$$\begin{aligned}
 N(k+b) &= \frac{\theta(z^{-1})}{\phi(z^{-1})\nabla^d} a(k+b) \\
 &= \psi(z^{-1})a(k+b) + \frac{T(z^{-1})}{\phi(z^{-1})\nabla^d} \hat{a}(k)
 \end{aligned} \quad (7.3)$$

where $\psi(z^{-1})$ is a polynomial of order $b-1$. The first term on the right contains only the effect of future disturbances and represents the minimum variance forecast error at time k , whereas the second term represents the forecast. Multiplying (7.1) by $\psi(z^{-1})\phi(z^{-1})\nabla^d \delta(z^{-1})$ and substituting (7.3) into the result yields

$$\begin{aligned}
 \delta(z^{-1})\theta(z^{-1})(Y(k+b) - \psi(z^{-1})a(k+b)) &= \delta(z^{-1})T(z^{-1})Y(k) \\
 &+ \omega(z^{-1})\psi(z^{-1})\phi(z^{-1})\nabla^d U(k)
 \end{aligned} \quad (7.4)$$

which is of the form

$$\delta(z^{-1})\theta(z^{-1})(Y(k+b) - \epsilon(k+b)) = \alpha(z^{-1})Y(k) + \beta(z^{-1})\nabla^d U(k) \quad (7.5)$$

where

$$\alpha(z^{-1}) = \alpha_0 + \alpha_1 z^{-1} + \dots + \alpha_{m_0} z^{-m_0} \quad (7.6)$$

$$\beta(z^{-1}) = \beta_0 + \beta_1 z^{-1} + \dots + \beta_{\ell_0} z^{-\ell_0} \quad (7.7)$$

and $\epsilon(k+b)$ is the b -step ahead forecast error $\psi(z^{-1})a(k+b)$.

The orders m_0 and ℓ_0 are

$$m_0 = r + \max(q-b, p+d-1) \quad (7.8)$$

$$\ell_0 = s + p + b - 1 \quad (7.9)$$

If the parameters of (7.5) were known, the control action

$$\nabla^d U(k) = \frac{-\alpha(z^{-1})}{\beta(z^{-1})} Y(k) = \frac{-\delta(z^{-1})}{\omega(z^{-1})} \cdot \frac{\Gamma(z^{-1})}{\psi(z^{-1})\phi(z^{-1})} Y(k) \quad (7.10)$$

would minimize $E(Y^2(k+b))$.

Suppose that $\delta(z^{-1})\theta(z^{-1}) = 1.0$. It is seen from (7.5) that the process output is now expressed as a direct function of the minimum variance controller parameters. If these parameters were unknown one might try to identify them from a model of the form

$$Y(k+b) = \alpha(z^{-1})Y(k) + \beta(z^{-1})\nabla^d U(k) + \epsilon(k+b) \quad (7.11)$$

where

$$\alpha(z^{-1}) = \alpha_0 + \alpha_1 z^{-1} + \dots + \alpha_m z^{-m} \quad (7.12)$$

and

$$\beta(z^{-1}) = \beta_0 + \beta_1 z^{-1} + \dots + \beta_\ell z^{-\ell} \quad (7.13)$$

The number of whole periods of delay (b) and the order of the pole (d) lying on the unit circle in the disturbance model are presumed known. Equation (7.11) admits linear least squares estimation which may be expressed recursively. The estimates of these controller parameters will be unbiased since $\epsilon(k+b) = \psi(z^{-1})a(k+b)$ is uncorrelated with the regressor variables ($Y(k), Y(k-1), \dots, \nabla^d U(k-1), \dots$) but they may not be efficient estimates because of the lack of independence of the $\epsilon(k)$'s. The parameters can be updated at every sampling interval and used in the control law

$$\nabla^d U(k) = \frac{-\hat{\alpha}(z^{-1})}{\hat{\beta}(z^{-1})} Y(k) \quad (7.14)$$

as if they were exactly known. Note that one is not trying to identify the process dynamic and stochastic model parameters, but only those combinations which appear in the minimum variance controller.

The restriction that $\delta(z^{-1})\theta(z^{-1}) = 1.0$ appears to limit the usefulness of this scheme. However, Astrom and Wittenmark (1973) proved two theorems which show that this algorithm has desirable asymptotic properties irrespective of the product $\delta(z^{-1})\theta(z^{-1})$. If the parameter estimates converge then the scheme reduces certain autocorrelations of the output and certain

cross-correlations between the input and output to zero. Furthermore, if the correct minimum variance controller structure is used, the parameters will converge to those of the true minimum variance controller. This latter result was somewhat surprising (for $\delta(z^{-1})\theta(z^{-1}) \neq 1$.) since the estimation equation (7.5) is no longer a function only of the minimum variance controller parameters. However, the following analysis gives an intuitive explanation as to why this is reasonable and why in fact the process may behave as though it were being generated by a model of the form (7.11) even though $\delta(z^{-1})\theta(z^{-1}) \neq 1$.

The system (7.5) may be written as

$$Y(k+b) = (1 + \mu_1 z^{-1} + \mu_2 z^{-2} + \dots)(\alpha(z^{-1})Y(k) + \beta(z^{-1})\nabla^d U(k)) + \epsilon(k+b) \quad (7.15)$$

where

$$(\delta(z^{-1})\theta(z^{-1}))^{-1} = (1 + \mu_1 z^{-1} + \mu_2 z^{-2} + \dots) \quad (7.16)$$

The controller parameters are estimated from the regression equation (7.11) and the control action (7.14) is taken at every interval. If the control (7.14) is therefore substituted into (7.15) then

$$\begin{aligned}
Y(k+b) &= \alpha(z^{-1})Y(k) + \beta(z^{-1})\nabla^d U(k) \\
&+ \mu_1 \left\{ \alpha(z^{-1})Y(k-1) - \frac{\hat{\alpha}(k-1, z^{-1})}{\hat{\beta}(k-1, z^{-1})} \beta(z^{-1})Y(k-1) \right\} \\
&+ \mu_2 \left\{ \alpha(z^{-1})Y(k-2) \left[\frac{\hat{\alpha}(k-2, z^{-1})}{\hat{\beta}(k-2, z^{-1})} \beta(z^{-1})Y(k-2) \right] \right\} \\
&+ \dots + \varepsilon(k+b)
\end{aligned} \tag{7.17}$$

Note that upon convergence of the parameters the extra terms in the square brackets tend to zero due to the control action (7.14) being taken at every interval. Furthermore, if the process is open loop stable (i.e., $\delta(z^{-1})$ has all its roots lying inside the unit circle) then $\mu(z^{-1})$ is a convergent series in z^{-1} and the μ_i weights decrease with increasing i , thereby further reducing the influence of the extra terms in the curly brackets. The process output therefore is well approximated by the regression model (7.11).

Extension of this theory to include feedforward control is straightforward, Wittenmark (1973), Cegrell and Hedqvist (1975) and Astrom et al (1977).

Most of the properties of these regulators are conditional upon convergence of the parameter estimates. Ljung and Wittenmark (1974a) have shown that a set of deterministic non-linear differential equations can be associated with the recursive

estimation and control equations. The solution of these differential equations gives the expected trajectories of the parameter estimates, and proving convergence of the estimation algorithm is equivalent to proving stability of the differential equations. Although convergence is not assured in all cases, simulations and process applications have shown that good control is usually achieved fairly rapidly even if the parameter estimates have not reached their final values.

7.2.2 Constrained Input Control and Non-minimum Phase Processes

Minimum variance control sometimes calls for excessively large variations in the manipulated variable, particularly when the control interval is chosen to be very short relative to the major time constant. This results from its attempt to cancel out the entire effect of the forecasted disturbance over b control intervals. In situations where this manipulation is too severe, it is common practice to calculate control laws which minimize the variance of Y subject to a constraint on the variance of $\nabla^d U(k)$, that is, to minimize

$$I_1 = E\{Y(k)^2 + \zeta^2 (\nabla^d U(k))^2\} \quad (7.18)$$

Note that if the disturbances are non-stationary ($d=1$) it is the variance of $\nabla^d U(k)$ that must be constrained since that of $U(k)$ is indefinitely large. The calculations required to solve this

linear quadratic control problems are well known and require spectral factorization, Wilson (1970), Whittle (1963), or the steady-state solution of a matrix Riccati equation, Astrom (1970). Self-tuning versions of these controllers, Astrom (1974), involve the use of an efficient recursive estimation method (preferably recursive maximum likelihood), Soderstrom et al (1978), to update the estimates of the process transfer function and stochastic noise model parameters at each interval followed by the on-line solution of the discrete spectral factorization problem or solution of the equivalent Riccati equations. Non-minimum phase systems (that is, systems with discrete models having roots of z^{-1} lying outside the unit circle) such as the chemical reactor treated in this thesis, are also conveniently handled by minimizing the criterion (7.18) with $\zeta'' > 0$. For $\zeta'' = 0$, one obtains the controller which has minimum variance among all controllers with finite variance for $\nabla^d U(k)$ as opposed to the minimum variance algorithm which would give a controller with infinite variance for $\nabla^d U(k)$. The spectral factorization solution in this case yields a relatively simple set of algebraic identities, Peterka (1972). For $\zeta'' > 0$, the solution is somewhat more involved as we saw in the previous chapter.

An alternative and much simpler approach to constraining the variations in the manipulated variable has been proposed by Clarke and Hastings-James (1971) and Clarke and Gawthrop (1975):

Rather than minimizing (7.18), they treat the simpler problem, MacGregor and Tidwell (1978), of minimizing an instantaneous index

$$I_2 = \{Y(k+b/k) + \zeta'(\nabla^d U(k))^2\} \quad (7.19)$$

where $Y(k+b/k)$ is the minimum variance forecast of $Y(k+b)$ made at time k . This criterion usually results in controllers which, for the same constraint on the variance of $\nabla^d U(k)$, have only slightly larger variance for the output, MacGregor and Tidwell (1977).

The resulting controller, however, can be expressed as a direct expression of the process model parameters and the constraining parameter ζ' . By showing that minimizing (7.19) is equivalent to minimizing

$$I_3 = E\{Y(k+b) + \zeta \nabla^d U(k)\}^2 = E\{\phi(k+b)\}^2 \quad (7.20)$$

where $\zeta = \zeta'/\omega_0$, Clarke and Gawthrop (1975) were able, by a direct extension of Astrom and Wittenmark's work, to develop a self-tuning controller to minimize (7.20). Using the definition of $\phi(k+b)$ from (7.20) the system equation (7.4) may be expressed as

$$\delta(z^{-1})\theta(z^{-1})\{\phi(k+b) - \varepsilon(k+b)\} = \alpha(z^{-1})Y(k) + (\beta(z^{-1}) + \zeta\delta(z^{-1})\theta(z^{-1}))\nabla^d U(k)$$

(7.21)

where $\alpha(z^{-1})$ and $\beta(z^{-1})$ are previously defined. If the parameters were known, the control action

$$\gamma(z^{-1})\nabla^d U(k) = -\alpha(z^{-1})Y(k) \quad (7.22)$$

would minimize (7.20) where $\gamma(z^{-1}) = \{\beta(z^{-1}) + \zeta\delta(z^{-1})\theta(z^{-1})\}$.

By direct analogy a self-tuning algorithm with the estimation model

$$\phi(k+b) = \alpha(z^{-1})Y(k) + \gamma(z^{-1})\nabla^d U(k) + \varepsilon(k+b) \quad (7.23)$$

and the control equation (7.22) will have the same properties as derived by Astrom and Wittenmark for the minimum variance case.

The order of $\gamma(z^{-1})$ may be larger than that of $\beta(z^{-1})$ in the minimum variance case, and if the process dynamic-stochastic models are unknown, the value of ζ that will reduce the variance of $\nabla^d U(k)$ by a given amount will be unknown, but can easily be found by on-line tuning. Note that $\zeta = \zeta'/\omega_0$ and since ζ' is positive, the sign of ζ must be that of ω_0 . Another point to note is that unlike ζ'' in the criterion (7.18) a finite, non-zero value of ζ is needed to ensure that one does not have parametric sensitivity or instability for non-minimum phase systems.

7.3 Application to the Control of a Catalytic-Chemical Reactor

The control of catalytic packed bed tubular reactors carrying out highly exothermic gas phase reactions represents one of the most challenging control problems in the chemical industry.

These reactors are highly non-linear, distributed parameter, multivariable processes and often exhibit non-minimum phase characteristics. Although it is usually desired to achieve control over the final conversion and product selectivities, on-line measurement of these via gas chromatographs, etc. is often too slow to be of any use in a direct feedback control scheme. Therefore, in the control of industrial reactors one usually relies on temperature measurements throughout the catalyst bed; a common situation being simple feedback from some measure of the hot-spot (maximum) temperature in the bed. This provides an indirect stabilization of the exit conversion and safeguards against dangerous temperature excursions. For more direct control over the exit conversion and product selectivities one must rely upon some form of inferential control based on complex multivariable models of the reactor (see, for example, Jutan et al (1977a,b,c), Wong (1977) and Wright et al (1977)). Furthermore, due to the extreme non-linearities in the process, either non-linear control schemes or else linear adaptive schemes, Tremblay and Wright (1977), are desirable if the operating region is to be changed, or, as is often true, the catalyst activity changes over time.

A schematic diagram of the butane hydrogenolysis reactor is shown in Figure 7.2. It consists of a single 4 cm diameter tube 28 cm long packed with finely divided catalyst particles (nickel on silica gel). The reactor wall temperature is controlled by the

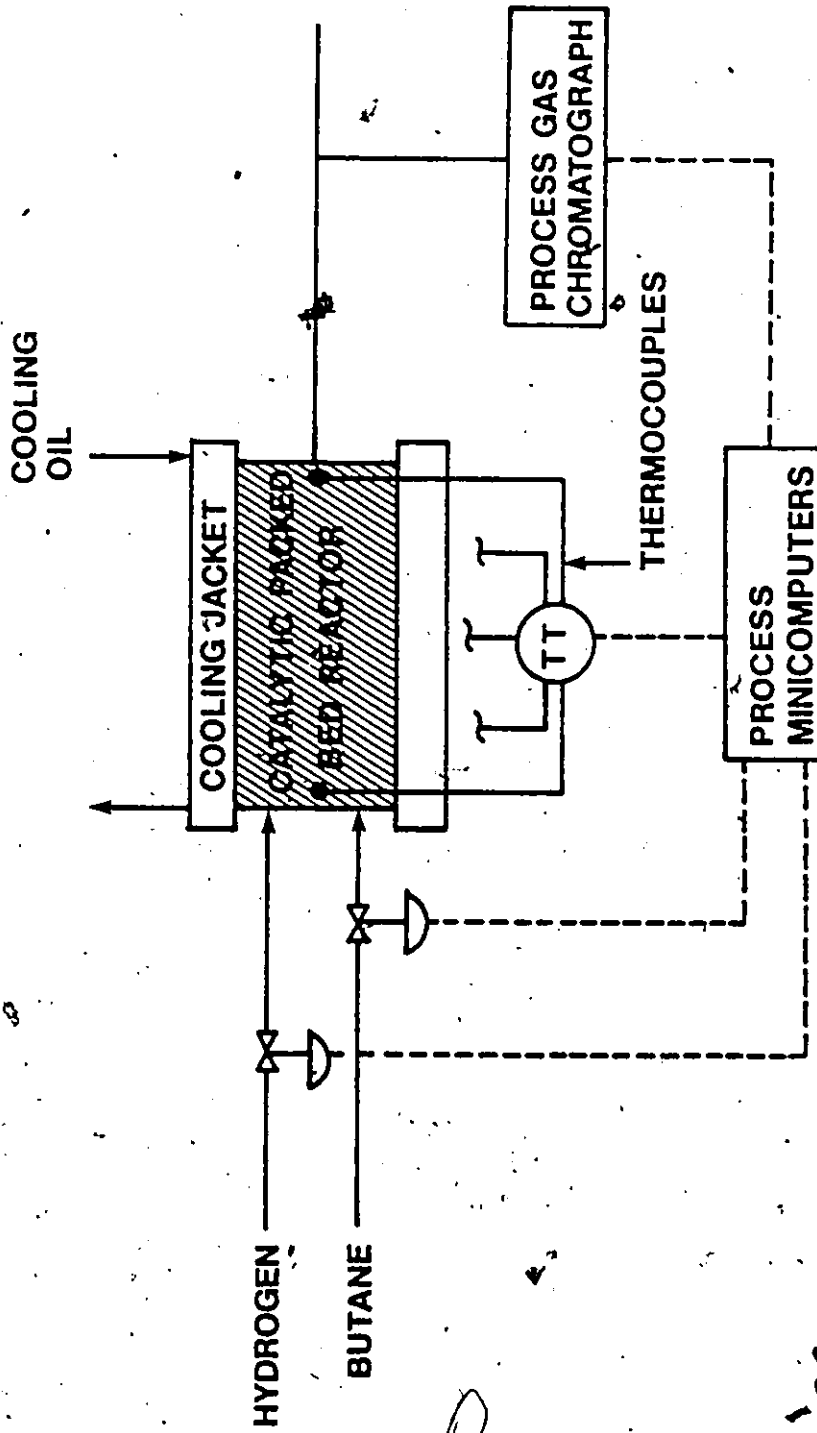


Figure 7.2: Reactor Control Configuration

countercurrent flow of heat transfer oil through the annulus of a cooling jacket. The flow rates of the two feed streams, hydrogen and butane, are controlled by a direct digital control algorithm once per second, and the self-tuning regulator manipulates the set-point of the butane flow loop on a 48 second interval. Temperatures measured at equispaced points along the central axis provide the primary measurements for control purposes. Exit concentration measurements obtained every 6 minutes using an on-line process chromatograph are too infrequent for direct use in a feedback control scheme. The reactor has been observed to have temperature runaways or to quench in a matter of minutes if left uncontrolled. Details of the reactor design, control software package and process computer interfacing are available elsewhere, Tremblay (1977).

7.3.1 Minimum Variance and Constrained Control of the Hot Spot Temperature: Run 1.

Preliminary investigations into the feasibility of using a self-tuning regulator (STR) for reactor control involved univariate control of the reactor hot spot temperature by manipulating the butane flow rate. No attempt was made to control the position of the hot spot, only its magnitude as determined from the thermocouple signals. The hot spot temperature used in the estimation and control algorithm was smoothed by a first order digital filter from temperature data collected every twelve seconds.

In Run 1, the estimation model (7.23) was used with $\ell = m = 3$ and $b = d = 1$, that is

$$\begin{aligned}\phi(k+1) &= \alpha_0 Y(k) + \alpha_1 Y(k-1) + \alpha_2 Y(k-2) + \beta_0 \nabla U(k) \\ &+ \beta_1 \nabla U(k-1) + \beta_2 \nabla U(k-2) + \epsilon(k+1) \\ &= \underline{x}'(k) \underline{\theta} + \epsilon(k+1)\end{aligned}\quad (7.24)$$

where

$$\phi(k+1) = Y(k+1) + \zeta \nabla U(k) \quad (7.25)$$

$$\underline{\theta}' = (\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2) \quad (7.26)$$

and

$$\underline{x}'(k) = (Y(k), Y(k-1), Y(k-2), \nabla U(k), \nabla U(k-1), \nabla U(k-2)) \quad (7.27)$$

The orders ℓ , m , b and d were chosen by postulating a structure for a dynamic-stochastic model that would account for previously observed process response although no preliminary identification was actually carried out.

The controller parameter vector $\underline{\theta}$ was estimated using the exponentially discounted recursive least squares algorithm, Soderstrom et.al (1978).

$$\underline{\theta}(k) = \underline{\theta}(k-1) + \underline{K}(k)(\phi(k) - \underline{x}'(k-1)\underline{\theta}(k-1)) \quad (7.28)$$

where

$$\underline{K}(k) = \underline{P}(k-1)\underline{x}(k-1) / (\lambda + \underline{x}'(k-1)\underline{P}(k-1)\underline{x}(k-1)) \quad (7.29)$$

and

$$\underline{P}(k) = \frac{1}{\lambda} \{ \underline{P}(k-1) - \underline{P}(k-1)\underline{x}(k-1)\underline{x}'(k-1)\underline{P}(k-1) \} \quad (7.30)$$

$$(\lambda + \underline{x}'(k-1)\underline{P}(k-1)\underline{x}(k-1))$$

and where $\lambda (\leq 1.0)$ is the exponential discounting factor in the least squares function

$$\sum_{s=1}^t \lambda^{t-s} \epsilon^2(s+b) \quad (7.31)$$

The results of one control run are summarized in Figures 7.3, 7.4, and 7.5, which have been divided into sections for ease of discussion. Each section represents a change in controller operation and will be discussed in order. Above each section in Figure 7.3 are some summary statistics and a plot of the estimated autocorrelation function for the data of that section.

Section A: Proportional-Integral Control The performance of a digital proportional plus integral (PI) algorithm with a control interval of 48 seconds provided the basis of comparison for the self-tuning regulator. Because of the inverse response characteristics of the reactor (at sampling intervals less than about 30 seconds), more frequent action did not appear to improve

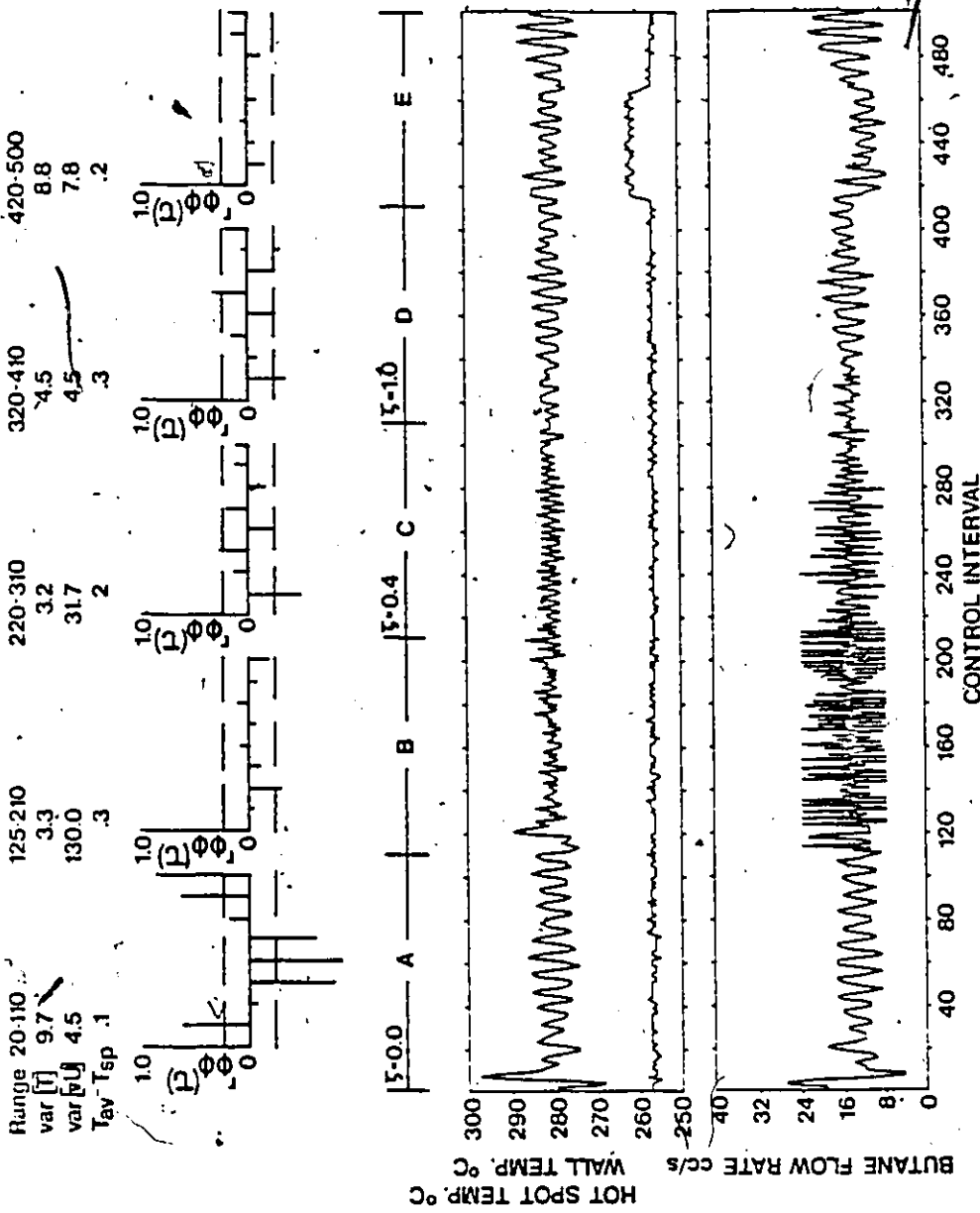


Figure 7.3: Data from Reactor Run 1

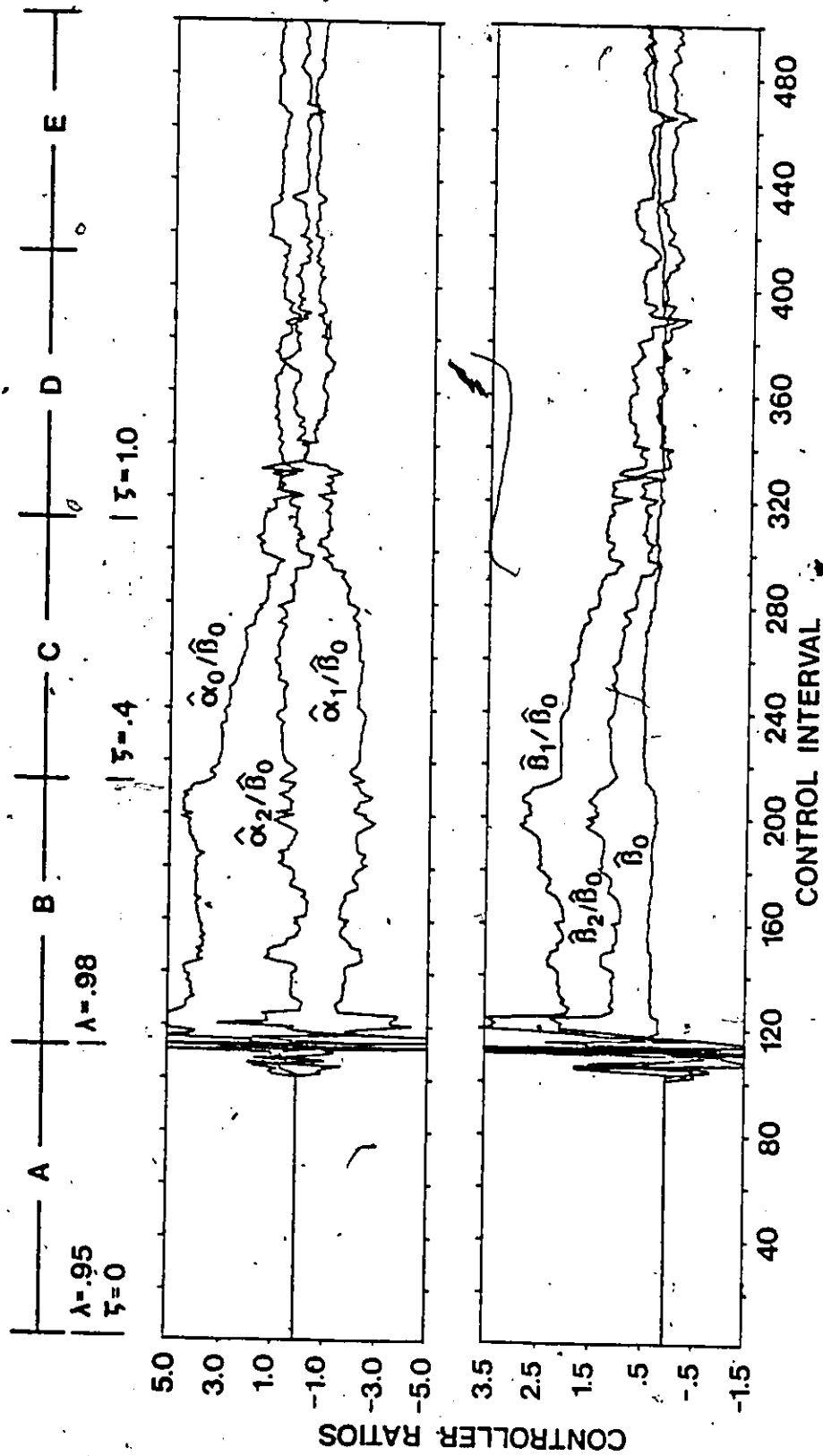


Figure 7.4: Parameter Ratios for Reactor Run 1

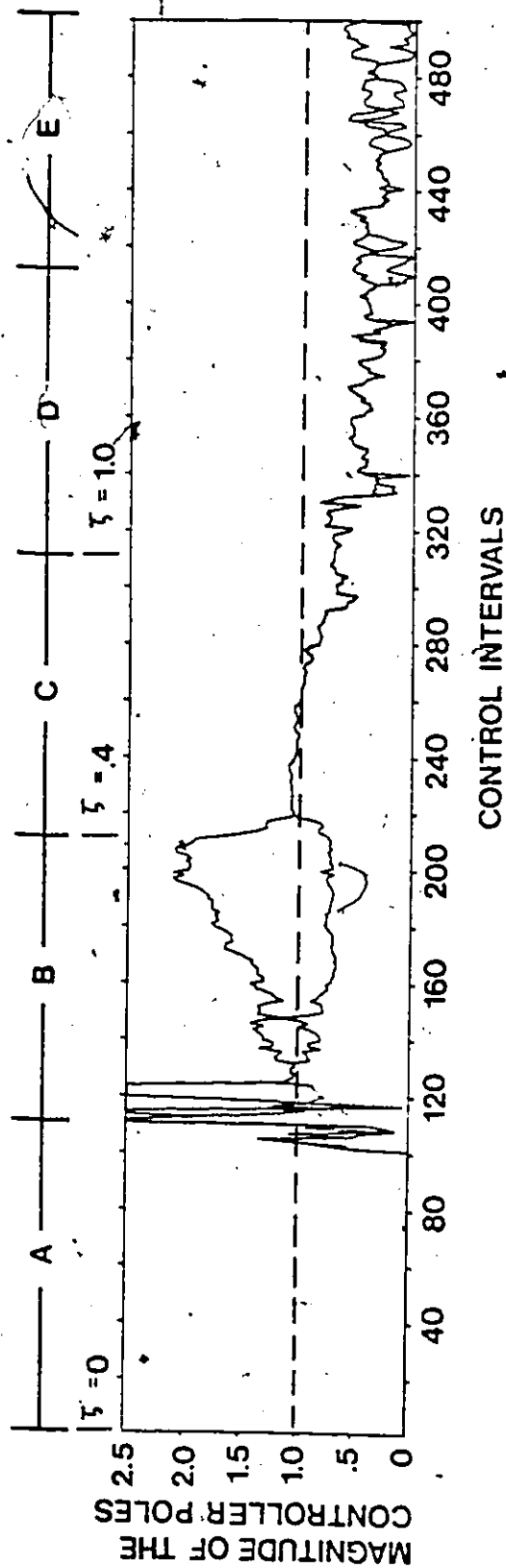


Figure 7.5: Controller Poles for Reactor Run 1.

the control. The response of the PI algorithm is shown in section A for 100 intervals. It should be mentioned that a considerable effort was required to tune this controller. The flow rate of the butane feed was constrained to the range 7 to 25 cm³/s throughout most of the run. It is obvious from the controller performance and the plot of the estimated auto-correlation function for the hot spot temperature that the controller is not optimal in a minimum variance sense.

Section B: Minimum Variance STR: At sample number 101, the estimation of the parameters in (7.24) was started, but with the PI controller still operative. Although this activation of the estimation algorithm prior to that of the STR controller ensures that the parameter estimates are non-zero when first used to compute the control signal, it is not recommended that this period extend too long since one will not be converging to the correct values. (This is evident upon substituting the fixed PI controller equation into (7.15) and comparing with (7.17).) The initial conditions for the estimation algorithm were $P(0) = 1000I$, $\underline{\theta}(0) = \underline{0}$, $\lambda = 0.95$ and $\zeta = 0.0$.

At sample number 111, the minimum variance STR ($\zeta = 0$) was started using the estimated parameters available at that time. (The value of the estimation discounting parameter λ was also changed to 0.98). The transition over from PI control to the STR is reasonably smooth. The hot spot temperature variance was very

quickly reduced to a level consistent with measurement error. ($\sigma^2 = 30^\circ\text{C}$) and the estimates of the controller parameters appear to have almost converged to steady values within about 20 intervals. The autocorrelation function of Y and the approximate 95% confidence bands (for white noise, Box and Jenkins (1970)) were calculated for the samples in the interval 125 to 210 and are shown above Section B in Figure 7.3. For optimality, 95% of the sample autocorrelation should lie within these limits. This plot reveals that the parameter estimates appear to have converged to the minimum variance controller parameters. Only one sample autocorrelation value lies near the confidence limit and a χ^2 -test on these indicates that the hot spot temperature sequence in this period is essentially white noise.

However, this control over the hot spot temperature calls for large alternating changes in the manipulated butane feedrate (see Figure 7.3). This is due to roots of the minimum variance controller lying outside the unit circle. (See the plot of the magnitude of these roots in section B of Figure 7.4.) Since $b=1$ (that is, less than one whole period of process dead time), we can say that this is the result of the discrete transfer function model being non-invertible. The control over the hot spot temperature remains nearly optimal in spite of the bang-bang type action on the manipulated butane flow rate, because limits are imposed on the flow rate actually allowed. However, these

excessive variations in the butane flow rate upset the production rate of the desired products of reaction (propane and ethane) and therefore are unacceptable.

Sections C and D: Clarke's Constrained STR: In order to reduce these variations in the manipulated variable Clarke's algorithm (equations 7.22 and 7.23) which minimizes the performance index in (7.19) was used. This change is readily accomplished on-line by simply changing the value of the constraining parameter ζ from zero to a finite value. At the beginning of Section C (sample number 211) the value of ζ was set equal to 0.4. The controller parameter estimates $\underline{\gamma}$ (Figure 7.4) make an immediate change (from those of $\underline{\beta}$) resulting in a corresponding move of the controller roots (Figure 7.5) towards new locations inside the unit circle. The variance of the manipulated flow rate is immediately reduced while that of the hot spot temperature remains almost unchanged. Note that the pattern of the hot spot variations has changed somewhat, being more autocorrelated than before. The large autocorrelation estimate of ϕ at lag one in the plot above Section C (Figure 7.3) tends to indicate that the parameter estimates have not converged in the period of Section C. A more rigorous analysis for convergence in the case of exponentially discounted least squares is treated in (7.8) and reveals that this is indeed the case. This is also apparent from the continued slow drift in the estimates after the initial more rapid change. This slow convergence was due to the

continued use of relatively slow updating ($\lambda = 0.98$) in the estimation algorithm over this period. In the equivalent period of an almost identical run the value of λ was changed on-line to 0.95 at the same time as ζ was changed to 0.4. The \underline{y} parameters converged much more rapidly to their new values. (See Figure 7.6 and compare with Section C of Figure 7.4).

In Section D, starting at sample number 311, the constraint on the manipulated butane feed rate was increased by setting $\zeta = 1.0$. This effectively pulls the roots of the controller further inside the unit circle (Figure 7.5), and reduces the variability in the butane flowrate further but at the expense of increased variation in the hot spot temperature. Again, the discounting factor ($\lambda = 0.98$) was not changed in this section.

Section E: A Step Disturbance in Reactor Wall Temperature:

The same control algorithm from the preceding section ($\zeta = 1.0$) was continued throughout Section E but at sample number 411 an extreme step load disturbance of 5°C in the reactor wall temperature was introduced and maintained for 50 intervals. The STR handles this load change extremely well with no visible upset nor offset in the hot spot temperature. Furthermore, the plot of the sample autocorrelation function $r_{\phi\phi}(k)$ over Section E (Figure 7.3) reveals that the controller is optimal for performance index I_2 or I_3 in this period. An identical load disturbance during PI control resulted in severe temperature

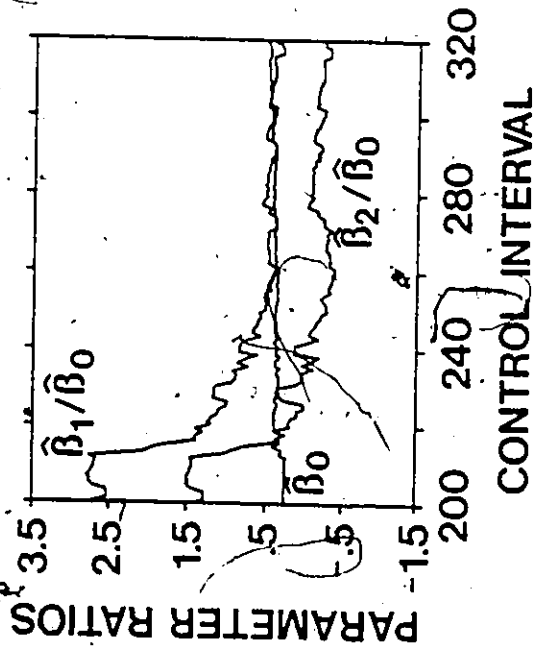


Figure 7.6: Clarke's Algorithm - An Alternative Run

excursions to over 315°C and very slow recovery. In still other runs, Jutan et al (1977c), the PI controller was seen to quench the reaction during such an upset.

Comments: For the constrained controller ($\epsilon = 0.4$), the variance of the hot spot temperature is almost identical to that of the unconstrained controller (see the summary statistics above Figure 7.3). This was not surprising in this application since, by constraining the manipulated input, we put less stress on the accuracy of the linear model (i.e., reduce parameter sensitivity) and minimize the effect of the severe non-linearities in the reactor.

Section F: Operation at Extreme Conditions: At a later point in this experiment, the set point of the hot spot temperature was stepped over 50 control intervals to 295°C, and held there for 50 control intervals. The 40 degree Celcius rise above the wall temperature in an extremely difficult region in which to operate the reactor. From Figure 7.6, we can see that the STR is able to stabilize the reactor in this region. We note the nonlinear behaviour of the hot spot temperature. After 50 control intervals at this new set point, the control algorithm is switched to the PI algorithm. The oscillations in the PI algorithm are much larger than the STR. Since the STR is able to stabilize the process in this region, it would

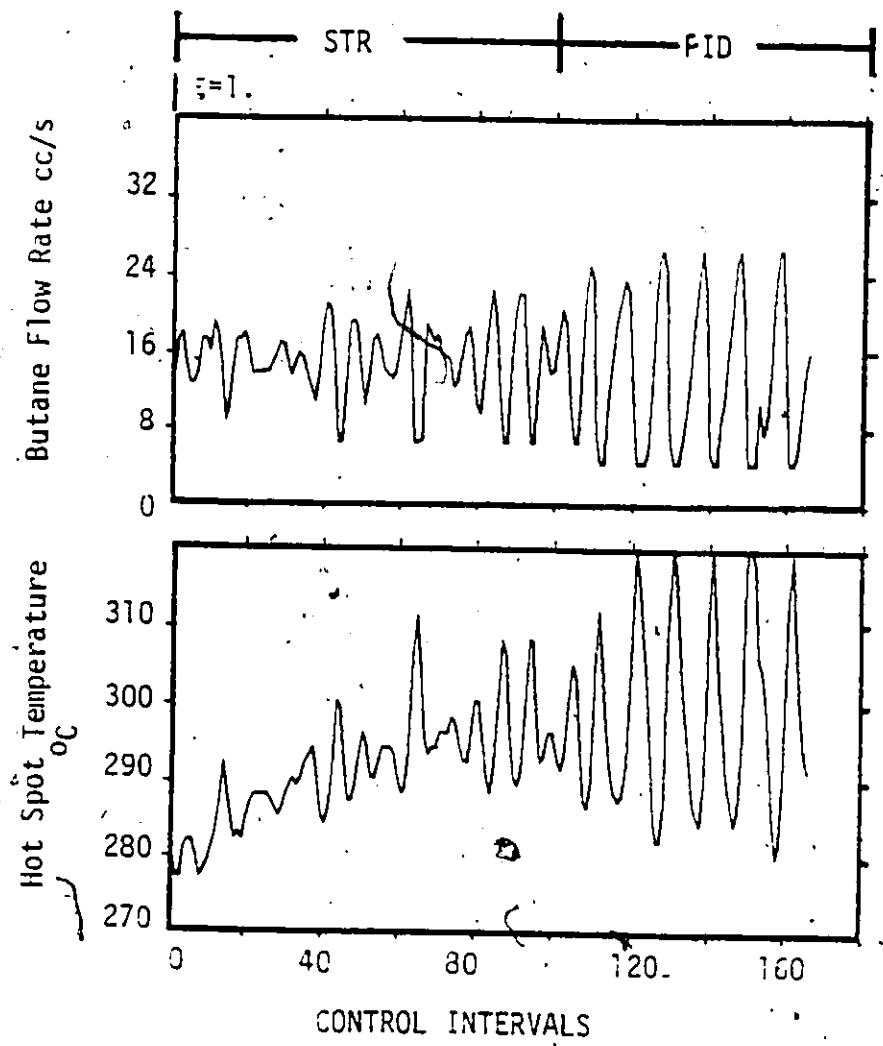


Figure 7.7: Operation at Extreme Conditions

now be possible to collect operating data in order to identify a complete dynamic/stochastic model of the process, Wong (1977), or estimate transport parameters for a mechanistic model, Jutan et al (1977c).

7.2.2 Run 2: Minimum Variance and Constrained Control of the Hot Spot Temperature: Run 2

The estimation model used in the previous section was again used in this run. The set point of the hot spot temperature was reduced from 280°C to 275°C , and the set point of the oil reduced from 250°C to 248°C . The results of this run are summarized in Figure 7.8, 7.9 and 7.10. The preliminary identification period was five intervals, Figure 7.8. Good control over the hot spot temperature was obtained very quickly, Section B, Figure 7.8. As in run 1, this occurred at the expense of large changes in the manipulated variable. At control interval 140, the discounting factor was changed from .98 to .95. Twenty intervals later the constraining factor was increased from 0 to .2. The controlled parameters immediately shifted to new values, Figure 7.9, Section C. The poles of the controller were also brought inside the unit circle very quickly, Figure 7.10, Section C. At sample number 205, the constraining factor was further increased to .4. As in Run 1, there was very little increase in the variance of the hot spot temperature as the input manipulations were constrained. In

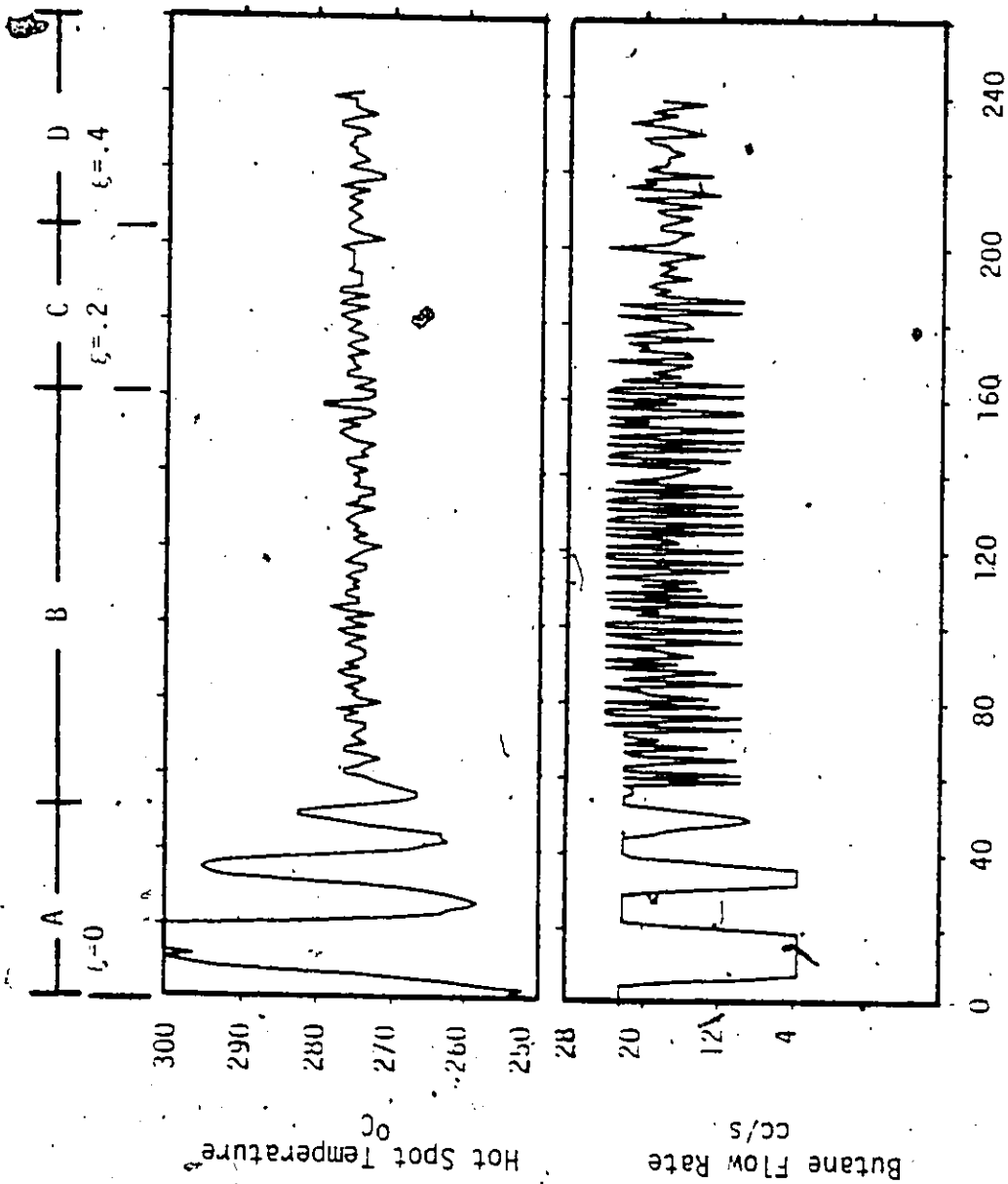


Figure 7.8: Data from Reactor Run 2

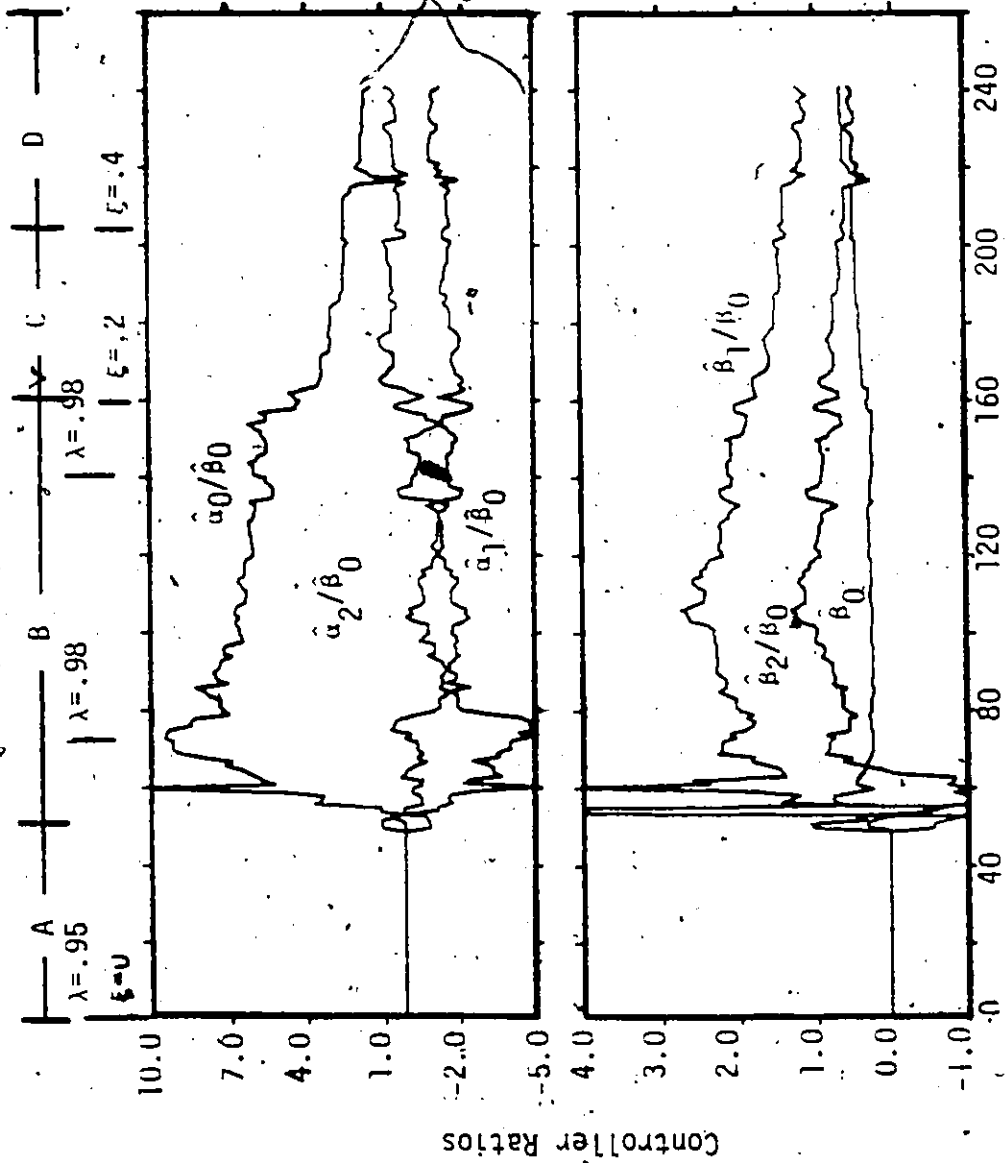


Figure 7.9: Parameter Ratios for Reactor Run 2

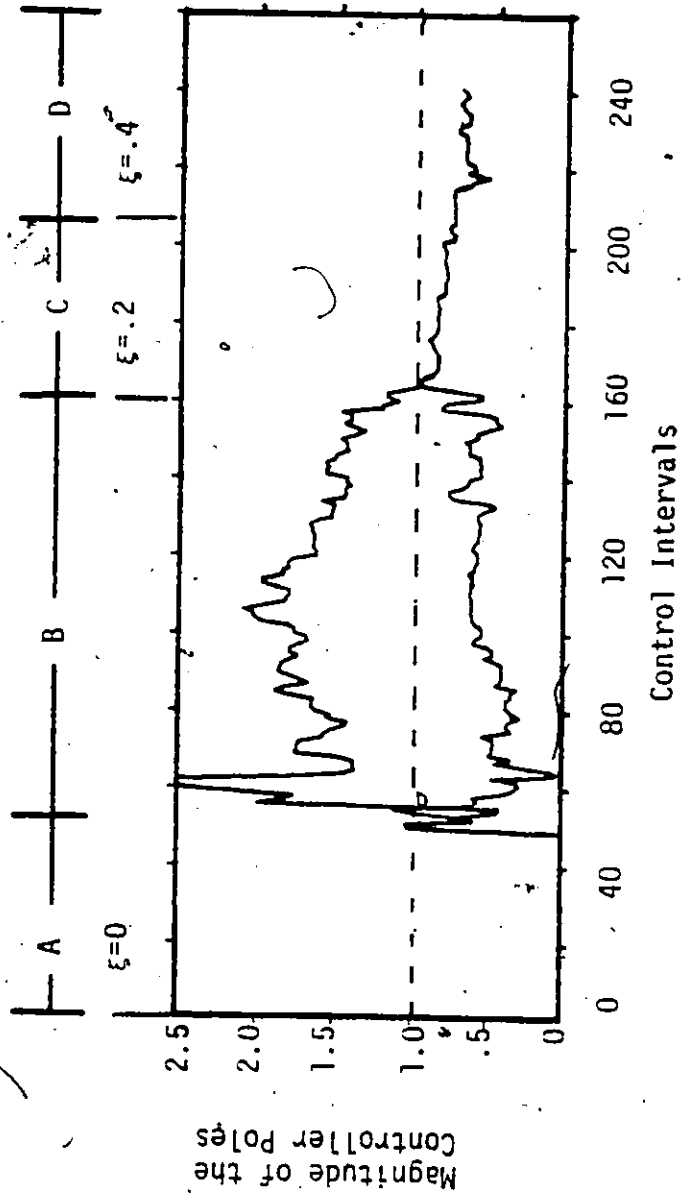


Figure 7.10: Controller Poles for Reactor Run 2

Section A of Figure 7.8, start up of the reactor is shown, and this figure gives an indication of how quickly the temperature can increase.

The self-tuning nature of these regulators was ideal for temperature control in the catalytic reactor application. The catalyst activity was observed to change significantly after shutdowns and when changes were made to more extreme conditions. The STR was able to retune the controller in these situations.

7.3.3 Tuning a PI Controller

A self-tuning algorithm was used to try to see if it could tune the obviously non-optimal PI controller form by using the regression model

$$Y(k+1) = \alpha_0 Y(k) + \alpha_1 Y(k-1) + \beta_0 \nabla U(k) + \varepsilon(k+1) \quad (7.32)$$

The algorithm was never able to tune this controller satisfactorily and usually became unstable before useful results could be obtained. Some of the problems undoubtedly resulted from the inverse response characteristics of the process. The parameter $\hat{\beta}_0$ was observed to change sign at several points causing immediate instabilities. Other problems resulted from a "parameter windup" phenomenon discussed in Section 7.4.7. However, the majority of the problems probably resulted from the simple fact that the PI controller structure is poor and will

only stabilize this highly exothermic reactor system for a very narrow range of parameter settings.

7.3.4 Control Over Effluent Propane Production

Attempts were also made to use a STR to obtain direct control over the effluent propane production rate. Since on-line measurements of the product production rates were available only every 6 minutes, an inferential estimator of propane production rate which used the available reactor temperatures and feedrate measurements was employed every 48 seconds. This estimator was then periodically updated on-line when the chromatographic data became available. This procedure was used successfully in other nonadaptive multivariable control studies on the reactor, Wong (1977). However, in a self-tuning or adaptive mode, this scheme did not work satisfactorily because the inferential estimator, Jutan et al (1977c), Wright et al (1977), contains the manipulated butane flowrate as one of the predictor variables. The values to which the parameters of the inferential estimator converge will therefore depend on the structure and parameters of the controller and will change if the controller does. (This is related to the classical identifiability problem in the presence of feedback.) These two parts of the algorithm (the STR and the inferential estimator) were highly interactive and led to instabilities.

7.4 Discussion

Based on experience obtained from this and a number of other applications, we discuss below some practical considerations that one may find useful when trying to implement one of these self-tuning regulators.

7.4.1 Integral Action and Controller Structure

In the linear process transfer function model (7.1) and the corresponding estimation models (7.11) or (7.23), the manipulated and controlled variables (U, Y) are deviation variables. The reference level of the controlled variable Y will be the desired set point. The corresponding reference level of the manipulated variable necessary to maintain the output at its set point may not be known and will change if the set point of the controlled variable is altered. Therefore, if the disturbance (7.2) is assumed to be stationary ($d=0$), the estimation equation (7.11) should contain an additional level parameter v to be estimated, that is

$$Y(k+b) = \alpha(z^{-1})Y(k) + \beta(z^{-1})U(k) + v + \epsilon(k+b) \quad (7.33)$$

The parameter v reflects the difference between a chosen reference value for the manipulated variable and its true steady-state value that corresponds to the set point of Y . If the steady-state value is not known exactly and the extra parameter v is not estimated, the controlled variable will have offset.

For non-stationary load disturbances or set point changes, the parameter v can be estimated to provide reset for the reference value of the manipulated variable in this proportional type of controller. If v is not included the self-tuning algorithm will attempt to eliminate this offset by forcing a root of $\hat{b}(z^{-1})$ towards unity thereby creating integral action indirectly. (See Figure 7 of Sastry et al (1977).) However, this can lead to ill-conditioning of the estimation space as well as parameter sensitivity, which occur whenever parameters lie near stability boundaries. A more realistic treatment of the control problem is to admit that non-stationary disturbances (deterministic or stochastic) are present in the system, in which case the disturbance model (7.2) will have a root equal to unity (i.e., $d=1$). The optimal controller structure then must contain integral action and is expressible in terms of $\nabla U(k)$ only. The additional v parameter is then unnecessary.

The orders of the polynomials (m, ℓ) in the optimal estimator and controller are given by (7.8) and (7.9) for the minimum variance case or by an equivalent expression resulting from (7.21) in the case of Clarke's constrained algorithm. Many processes can be characterized by relatively simple discrete transfer functions. By assuming a simple non-stationary disturbance structure

$$\left(N(k) = \frac{(1 - \theta z^{-1})}{(1 - z^{-1})} a(k) \right. \quad (7.34)$$

one can then make a tentative choice of the controller order. This stochastic model (7.34) is often capable of representing quite well the type of non-stationary disturbances arising in industry, Box and Jenkins (1970). After sufficient data have been collected, it is then possible to test for optimality and modify the structure if necessary.

If the number of periods of delay (b) in the process is large, then the optimal order (λ) of the $\beta(z^{-1})$ polynomial (7.9) will be large in order to provide dead time compensation. A large number of parameters might then have to be estimated. In this case, one might consider using a longer control interval. If this is undesirable, one could reduce the number of parameters by using models of the form

$$Y(k+b) = \alpha(z^{-1})Y(k) + \beta(z^{-1})(U(k) - (1-\theta)U(k-b)) \quad (7.35)$$

An estimate of θ would be required. As observed in the previous chapter, this regulator structure results when one assumes a disturbance model of the form (7.34) since $\psi(z^{-1})\nabla U(k)$ can be expressed as $U(k) - (1-\theta)U(k-b)$.

7.4.2 Deterministic Disturbances

Minimum variance (dead-beat) and linear quadratic controllers for deterministic (e.g., step) load and set point disturbances can be shown to be identical to those designed for

the equivalent stochastic disturbances when the stochastic disturbance model rational polynomial $\theta(z^{-1})/\phi(z^{-1})\nabla^d$ is the generating function of the deterministic disturbance, Chapter 8. For example, an optimal controller designed for a random walk type of stochastic disturbance

$$N(k) = \frac{1}{1-z^{-1}} a(k) \quad (7.36)$$

is identical to the optimal controller for a step load disturbance (at the output) which has a generating function $1/(1-z^{-1})$. Therefore, the self-tuning regulator will naturally converge to the optimal deadbeat or linear quadratic control algorithm for deterministic disturbances in the process if these are dominant, Wittenmark (1973). The only major additional concern in this case is that the disturbances must be frequent enough to excite the system so that the parameter estimates can be obtained and the P matrix does not become excessively large due to lack of information. The most satisfactory mode of using the STR in these situations is to allow tuning of the controller over a few periods where such disturbances are present and then turn off the estimator portion of the regulator.

7.4.3 Choice of the Control Interval

The choice of the control interval is one of the most important decisions which affects the performance of the self-tuning algorithms, and yet is the one consideration on which

there is little theory or guidance offered in the literature. Note that a distinction is made between the sampling and the control interval. The former can be much shorter with the controller working on filtered values of the measurements obtained between the control intervals. MacGregor (1976, 1977) has shown how a theoretical analysis can be used to predict the effect of changing the control interval on the performance of stochastic controllers. For these stochastic disturbances, it was noted that little is to be gained by sampling faster than the process dead time. However, general rules are difficult to formulate. It is probably safe to state that in industry the control frequency is generally much faster than is necessary for good control. Controlling and updating too rapidly can have a number of adverse effects on self-tuning regulators: The order of the optimal controller often increases. Heavy constraining on the manipulated variable may be necessary, MacGregor (1976). From a parameter estimation point of view, at some point the extra information obtained by sampling faster is marginal. Estimation difficulties then could result from ill-conditioning of the estimation space and instability of the P matrix when discounting is used. Furthermore, if the process dead time changes, the discrete parameter b may change. This has been shown to cause problems with stability of the self-tuning regulators, Wittenmark (1973).

7.4.4 Parameter Identifiability

It is well known that parameter identifiability problems may arise under fixed linear feedback control with no added input signal, Gustavsson et al (1977). In fact, in such a situation, the parameters in the estimation model (7.11) cannot all be uniquely estimated. However, if the dead time is known, the minimum variance controller parameters (α_i/β_0 , β_i/β_0) can be uniquely identified, Bohlin (1971). A frequent solution in many applications has been to fix the β_0 parameter in the estimation model (7.11). Since the self-tuning regulator is time varying, one can usually estimate all the parameters as we have done in the reactor applications. Once the parameters have essentially converged, one can easily fix $\beta_0 = \hat{\beta}_0$ at this point to avoid potential problems if the updating algorithm is to be continued. By estimating β_0 in this way one can also avoid the potential problems of stability and slow convergence resulting from a poor choice of a fixed β_0 , Wittenmark (1973).

7.4.5 Initial Choice of the Parameters

Initial estimates $\underline{\theta}(0)$ and $P(0)$ are needed to start the recursive estimation scheme (7.28), (7.30). From a Bayesian viewpoint $\underline{\theta}(0)$ represents the prior expectation of $\underline{\theta}$ and $P(0)$ represents a matrix proportional to the covariance matrix of the prior distribution of $\underline{\theta}$. Therefore, one can either achieve very rapid updating in the situation of little prior knowledge

(large values for $P(0)$) or very slow but smooth updating in the situation of substantial prior knowledge.

A common practice in the case of poor prior knowledge is to start the parameter estimation algorithm using the model (7.11) during the time that a fixed controller, say a PI algorithm, is still operating, and subsequently to activate the controller (7.14). This procedure was used in the reactor runs as noted earlier. However, it is important to recognize that during the time that the PI controller is operating, the parameter estimates ($\hat{\alpha}$, $\hat{\beta}$) will not be converging towards the true parameter values appropriate for the controller (7.14). This is evident in the development and discussion of equation (7.17) where it is apparent that convergence depends upon estimating the minimum variance controller parameters and simultaneously implementing the controller. Therefore, this initial period should not be too lengthy especially if no discounting factor ($\lambda = 1.0$) is used in the least squares algorithm.

7.4.6 RLS Discounting Factor,

The exponentially discounted least squares algorithm in equations (7.28) and (7.30) with a value of $\lambda < 1.0$ is very useful for tracking the parameters in time varying systems and during the initial transient period of tuning. In fact, values of $\lambda < 1.0$ have been used in almost all STR applications. Since the asymptotic window length or the effective number of

observations included in the estimates is $(1-\lambda)^{-1}$, the usual range of λ is between 0.95 and 1.0. In choosing λ one is effectively trying to trade off the speed of adaptation against the amount of variation in the estimates. It is usually possible to improve the overall performance of the self-tuning controller by periodically changing the value of λ on-line, decreasing it when known changes (e.g., in set point or in the constraining parameter ζ) are being made and then increasing it again when the parameters have nearly converged to their new values.

7.4.7 Parameter Windup

During this study and in at least one industrial study, a problem was encountered which we term "parameter windup" by analogy with the similar phenomenon of reset windup in conventional PI control. If this situation is encountered and adequate precautions not taken it can lead to rapid deterioration of the regular performance.

This problem occurs in the recursive estimation algorithm of the minimum variance and Clarke type self-tuning controllers whenever the manipulated variable saturates at one of its limits (due to actual or imposed constraints). Obviously it is important to account for this saturation in the estimation equations (7.11) and (7.23) so that the parameter estimates do not change in response to $\nabla U(k)$ values called for by the

controller but not actually implemented. The usual procedure for dealing with this has been to insert only the actually implemented control action into the estimation equation. However, this does not solve the windup phenomenon in which the controller parameter estimates begin to diverge from their desired values. This divergence is particularly pronounced whenever $U(k)$ hits and remains at a limit for several periods.

In the above situation two things happen. First, since no new information is being gathered on the β (or γ) parameters in the estimation model during this period, it could lead to problems when using recursive least squares estimation with rapid discounting (say $\lambda = .90$ to $.95$). Those elements of the P matrix associated with the β parameters could increase rapidly leading to a jump in the estimates whenever the situation returns to normal. Secondly, the more serious problem occurs with the α parameters. As the past ∇U 's are set to zero (because of the inability to implement the desired control action during the period of saturation), the true system model of equation (7.15) becomes

$$\begin{aligned} Y(k+b) &= \frac{\alpha(z^{-1})}{\delta(z^{-1})\theta(z^{-1})} Y(k) + \epsilon(k+b) \\ &= \mu(z^{-1})\alpha(z^{-1})Y(k) + \epsilon(k+b) \end{aligned} \quad (7.37)$$

that is, the estimation equation (7.11) is no longer valid. The

terms in $\mu(z^{-1})_{\alpha}(z^{-1})Y(k-j)$ and $\mu(z^{-1})_{\beta}(z^{-1})\nabla U(k-j)$ for $j = 1, 2, \dots$ no longer serve to cancel one another as they did in equation (7.17) when control action was being taken at each interval. The estimates of the α parameters will therefore start to converge towards the leading coefficients of $\mu(z^{-1})_{\alpha}(z^{-1}) = T(z^{-1})/\theta(z^{-1})$. In other words, the estimation model (7.11) and the theorems 5.1 and 5.2 of Astrom and Wittenmark which relate to it are no longer valid because control action is not being taken at every interval due to the saturation.

Clearly the same arguments apply to Clarke's algorithm given in (7.21) and (7.23) except that an additional problem arises which could aggravate this parameter windup. Upon saturation, the dependent variable $\phi(k+b) = (Y(k+b) + \zeta \nabla^d U(k))$ in the regression equation becomes equal to $Y(k+b)$ (if $d = 1$). The algorithm will then behave just as the minimum variance algorithm under saturation since the constraining term has disappeared. If this $\zeta \nabla U(k)$ term is being relied upon to stabilize a non-minimum phase system the shift in the parameter value could have a rather pronounced effect on the controller performance.

If saturation of the manipulated variable might be a common occurrence in a given application one should probably avoid use of the minimum variance and Clarke algorithms that are based on direct estimation of controller parameters. If these are used, one might consider discontinuing the parameter estimation

upon reaching a limit, until such time as the situation has returned to normal and the $\underline{x}(k)$ vector contains only elements of $Y(k-j)$ and $\nabla U(k-j)$ which resulted from normal controller operation. This simplistic approach does not entirely solve the problem since the effect of the controller saturation on the true process model (7.11) can be seen to reach further back than just the elements in the $\underline{x}(k)$ vector. As a result, one usually observes a bump in the parameter estimates when the estimator is turned back on.

One can avoid the parameter windup phenomenon altogether by using those schemes which rely upon direct on-line estimation of the process transfer function and stochastic noise model parameters. The controller parameters are calculated directly on-line from the minimum variance identity or solution of on-line Riccati equations using the model parameters at each interval. This potentially raises the problem of identifiability under closed loop as the parameters converge. Note that although the process parameters may not be uniquely identified in such cases, the controller parameters, which are a nonlinear subset of these, are usually identifiable, Bohlin (1971).

7.4.8 Testing for Parameter Convergence and Controller Optimality

Theorems 5.1 and 5.2 of Astrom and Wittenmark (1973) provide a method of testing for convergence of the parameter estimates and for optimality of the controller by performing

hypothesis tests on certain autocorrelations of the output, $\rho_{\phi\phi}(k)$, and cross correlations between the input and output, $\rho_{u\phi}(k)$.

If the controller is optimal, then, Astrom and Wittenmark (1973)

$$\rho_{\phi\phi}(k) = 0, k \geq b \quad (7.38)$$

Bartlett (1946), has shown that for a stationary normal process, the sample autocorrelations

$$r_{\phi\phi}(k) = \frac{1}{N} \sum_{j=1}^{N-k} \phi(j)\phi(j+k) \quad (7.39)$$

are asymptotically normally distributed with variance given by

$$\text{Var}(r_{\phi\phi}(k)) = \frac{1}{N} \sum_{j=-\infty}^{\infty} (\rho_j^2 + \rho_{j+k}\rho_{j-k} - 4\rho_k\rho_j\rho_{j+k} + 2\rho_j^2\rho_k^2) \quad (7.40)$$

where $\rho_j = \rho_{\phi\phi}(j)$.

If the controller is optimal implying that (7.38) is true, this reduces to

$$\text{Var}(r_{\phi\phi}(k)) = \frac{1}{N} (1 + 2 \sum_{j=1}^{b-1} \rho_j^2) \quad k \geq b \quad (7.41)$$

As suggested by Box and Jenkins (1970), the $b-1$ nonzero, ρ_j 's can be replaced by their estimates $r_{\phi\phi}(k)$ and then each of the $r_{\phi\phi}(k), k \geq b$ can be compared with their two standard deviation

limits to test the hypothesis that they are zero. Box and Jenkins also given an overall Chi-squared test for this hypothesis.

If the parameters have converged then for the case of ordinary non-discounted recursive least squares theorem 5.1 of Astrom and Wittenmark shows that

$$\begin{aligned} \rho_{\phi\phi}(k) &= 0 & k &= b, \dots, b+m \\ \rho_{u\phi}(k) &= 0 & k &= b, \dots, b+l \end{aligned} \quad (7.42)$$

Therefore, using Bartlett's variance formulae the same tests described above can be used to test for convergence of the parameters.

If exponentially discounted least squares is used to estimate the parameters, a rederivation of Astrom and Wittenmark's theorem 5.1 shows that if the parameters have converged the expectations of the following autocorrelations and cross-correlations should be zero (Appendix 4) instead of those in (7.42).

$$\begin{aligned} E(\bar{r}_{\phi\phi}(k)) &= 0 & k &= b, \dots, b+m \\ E(\bar{r}_{u\phi}(k)) &= 0 & k &= b, \dots, b+l \end{aligned} \quad (7.43)$$

where

$$\bar{r}_{\phi\phi}(k) = \frac{1}{n} \sum_{i=1}^{N-k} \lambda^{N-i} \phi(i)\phi(i+k) \quad (7.44)$$

With $n = (1-\lambda)^N / (1-\lambda)$ this provides an unbiased estimate of $\rho_{\phi\phi}(k)$. $\bar{r}_{u\phi}(k)$ is similarly defined.

Following Bartlett's approach the asymptotic variance of these sample autocovariances are given by (Appendix 4)

$$\text{Var}(\bar{r}_{\phi\phi}(k)) = \frac{1}{n^2} \sum_{j=-n+1}^{n-1} (\rho_j^2 + \rho_{j+k}\rho_{j-k} - 4\rho_k\rho_{j+k}\rho_j + 2\rho_k^2\rho_j^2) \cdot \lambda^{|j|} \frac{(1-\lambda^{2n})}{(1-\lambda^2)} \quad (7.45)$$

For values of λ substantially less than 1.0, this variance can be much larger than that given by (7.39).

This test for convergence was applied to sections C and D of reactor run 1. The sample weighted autocorrelations $\bar{r}_{\phi\phi}(k)$ are plotted together with their approximate two standard deviation limits (calculated on the assumption that (7.38) is true, Figure 7.11. These limits are much larger than those for the unweighted sample autocorrelations ($2/\sqrt{38}$ vs $2/\sqrt{90}$) shown above sections C and D of Figure 7.3. From this analysis we would conclude that the parameter estimates have not converged in section C, while in section D they appeared to have converged.

7.5 Summary

The self-tuning control theory discussed in this paper provides a direct on-line means of implementing univariate stochastic controllers. This is an alternative to the usual

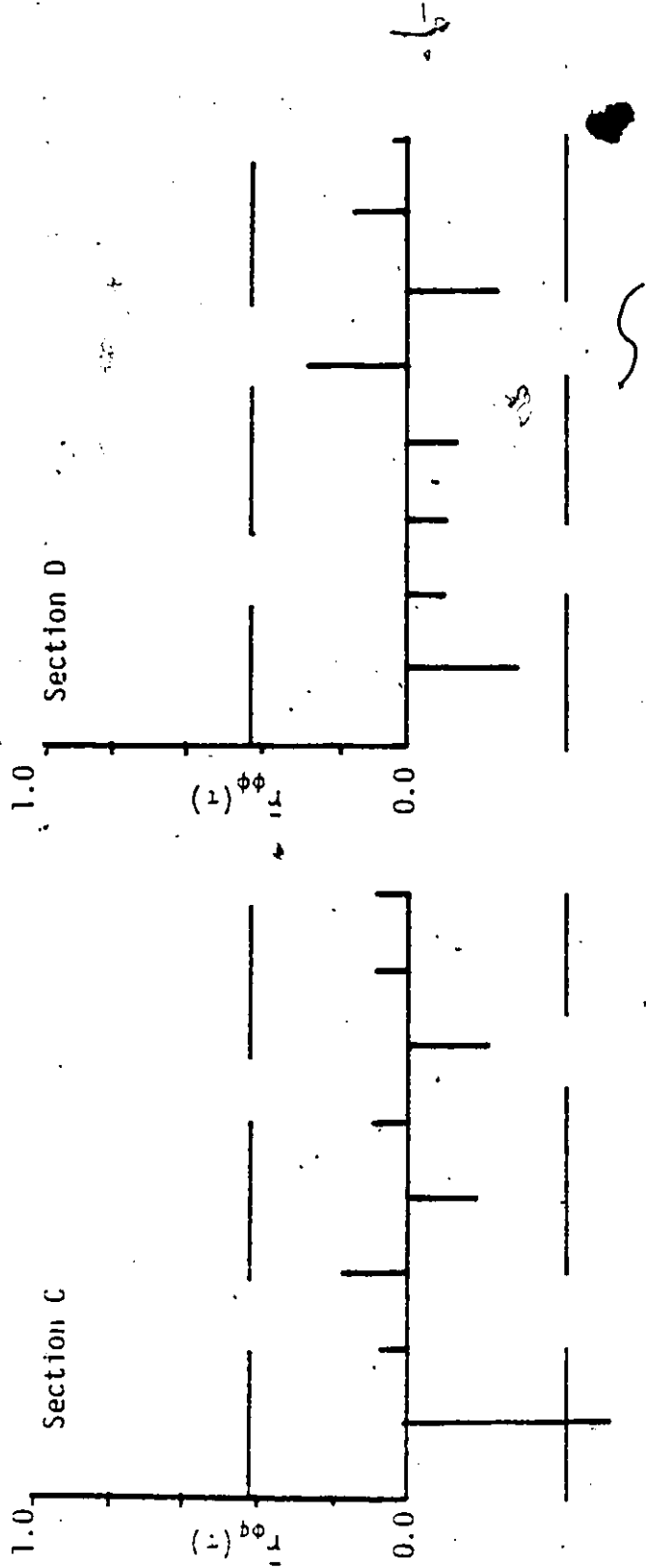


Figure 7.11: Alternative Run Weighted Autocorrelation Functions for Reactor Run 1

procedure of developing dynamic-stochastic models off-line using a set of data collected from the process and designing the controller from there. The strategy is easily implemented and it can be used in an adaptive mode allowing the controller to track changing process conditions.

Various forms of these controllers were implemented on a pilot scale packed bed catalytic reactor. The algorithm was shown to tune itself rapidly, and to adapt to changing catalyst activity and operating levels. Constraint on the variation in the manipulated variable was easily handled by Clarke's modification of the basic STR algorithm. The quality of control of the reactor hot spot temperature was considerably better than that achievable by conventional PI control. Although better control over the more economically meaningful production rates can be achieved using modern multivariable stochastic control theory, Jutan et al (1977a,b,c), the simplicity and adaptive nature of these STR schemes in comparison with the latter approach makes them very attractive.

CHAPTER 8

MODELLING AND CONTROL OF DETERMINISTIC AND
STOCHASTIC PROCESSES8.1 Introduction

One of the most appealing aspects of the linear quadratic approach, is the ease with which additive stochastic disturbances are incorporated in the filtering and control algorithms. Not all disturbances affecting processes are of a stochastic nature. Steps, ramps, exponential rises to a new level, and periodic disturbances such as sinusoids, are examples of deterministic load disturbances that might be encountered. To detect and control these disturbances, they must be modelled and included in the description of the process. In this chapter, attention will be focused on the modelling and control of processes subject to deterministic and stochastic disturbances. Optimal control algorithms designed for stochastic processes, are seen to be optimal control algorithms for deterministic processes of "equivalent" structure. Difficulties in the control of processes subject to both types of disturbances are investigated.

To motivate some of these ideas, consider the following simple example. A process transfer function is described by

$$Y(k) = \frac{g(1-s)}{1-\delta z^{-1}} U(k-2) + N(k) \quad (8.1)$$

where z^{-1} is the backward shift operator. It is desired to find a linear feedback controller, $D(z^{-1})$, so that when the process (8.1) is subject to a sudden constant load disturbance of unknown magnitude, a_0 , at the process input, the process output, $Y(k)$, is brought to zero and held there in the fewest number of control intervals. The controller that will accomplish this is known as a minimal prototype or deadbeat controller. A block diagram of this process is shown in Figure 8.1. The load variable has the transfer function, Raggazini and Franklin (1958),

$$N(k) = \frac{g(1-\delta)a(k)}{(1-\delta z^{-1})(1-z^{-1})} \quad (8.2)$$

$a(k)$ is zero for all time, except at the instant when the load disturbance first affects the process, at which point it has the value a_0 . (This is equivalent to the z-transform of the effect of an input step at time $k=0$ on the process output.) Using the method of undetermined coefficients (a purely algebraic approach), the controller that accomplishes this is, Smith (1967),

$$U(k) = \frac{-(1+\delta+\delta^2) + \delta(\delta+1)z^{-1}}{(1-z^{-1})(1+(1+\delta)z^{-1})} Y(k) \quad (8.3)$$

Now suppose that we designed a minimum variance controller for the process (8.1) and noise model (8.2), under the different assumption that the $a(k)$'s were from a normal population. The resulting controller is identical to (8.3). In this example, the

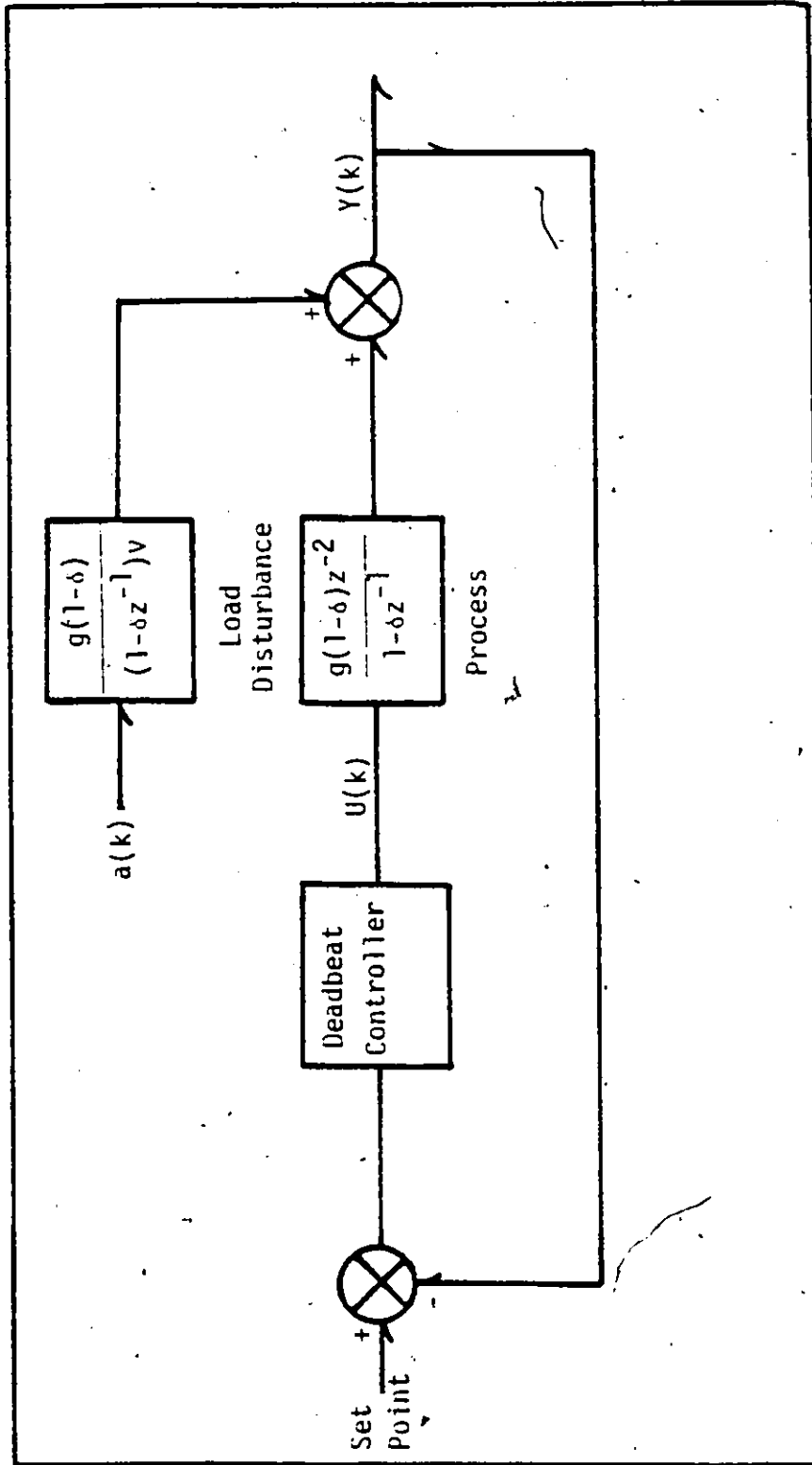


Figure 8.1: Deadbeat Controller for Load Disturbance

structure of the controller did not depend on the statistical properties of the $a(k)$'s. Furthermore, this same duality between control for deterministic and stochastic disturbances can be shown to hold for the more general case of Linear-Quadratic-Control (LQC), that is, where one minimizes the output variance subject to a restriction on the input variance. Wilson (1970) gives a clear presentation of this duality using the spectral factorization (or more correctly covariance generating function factorization in discrete cases) approach of Wiener (1949) and Whittle (1963).

As we observed in Chapter 6, a minimum variance controller is designed to cancel out the minimum variance forecast of $N(k+b)$ given only information to time k , that is

$$\frac{w(z^{-1})}{\delta(z^{-1})} U(k) + \hat{N}(k+b/k) = 0 \quad (8.4)$$

The structure of the controller obviously depends on the forecast. An important point in the duality of stochastic and deterministic controllers is that for linear forecasts, the structure of the forecast is independent of the distribution from which the $a(k)$'s come. To construct this linear forecast from a statistical argument, the $a(k)$'s need only be from a mean zero process with finite variance, Box and Jenkins (1970).

We can consider both the deterministic and stochastic disturbances to be modelled by linear difference equations. The solution of these difference equations consists of a complimentary

solution, $C(k+b/k)$ and a particular solution, $P(k)$, that is

$$N(k+b) = P(k) + C(k+b/k) \quad (8.5)$$

The complimentary solution involves initial conditions and as such utilizes information available to time k . The particular solution is the forced solution and incorporates the information ($a(k)$'s) that enter the process from time $k+1$ to $k+b$. For stochastic disturbances, it is well known that $C(k+b/k)$ is the linear minimum variance forecast $\hat{N}(k+b/k)$, Box and Jenkins (1970). For deterministic disturbances of the form (8.2), $\hat{N}(k+b/k)$ in (8.5) is the complimentary solution for this difference equation with the initial conditions determined from $a(k)$, $a(k-1)$ Therefore, the same minimum variance, or minimal prototype controller will result from disturbances of the same structure (transfer function) irrespective of the distribution of the shocks $a(k)$.

The notion that process disturbances can be represented as difference equations, having complimentary and particular solutions, is fundamental to the modelling and control of processes subject to both stochastic and/or deterministic disturbances. This topic is explored in more detail in the next section.

8.2 Modelling of Stochastic and Randomly Occurring Deterministic Disturbances

Multivariate disturbances of a random nature can often be modelled by vector stochastic difference equations of the form

$$\nabla^d \phi(z^{-1}) \underline{d}(k) = \theta(z^{-1}) \underline{a}(k) \quad (8.6)$$

or equivalently

$$\phi(z^{-1}) \underline{d}(k) = \theta(z^{-1}) \underline{a}(k) \quad (8.7)$$

The $(a(k))$'s are really a sequence of normally distributed random variables or shocks, uncorrelated in time. ∇ is an abbreviation for $(1-z^{-1})I$. $\theta(z^{-1})$ and $\phi(z^{-1})$ are matrix polynomials of order p and q . The modulus of the determinant of these polynomials is assumed to be less than one. In the univariate case, this corresponds to the polynomials having all their zeroes inside the unit circle in z . The dimension of the disturbance is taken as n_d . Equation (8.6) is a parametric representation of the disturbances affecting the process. Box and Jenkins (1970) and Astrom (1970) discuss techniques for estimation (p,d,q) and the parameters of the model for univariate time series. Multiple time series are discussed by Quenouille (1957) and Kashyap and Rao (1976).

In this section, we examine in more detail the difference equation representations for stochastic and deterministic disturbances introduced in section 8.1. This motivates a discussion on the modelling and control of processes subject to stochastic and deterministic disturbances.

To introduce these ideas, consider the univariate stochastic difference equation

$$d(k) - d(k-1) = a(k) \quad (8.8)$$

The solution of this difference equation at time origin k_0 consists of a complimentary and particular solution, and is given by

$$d(k) = b(k_0) + \sum_{j=k_0+1}^k a(j), \quad k > k_0 \quad (8.9)$$

$b(k_0)$ is obtained from the initial conditions and with no loss of generality may taken as

$$b(k_0) = \sum_{j=-\infty}^{k_0} a(j) \quad (8.10)$$

The complimentary solution $b(k_0)$ is a prediction for $d(k)$ using known information available up to and including k_0 . This prediction, denoted by $d(k/k_0)$ will be in error by the amount $\sum_{j=k_0+1}^k a(j)$. The complimentary solution is the conditional mean of $d(k)$ given only information to time k_0 , Box and Jenkins (1970), and as such minimizes the variance of the prediction error. The prediction error is the particular solution to (8.8).

If the origin of the projection changes from k_0 to k_0+1 then

$$d(k) = b(k_0+1) + \sum_{j=k_0+2}^k a(j), \quad k > k_0+1 \quad (8.11)$$

where the coefficient $b(k_0+1)$ is given by

$$b(k_0+1) = b(k_0) + a(k_0+1) \quad (8.12)$$

The effect of the shock $a(k_0+1)$ is to alter the level of the disturbance $d(k)$. Using (8.12) the original stochastic difference equation may be written in the form

$$d(k) = z(k) \quad (8.13)$$

where

$$z(k+1) = z(k) + a(k+1) \quad (8.14)$$

These two equations represent a state space description for a sequence of shocks entering the level of the disturbance process at each time interval.

Let us consider another example. The stochastic difference equation is

$$d(k) - (1+\phi)d(k-1) + \phi d(k-2) = a(k) - \theta a(k-1) \quad (8.15)$$

with $-1 \leq \theta, \phi \leq 1$. The left hand side of (8.15) describes the exponential rise of $d(k)$ to a new level. The solution to (8.15) is given by

$$d(k) = b_1(k_0) + b_2(k_0)\phi^{k-k_0} + \sum_{j=k_0+1}^k \left(\frac{1-\theta}{1-\phi} + \frac{\theta-\phi}{1-\phi} \cdot \phi^{k-j} \right) a(j) \quad (8.16)$$

$$k > k_0$$

Using the same approach as in the previous example, one can obtain

the following state space description of the stochastic disturbance

$$d(k+1) = b_1(k) + b_2(k) \quad (8.17)$$

where

$$b_1(k+1) = b_1(k) + \frac{1-\theta}{1-\phi} a(k+1) \quad (8.18)$$

and

$$b_2(k+1) = \phi b_2(k) + \frac{\theta-\phi}{1-\phi} a(k+1) \quad (8.19)$$

The effect of the shock is to alter the time evolution of $d(k)$ from what would be calculated using the complementary solution. The shocks enter as an additive term to this solution. The moving average and autoregressive parameters determine how the shocks filter through to the difference equations for $b_1(k)$ and $b_2(k)$. The variance of the $a(k)$'s determines the magnitude of the changes in $b_1(k)$ and $b_2(k)$ arising from the forced solution.

The term $b_1(k_0) + \phi^{k-k_0} b_2(k_0)$ in (8.16) is the complementary solution, $C(k/k_0)$ to (8.15). From (8.16) we see that it is also the conditional mean for $d(k)$ given information up to and including time k_0 . However, in Chapters 6 and 7, the conditional mean was obtained as the term $\Gamma(z^{-1})a(k_0)/(v(1-\phi z^{-1}))$ in the expansion

$$\frac{(1-\phi z^{-1})}{v(1-\phi z^{-1})} d(k) = \psi(z^{-1})a(k) + z^{-(k-k_0)} \frac{\Gamma(z^{-1})}{v(1-\theta z^{-1})} a(k) \quad (8.20)$$

The two expressions for the conditional mean must be equal.

This equivalence is obtained by using (8.18) and (8.19) to express $b_1(k_0)$ and $b_2(k_0)$ in terms of the shocks $a(k_0), a(k_0-1), \dots$

If this is done we find that

$$c(k/k_0) = \hat{d}(k/k_0) = \frac{1-\theta}{1-\phi} \cdot \frac{1}{(1-z^{-1})} \cdot a(k_0) + \frac{\theta-\phi}{1-\phi} \cdot \frac{\phi^{k-k_0}}{(1-\phi z^{-1})} \cdot a(k_0) \quad (8.21)$$

For any value of $k > k_0$, it is straightforward to show that $T(z^{-1})a(k_0)/(\nabla(1-\phi z^{-1}))$ can be written in the form (8.21).

From the examples just considered, we have seen that the minimum variance forecast, $\hat{d}(k/k_0)$, for $d(k)$ can be obtained as the solution to a homogenous difference equation. This construction, as opposed to the factorization approach of Box and Jenkins (1970) and Astrom (1970), gives us a greater insight into the structure of the forecast and explicitly indicates what effect the lead time $(k-k_0)$ has on the forecast. Other examples of stochastic difference equations are discussed in Box and Jenkins (1970) and Abraham and Box (1978).

The difference equations (8.6) or (8.7) can always be written in the form

$$\underline{d}(k) = (I_{nd} \ 0 \ \dots \ 0) \underline{z}(k) \quad (8.22)$$

and

$$\underline{z}(k+1) = \begin{bmatrix} \phi_1 & I_{nd} & \dots & \dots & 0 \\ \phi_2 & 0 & I_{nd} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{p+d} & 0 & \dots & \dots & I_{nd} \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix} \underline{z}(k) + \begin{bmatrix} I \\ -\theta_1 \\ \vdots \\ \vdots \\ \theta_{p+d} \end{bmatrix} a(k+1) \quad (8.23)$$

or more compactly as

$$\underline{d}(k) = C\underline{z}(k) \quad (8.24)$$

and

$$\underline{z}(k+1) = D\underline{z}(k) + \underline{r}a(k+1) \quad (8.25)$$

We note that the matrix D involves only the coefficients of the autoregressive equation.

The shocks in (8.6) enter at every interval. Let us suppose that the $a(k)$'s do not enter at every time interval, but rather appear occasionally, if at all. To clearly distinguish the $a(k)$'s from these occasional shocks, the latter will be denoted by $\underline{\alpha}(k)$. If the moving average term $\underline{\theta}(z^{-1})=I$, the description of the process disturbance is given by

$$\phi(z^{-1})\underline{d}(k) = \underline{\alpha}(k) \quad (8.26)$$

If $\underline{\alpha}(k)$ were zero for all time, the disturbance would evolve from its initial conditions. Given $p+d$ starting values $d(k)$, $\underline{d}(k+1)$, . . . $\underline{d}(k+p+d)$, the future behaviour of the disturbance could be exactly predicted if no shocks $\underline{\alpha}(k)$ were to affect the process. With an appropriate choice of the autoregressive parameters, level changes, and exponential rises to a new level could be described. If $\phi(z^{-1})$ is allowed to have roots on the unit circle, periodic disturbances such as sinusoids can be modelled. Since there is no uncertainty in predicting the future value of these processes, once the starting values have been determined, such equations describe deterministic disturbances.

In most processes, however, neither the time at which the disturbance affects the process nor its magnitude are known prior to the occurrence of this event. Therefore, to allow for these types of disturbances, which we shall call randomly occurring deterministic disturbances, it is essential that we acknowledge in the mathematical model, non zero values of the shocks $\underline{\alpha}(k)$. For most of the time when there are no disturbances $\underline{\alpha}(k) = 0$. When a shock occurs, the disturbance evolves according to its difference equation until a new shock again affects the process.

For example, if we were to model an exponential rise to a new level, the state equations would be

$$d(k) = (1 \quad 0) \begin{bmatrix} z_1(k) \\ z_2(k) \end{bmatrix} \quad (8.27)$$

and

$$\begin{bmatrix} z_1(k+1) \\ z_2(k+1) \end{bmatrix} = \begin{bmatrix} \phi+1 & 1 \\ -\phi & 0 \end{bmatrix} \begin{bmatrix} z_1(k) \\ z_2(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \alpha(k+1) \quad (8.28)$$

These randomly occurring disturbances can be modelled by the state equations

$$\underline{d}(k) = \underline{C}\underline{z}(k) \quad (8.29)$$

and

$$\underline{z}(k+1) = \underline{D}\underline{z}(k) + \underline{\Gamma}\alpha(k+1) \quad (8.30)$$

We immediately notice the similarity between the equations, and those used to model stochastic disturbances (8.24) and (8.25). The fundamental difference between modelling stochastic and randomly occurring deterministic disturbances is in the distribution and frequency of the shocks affecting the process. The continuous analog of such disturbances are differential equations. Johnson (1971, 1972a, 1974a) discusses these representations in detail.

8.3 Control of Processes Subject to Deterministic Disturbances

The key to controlling processes subject to deterministic or stochastic disturbances is to include a proper mathematical description of the disturbance in the process model. The effect

of the disturbances on the state equations are assumed to be described by

$$\underline{x}(k+1) = A\underline{x}(k) + G\underline{u}(k-f) + \underline{F}d(k) \quad (8.31)$$

The disturbances are modelled by the equations (8.29) and (8.30). There are n states, r manipulated variables, f ($f \geq 0$) pure periods of transport delay, and n_d disturbances. As well, we may have a measurement equation of dimension m ,

$$y(k) = H\underline{x}(k) \quad (8.32)$$

The process disturbances have been modelled as an additive term. This corresponds to a linear(ized) representation of the process plus disturbances. If need be, we can also include the effect of the disturbances in the measured outputs.

The control problem is to detect the presence of a disturbance and to eliminate, or minimize its influence on either the states or the process outputs if possible. There have been a number of studies of this problem. Most attention has been focused on the case of continuous dynamics where F in (8.31) is of rank n (the number of states), Davison and Smith (1970), Davison (1972). The papers by Davison are very difficult to read and the control strategies are not well motivated. One almost needs to know the final result before contemplating the controller design. Necessary and sufficient conditions for the existence of a controller to asymptotically eliminate the effect of disturbances

on the process outputs are given, but they lack an intuitive interpretation. Little guidance is offered on how to proceed if it is impossible to satisfy the necessary and sufficient conditions.

In many process applications, we do not have independent disturbances affecting all the states. In the reactor study, a major load variable is the wall temperature. As shown in Chapter 3, this affects all the states (temperatures along the reactor) simultaneously. The conditions for the existence of solutions to the control problem in the aforementioned papers do not appear to be readily extended to the more general case when the rank of F is less than n . Furthermore, the 'robustness' properties of the controllers are not applicable when the rank of F is less than n , Davison (1972).

Johnson (1968, 1970a, 1970b, 1971, 1972a, 1972b, 1973, 1974a) has developed an approach to control processes subject to deterministic disturbances. One considers controlling the process described by (8.31) as if the states $\underline{x}(k)$, and disturbances $\underline{d}(k)$ were exactly known. This implies that $\underline{\alpha}(k)$ is zero for all time. An observer is then constructed for these variables. A certainty equivalence theorem is proven to show that one can in fact control the process using the estimated states instead of the true values. Johnson's 1971 paper is a nice exposition of the method. Unfortunately reduced order observers were used to reconstruct the states, and this detracts from the insight that is

obtained when full order observers are used. The 1973 paper, presented at a University of Waterloo symposium, although not widely available, is perhaps the clearest explanation of his work. The technique was developed for continuous systems, but is readily extended to the discrete time case.

8.3.1 Elimination of the Disturbances on the States

For the moment, let us assume that $C \neq I$, i.e., we are primarily interested in controlling deterministic disturbances entering directly into some of the states. As proposed by Johnson, one can partition the control action into two components

$$\underline{u}(k) = \underline{u}^R(k) + \underline{u}^D(k) \quad (8.33)$$

$\underline{u}^R(k)$ is to regulate the states and $\underline{u}^D(k)$ is used to counteract the effect of the disturbances. If $\underline{u}^D(k)$ is chosen so that

$$G\underline{u}^D(k) + FC\underline{z}(k+f) = \underline{0} \quad (8.34)$$

then the disturbance will have no effect on the original process. (Since $\underline{d}(k)$ is presumed known, and its time evolution is also known, we are able to predict its value at time $k+f$). The necessary conditions needed to accomplish this, regardless of the value of $\underline{d}(k)$, are that, Johnson (1971)

$$\text{rank}(G|FC) = \text{rank}(G) \quad (8.35)$$

In other words, in order to collapse $G\underline{u}(k-f)$ onto $-FC\underline{z}(k)$, \underline{u} must

span the vector space in which the disturbance affects the process. If there are as many disturbances as there are states, then there must also be as many controls as there are states. For (8.35) to hold, it is necessary, but not always sufficient, that the number of controls equal the number of disturbances affecting the process. If (8.35) holds, the control is given by

$$\underline{u}(k) = -L_1(k)\underline{x}(k) - L_2(k)\underline{z}(k+f) \quad (8.36)$$

$L_1(k)$ is designed to regulate the states $\underline{x}(k)$, as if no disturbances were present. Techniques such as pole placement or linear quadratic strategies can be used to accomplish this. When G is of maximal rank (the rank equal to the number of controls), $L_2(k)$ is given by

$$L_2(k) = (G'WG)^{-1}G'WFC \quad (8.37)$$

where W is any nonsingular weighting matrix. The elements of $L_2(k)$ are not unique as claimed by Johnson (1971).

When (8.35) does not hold, Johnson (1971) suggests that one minimize some norm of $G\underline{u}^D(k-f) + F\underline{d}(k)$. If the Euclidian norm is used, $\|G\underline{u}^D(k-f) + F\underline{d}(k)\|$, the pseudoinverse of G minimizes this quantity, and the control $\underline{u}^D(k)$, when G is of maximal rank, is given by (8.37) with $W=I$. It is important to realize that in this case, we cannot remove the total effect of the disturbances on all the states, and special care must be taken in specifying $L_1(k)$. This is discussed in more detail in section 8.5.

8.3.2 Elimination of the Disturbance on the Outputs

In some instances, one is interested in eliminating the effect of the disturbances on the measured outputs, while at the same time ensuring that the states are bounded for all bounded disturbances. The solution is somewhat more complicated in this case. There is a fundamental error in Johnson's (1974b) paper. A later paper, Johnson (1975), expanded on the topics presented in the 1974b paper but did not entirely correct the fault. The correction will be shown.

From (8.31) and (8.32) we see that the effect of the process disturbances on the output, call this $\underline{y}^D(k)$, is given by

$$\underline{y}^D(k) = H \sum_{j=0}^{k-1} A^j (G \underline{u}^D(k-1-f-j) + FC \underline{z}(k-1-j)) \quad (8.38)$$

We recall that the null space of a matrix P is the set of vectors \underline{q} such that $P\underline{q} = \underline{0}$. If we seek a time-invariant controller of the form

$$\underline{u}^D(k) = -L_2 \underline{z}(k+f) \quad (8.39)$$

then $-GL_2 + FC$ must lie in the null space of HA^j if (8.38) is to vanish for all j , Johnson (1974), Wonham (1974). A necessary and sufficient condition for this to occur is that, Johnson (1974), Wonham (1974)

$$\text{rank}(Y'(G|FC)) = \text{rank}(Y'G) \quad (8.40)$$

where Y is given by

$$Y = (H' | A'H' | A^2'H' | \dots | A^{n-1}H')$$
 (8.41)

When H is a square $n \times n$ nonsingular matrix condition (8.40) is identical to (8.35). If condition (8.40) holds, a time-invariant matrix which ensures that $-GL_2 + FC$ remains in the null space of HA^j is given, by Johnson (1974b)

$$L_2(k) = (Y'G)^{\#} Y'FC$$
 (8.42)

where $(Y'G)^{\#}$ denotes the pseudoinverse of $(Y'G)$.

To guarantee that the states are bounded for all bounded disturbances, $L_1(k)$ is designed so that $A - GL_1(k)$ has its eigenvalues inside the unit circle. To arbitrarily assign these (A, G) must be controllable.

However, having now designed a controller to stabilize the states, it may not be possible to eliminate the effect of the disturbances on the process outputs even if condition (8.40) holds. To see this, we note that the effect of the process disturbances on the outputs is now given by

$$y^D(k) = H \sum_{j=0}^{k-1} (A - GL_1(j))^j (-GL_2 + FC) d(k-1-j)$$
 (8.43)

The controller L_2 was designed so that $-GL_2 + FC$ was in the null space of HA^j for $j=0,1,2,\dots$. This does not ensure that $-GL_2 + FC$ is in the null space of $H(A - GL_1(k))^j$. To see this, consider the

following example, Let the state equation be

$$\underline{x}(k+1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \underline{x}(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot d(k) \quad (8.44)$$

$$d(k+1) = d(k) + \alpha(k+1) \quad (8.45)$$

$$y(k) = (0 \ 1) \underline{x}(k) \quad (8.46)$$

It is obvious that the effect of the disturbances on the output can be removed with the control $\underline{u}^D(k) = -d(k)$. As expected, condition (8.40) holds. If we now design a time-invariant controller, L_1 , to place on both eigenvalues of $A - GL_1$ at .707, it is easy to verify that the disturbance $d(k)$ propagates to the measured output in spite of the action of $\underline{u}^D(k)$.

The obvious modification to Johnson's (1974b) work, when the state controller is time-invariant, is to design the stabilizing controller for the states, and then examine whether it is possible to remove the effect of the disturbances on the process outputs by examining condition (8.40) with $A - GL_1$ replacing A . This is somewhat unsatisfactory as it is not apparent under what conditions a solution exists.

To pursue this briefly, we can pose the problem in the following manner. Suppose it is possible to find an L_2 such that

$$H(-GL_2 + FC) = 0 \quad (8.47)$$

is satisfied. Then what are the restrictions on the time invariant state controller L_1 such that

$$H(A-GL_1)^j(-GL_2+FC) = 0, j=1,2,\dots \quad (8.48)$$

holds? We recall from linear algebra that if \underline{h}_i is a left eigenvector of $A-GL_1$ and λ_i the corresponding eigenvalue then

$$\underline{h}_i(A-GL_1) = \lambda_i \underline{h}_i \quad (8.49)$$

For $(A-GL_1)^j$ to be in the null space of H , L_1 must be chosen so that the m rows of H are m eigenvectors of $A-GL_1$. If this can be done then

$$H(A-GL_1)^j(-GL_2+FC) = \Lambda H(-GL_2+FC) \quad (8.50)$$

This expression vanishes if we can find an L_2 such that (8.47) holds. Λ is an $m \times m$ diagonal matrix whose elements are the eigenvalues of $A-GL_1$ having the eigenvectors H . If (A,G) is controllable, the n eigenvalues of $(A-GL_1)$ can be arbitrarily assigned. It would be unlikely though that using a linear quadratic strategy, they would be chosen so that (8.50) holds. Rather than arbitrarily specify n eigenvalues of $(A-GL_1)$, (8.49) indicates that we may be able to arbitrarily specify m eigenvalues by choosing L_1 such that

$$HGL_1 = HA - \Lambda_m H \quad (8.51)$$

The m prescribed eigenvalues of $A-GL_1$ are denoted by Λ_m . Equation (8.51) is a set of mn equations in nr unknowns (L_2). For a solution to exist, there cannot be more equations than unknowns. A necessary condition for a solution is that

$$(\text{number of unknowns}) r \geq m (\text{outputs to be controlled}) \quad (8.52)$$

This result is not unexpected. As well, we require (8.47) to vanish. A necessary and sufficient condition for this is that

$$\text{rank } (HG|HFC) = \text{rank } (HG) \quad (8.53)$$

From this discussion, we can see that if we are to drive m outputs to zero, it is not possible to insure the stability of all the states unless there are $r=n$ controls.

For the example (8.44-8.46), it can be shown, using (8.47) and (8.51), that the controllers $u^D(k) = -d(k)$ and $u^D(k) = -(0, 1-\lambda)$, will eliminate the disturbance in the outputs for all time, while at the same time placing one eigenvalue of $(A-GL_1)$ at λ . The second eigenvalue is fixed at one, and cannot be altered if we are to eliminate the disturbance in the output for all time.

8.3.3 Alternate Approaches to Disturbance Minimization

Before concluding this section, we comment on two other approaches. To control a discrete system subject to a step disturbance in all the states, Bradshaw and Porter (1978) took the following approach. Since it is 'known' that integral action is required to eliminate the offset, the structure of the controller is

$$\underline{u}(k) = -L_1 \underline{x}(k) - L_2 \underline{q}(k) \quad (8.54)$$

$$\underline{q}(k+1) = \underline{q}(k) + \underline{y}(k) \quad (8.55)$$

The controller gains are chosen so that the controller plus process

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{q}(k+1) \end{bmatrix} = \begin{bmatrix} A-GL_1 & -GL_2 \\ H & I \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{q}(k) \end{bmatrix} + \begin{bmatrix} I \\ 0 \end{bmatrix} \underline{d}(k) \quad (8.56)$$

is asymptotically stable. If this can be accomplished, and certain alternate rank conditions hold, then $\underline{y}(k) \rightarrow \underline{0}$ asymptotically. Necessary and sufficient conditions for the existence of this controller are given. It is also assumed that FC is of rank n . Is this approach fundamentally different from the more general approach of Johnson? The keyword to note is that in this method, the effect of the disturbance on $\underline{y}(k)$ is brought to zero asymptotically. This means that $(A-GL_1)^j(-GL_2 + FC)$ is no longer required to be in the null space of H for all j . From (8.43), we see instead, that $(I-A+GL_1)^{-1}(-GL_2+FC)$ must lie in the null space of H . Given a state controller L_1 , a necessary and sufficient condition for this is that

$$\text{rank}(H(I-A+GL_1)^{-1}G) = \text{rank}(H(I-A+GL_1)^{-1} \cdot (G|FC)) \quad (8.57)$$

When FC is of rank n , (8.57) becomes

$$\text{rank}(H(I-A+GL_1)^{-1}G) = m \quad (8.58)$$

Since the bracketed term in (8.58) is an $m \times r$ matrix, a necessary condition for (8.58) to hold is that the number of controls must

equal or exceed the number of variables to be asymptotically driven to zero. The more general approach of Johnson encompasses the specific method of Bradshaw and Porter (1978). For example, in (8.44-8.46); it is possible to arbitrarily assign the eigenvalues of $(A-GL_1)$ and asymptotically eliminate the effect of the disturbance $d(k)$ (if it does not grow in time). Although it is possible to remove the offset asymptotically, it is impossible, as we have seen, to eliminate the offset in a short (one stage) time frame and still arbitrarily assign the eigenvalue of $A-GL_1$.

Several comments can be made on the approach of Bradshaw and Porter (1978), Davison and Smith (1970), Davison (1972). It is assumed that all the states are measured. If they are not, then they must be reconstructed using an observer. To obtain meaningful estimates of the states, the disturbances must be included in the observer design. Once we obtain estimates of the disturbances, why not then use them directly in the controller design? Furthermore, it is not obvious what the controller structure should be for, as an example, an exponential rise to a new level. The structure of the controller evolves in a straight forward fashion once a mathematical model of the disturbance has been formulated.

Another technique proposed to guard against step disturbances, Johnson (1968), Tomizuka and Rosenthal (1979), is to re-arrange the state model and express the control in terms of $\underline{v}(k)$. One then designs the controller to minimize an objective

function of the form

$$J_1 = \frac{1}{N} \sum_{k=0}^N (\underline{x}'(k)Q\underline{x}(k) + \nabla\underline{u}'(k-f-1)R\nabla\underline{u}(k-f-1)) \quad (8.59)$$

This approach always assumes that there is a step disturbance affecting the process, and can lead to poor performance when no disturbance is present, Liou et al (1973).

8.3.4 Applications to the Butane Hydrogenolysis Reactor

In Chapter 3, wall temperature effects were included in the reactor model. The matrices (A,G,F) are reported in Appendix 3. Using (8.35), we find that it is impossible to totally eliminate the effect of a wall temperature disturbance on the temperatures at the interior collocation points by manipulating the flow rates of butane and hydrogen. Can the effect on the effluent concentrations of a wall temperature disturbance be eliminated? The concentrations at the reactor exit are expressed as

$$\underline{c}(k) = H_I \underline{x}(k) + D_I \underline{u}(k-1) + F_I d(k-1) \quad (8.60)$$

The matrices of parameters (H_I, D_I, F_I) are reported in Appendix 3. Although (8.60) is not of the form examined thus far, conditions for the removal of the effect of the disturbance are readily obtained. A necessary condition to eliminate the effect of a wall temperature disturbance on the effluent concentrations, for all time, is that (8.53) hold with HG replaced by $H_I G + D_I$, $C=1$ and

HF replaced by $HF + F_I$. Examination of this rank condition shows that it does not hold.

The next step is to examine whether it is possible to remove the effect of a step disturbance asymptotically, using time invariant feedback controllers L_1 and L_2 . Substituting the control equation (8.36) into the state equation (8.31), with $f=0$, and (8.60), we find that for a step disturbance at $k=0$,

$$C(\infty) = (A - GL_1)^{\infty} \underline{x}(0) + (H_I - D_I L_1)(I - A + GL_1)^{-1}(-GL_2 + F)d(0) \\ + (F_I - D_I L_2)d(0) \quad (8.61)$$

If L_1 is designed to insure stability of the temperatures along the reactor, then the first term in (8.61) approaches zero asymptotically. Let $P = (H_I - D_I L_1)(I - A + GL_1)^{-1}$. To asymptotically eliminate the effect of the disturbance on the effluent concentrations, L_2 must satisfy (for any value of $d(0)$),

$$(PG + D_I)L_2 = PF + F_I \quad (8.62)$$

A necessary and sufficient condition for a solution to (8.62) is that the

$$\text{rank}(PG + D_I | PF + F_I) = \text{rank}(PG + D_I) \quad (8.63)$$

In a previous study, Jutan et al (1977c), a time invariant feedback controller, L_1 , was designed using an infinite time linear quadratic technique to regulate the effluent concentrations.

Having now designed a state feedback controller to compensate for impulse type disturbances (or equivalently, initial value departures of the state from an equilibrium value), is it possible to find an L_2 to asymptotically eliminate the offset in the effluent concentrations arising from a step disturbance in the wall temperature? Using L_1 given in Appendix 3, we find that

$$(PG+D_I | F_I+PF) = \begin{bmatrix} .064 & -.022 & | & .279 \\ .106 & -.035 & | & .908 \\ .019 & -.008 & | & 1.191 \end{bmatrix}$$

The rank of this matrix is three. Thus (8.63) does not hold and it is impossible to eliminate (given the choice for L_1), the offset in all three effluent concentrations. It is, however, possible to totally eliminate the offset in two of the species and minimize the offset in the third species by choosing

$$\underline{u}^D(k) = -(PG+P_I)^{\#}(F_I+PF)d(k) \quad (8.64)$$

Since the matrices (A, G, F, H_I, D_I, F_I) are obtained from a linearization, and discretization technique, it is doubtful that the rank conditions would ever exactly hold.

This concludes a rather lengthy discussion on the design of controllers to compensate for deterministic disturbances. As we have seen, once the disturbance has been modelled, the structure of the controller that will compensate to this type of upset is

immediately specified. The important aspect in the design of these controllers is, therefore, to adequately model the process disturbances.

8.4 Design of the Feedback Gains Using a Quadratic Criterion

The feedback controller gains $L_1(k)$ and $L_2(k)$ must be specified. Johnson (1971) suggests that $L_1(k)$ be chosen to regulate the undisturbed process and $L_2(k)$ be chosen to satisfy (8.34). It is of interest to examine the specification of these gains by minimizing a quadratic performance criterion.

The dynamic model of the process and disturbances (8.29-8.32) can be written more compactly as

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{z}(k+1) \end{bmatrix} = \begin{bmatrix} A & FC \\ 0 & D \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{z}(k) \end{bmatrix} + \begin{bmatrix} G \\ 0 \end{bmatrix} \underline{u}(k-f) + \begin{bmatrix} 0 \\ \Gamma \end{bmatrix} \underline{\alpha}(k+1) \quad (8.65)$$

and

$$\underline{y}(k) = \begin{bmatrix} H & 0 \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{z}(k) \end{bmatrix} \quad (8.66)$$

or

$$\underline{x}^*(k+1) = A^* \underline{x}(k) + G^* \underline{u}(k-f) + \Gamma^* \underline{\alpha}(k+1) \quad (8.67)$$

and

$$\underline{y}(k) = H^* \underline{x}^*(k) \quad (8.68)$$

We wish to find the sequence of controls $\{\underline{u}(1), \underline{u}(2), \dots, \underline{u}(N-1)\}$ to minimize the objective function

$$J_2 = \frac{1}{N} \{ \underline{x}^{*'}(N) Q^*(N) \underline{x}^*(N) + \sum_{k=\phi}^{N-1} (\underline{x}^{*'}(k) Q^*(k) \underline{x}^*(k) + \underline{u}'(k-f-1) R \underline{u}(k-f-1)) \} \quad (8.69)$$

We recognize that the disturbances are uncontrollable states in our process and a feedback controller can only try and compensate for their effects on the process. Thus $Q^*(k)$ should have the following structure

$$Q^*(k) = \begin{bmatrix} Q(k) & 0 \\ 0 & 0 \end{bmatrix} \quad (8.70)$$

where $Q(k)$ is an $n \times n$ positive semidefinite matrix. When $Q(k)$ is time invariant, we find that, by making use of the partitioned structure of $(-A^*, G^*, Q^*)$, the solution which minimizes (8.69) is given by

$$\underline{u}(k) = -L_1(k) \underline{x}(k+f) - L_2(k) \underline{z}(k+f) \quad (8.71)$$

with

$$L_1(k) = (R + G^* S_{11}(k+1) G)^{-1} G^* S_{11}(k+1) A \quad (8.72)$$

and

$$L_2(k) = (R + G'S_{11}(k+1)G)^{-1}G'(S_{11}(k+1)FC + S_{12}(k+1)D) \quad (8.73)$$

The $n \times n$ matrix $S_{11}(k+1)$ satisfies the matrix Riccati equation

$$S_{11}(k) = A'S_{11}(k+1)(A - GL_1(k)) + Q \quad (8.74)$$

with starting value $S(N) = Q(N)$.

This is the Riccati equation one would solve for the process with no disturbances. After some manipulation, one finds that the $n \times n_d$ matrix $S_{12}(k+1)$ satisfies the matrix difference equation

$$S_{12}(k) = (A - GL_1(k))'(S_{11}(k+1)FC + S_{12}(k+1)D) \quad (8.75)$$

with starting value $S_{12}(N) = 0$. As we can see, there is no need to solve the complete matrix Riccati equation associated with the augmented system (8.67).

In most applications, one is interested in the steady state solution for the controller gains, that is, the limiting solution of (8.69) as $N \rightarrow \infty$. Some care must be taken to insure that this limit exists. Solutions of the matrix Riccati equations have been studied, and Kwakernaak and Sivan (1972) summarize the results. A sufficient condition for the matrix Riccati equation to converge to a unique positive definite value, $S_{11}(\infty)$, is that the pair (A, G) be controllable and $|S(N)| > 0$. When these conditions are satisfied, $A - GL_1(\infty)$ has all its eigenvalues inside the unit circle. The condition of controllability can be relaxed to

stabilizable if the uncontrollable states of A are stable. If Q is taken as $H'Q^{**}H$, where Q^{**} is a weighting matrix for the process outputs, then (A,H) must be observable in addition to the above requirements. Observability can be replaced by detectability if the unobservable states are stable. We will assume that these conditions are satisfied.

If $S_{11}(k+1)$ converges to a steady state value, then (8.75) can be considered at some point in time to be a constant coefficient difference equation with non-zero initial conditions. If the modulus of the eigenvalues of D are less than one, (8.76) will converge to unique value. If the eigenvalues of D have modulus greater than or equal to one, $S_{12}(k)$ may still converge, although the solution must be examined for the particular choice of D .

The convergence of $S_{12}(k)$ does not insure that we eliminate the effect of the disturbances in the states (or process outputs), or insure that our objective function (8.69) is bounded. If we have used a time invariant feedback controller, the solution to the state equations (8.67) when $\underline{a}(k)$ is zero for all time is

$$\underline{x}(k) = (A - GL_1)^k \underline{x}(0) + \sum_{j=0}^{k-1} (A - GL_1)^j (-GL_2 + FC) D^{k-1-j} \underline{z}(0) \quad (8.76)$$

The first term approaches zero asymptotically, if (A,G) is controllable. If D has its eigenvalues inside the unit then

$\underline{x}(k) \rightarrow 0$ as k gets large regardless of the choice for $L_2(\infty)$. This may appear to be an uninteresting example since the state regulator portion of the control $\underline{u}^R(k)$ will eventually return the state to the origin without the help of $\underline{u}^D(k)$. The effect of the latter component is to reduce the influence of the disturbance over the time interval in which it affects the process.

If the disturbance is nonstationary, the offset in the states cannot be removed by using the infinite time objective function (8.69) because the deviations of the manipulated variables from fixed levels are penalized. (One of the techniques proposed by Davison (1972) to compensate for external disturbances involves minimization of an objective function where \underline{u}^R is constrained.) The offset can be removed when $R=0$, in which case it can be shown that

$$L_2(k) = (G'S_{11}(k+1)G)^{-1}G'S_{11}(k+1)FC \quad (8.77)$$

This equation is of the form (8.37) with $W=S_{11}(k+1)$. However, if the disturbances do not grow in time (step disturbances are included in this class), the objective function (8.69) will be bounded for all R , $|R| > 0$, since we have divided by N .

For all nonstationary disturbances, $\underline{u}(k-f-1)$ in the objective function (8.69) must be replaced by a function of the form

$$\underline{u}^*(k-f-1) = \psi(z^{-1})\underline{u}(k-f-1) \quad (8.78)$$

where $\psi(z^{-1})$ is a matrix polynomial in the backward shift operator

z^{-1} . The structure of $\psi(z^{-1})$ is obtained by noting that if the offset is to be eliminated, the variable $WGu(k)$, where W is the generalized inverse of F , must satisfy the same difference equation as the disturbance $d(k)$. To use the linear quadratic criterion directly, the process model must be expressed in terms of the manipulated variable $u^*(k)$. It is not always straight forward to express the model in the variable $u^*(k)$. Important exceptions occur when the rank of F is n , when there is a single manipulated variable and disturbance, or when we wish to guard against step disturbances. If the disturbance contains stationary and nonstationary components, only the nonstationary components must be included in the definition of $u^*(k)$. The manner in which this is accomplished is not obvious.

Penalizing the deviations from fixed levels $(u'(k-f-1)Ru(k-f-1))$, or changes in the control action $(\nabla u'(k-f-1)R\nabla u(k-f-1))$ may be meaningful. However, when we replace $u'(k-f-1)Ru(k-f-1)$ in (8.69) by an arbitrary function of the manipulated variable, it is difficult to give a physical interpretation to the objective function.

In this section, we have explored the use of a linear quadratic criterion to specify the controller gains. If the disturbances are nonstationary, their effect cannot be eliminated using an infinite time linear quadratic criterion when a constraint is placed on the magnitude of the actual level of the control. What is the advantage of using an LQ strategy to specify L_2 ? In

instances where it is impossible to eliminate the effect of the disturbances using an algebraic technique, i.e., the rank conditions do not hold, we can use an LQ strategy to minimize the effect of the disturbances. Although the discussion referred to state control, similar comments apply to output control where one replaces Q by $H^*Q^{**}H$.

8.4 Reconstruction of the States and Disturbances by Observers

The problem was originally posed as one where the disturbances were unmeasurable. For deterministic process, the states and disturbances can be reconstructed using an observer, Luenberger (1964). Observers for the augmented system (8.65) and (8.66) can be constructed as, Johnson (1970a)

$$\hat{\underline{x}}^*(k+1) = A^*\hat{\underline{x}}^*(k) + G^*\underline{u}(k-f) + K(\underline{y}(k) - \hat{\underline{y}}(k)) \quad (8.79)$$

and

$$\hat{\underline{y}}(k) = H^*\hat{\underline{x}}^*(k) \quad (8.80)$$

$\hat{\underline{x}}^*(k)$ denotes the estimate of our process and disturbance states.

The matrix K is chosen so that the system of equations

$$\underline{x}^*(k+1) - \hat{\underline{x}}^*(k+1) = (A^* + KH^*)(\underline{x}^*(k) - \hat{\underline{x}}^*(k)) + r^*\underline{a}(k+1) \quad (8.81)$$

are asymptotically stable, and have desired transient behaviour.

To reconstruct the states and disturbances, the pair (A^*, H^*) must be completely observable. O'Reilly (1979) has shown

that necessary and sufficient conditions for the pair (A^*, H^*) to be completely observable are:

1. (A, H) be completely observable;
2. (C, D) be completely observable;
3. $\text{rank} \begin{bmatrix} A & F \\ H & 0 \end{bmatrix} = n + n_d$ (8.82)

A necessary condition for (8.82) to hold is that $m \geq n_d$, i.e., the number of outputs must equal or exceed the dimension of the unmeasured disturbances.

Instead of using $\underline{x}(k)$ and $\underline{w}(k)$ in the controller, $\hat{\underline{x}}(k)$ and $\hat{\underline{w}}(k)$; as determined by the observer are substituted. It is straight forward, as Johnson (1970a) has shown, to prove a certainty equivalence principle for this system. This guarantees that the poles of the closed loop system can be assigned arbitrarily using the estimated values instead of the true values of the state and disturbance.

It is instructive to substitute the equation for the observer into the control equation (8.36). We obtain

$$\begin{aligned} \underline{u}(k) &= -L_1 \hat{\underline{x}}(k) - L_2 \hat{\underline{z}}(k+f/k) \\ &= -L_1 \hat{\underline{x}}(k) - L_2 D^f (D^k \hat{\underline{z}}(0) + \sum_{j=0}^{k-1} D^{k-1-j} K^D (\underline{y}(j) - \hat{\underline{y}}(j))) \end{aligned} \quad (8.83)$$

where K^D is the matrix of elements from K that correspond to the

disturbance observer. $\hat{z}(k+f)$ has been replaced by a forecast $\hat{z}(k+f/k)$, and this has been taken as $D^f \hat{z}(k)$. Assume for the moment that we wish to guard against a step, and that $\hat{z}(0)=0$. Then the control is given by

$$\underline{u}(k) = -L_1 \hat{x}(k) - L_2 K^D \sum_{j=0}^{k-1} (y(j) - \hat{y}(j)) \quad (8.84)$$

We recognize this as a multivariable proportional plus integral type controller.

The key to designing a controller to guard against a particular form of disturbance is to model the disturbance correctly and include it in the description of the process.

8.5 The Certainty Equivalence Property in Stochastic Control

In section 8.2, we saw that both stochastic and deterministic disturbances could be represented by equations (8.29) and (8.30). The essential difference between the stochastic and deterministic representations is that the 'occasional' shock $\underline{\alpha}(k)$, is replaced by a sequence of shocks $\underline{\alpha}(k)$ occurring at every control interval. The 'infrequent' impulses $\underline{\alpha}(k)$ were completely ignored in both the controller design and construction of the observer. The rationale for this, is that if the impulses are far enough apart in time then we can effectively reconstruct the states of the process and compensate for the disturbance before the next impulse arrives.

For linear processes subject to additive disturbances of the form (8.29) and (8.30), controllers designed to minimize the objective function

$$J = \frac{1}{N} E\{\underline{x}^*(N)' Q^*(N) \underline{x}^*(N) + \sum_{k=0}^{N-1} (\underline{x}^*(k)' Q^*(k) \underline{x}^*(k) + \underline{u}'(k-f-1) R \underline{u}(k-f-1))\} \quad (8.85)$$

are of the form

$$\underline{u}(k) = -L(k) \underline{\hat{x}}^*(k/\tau) \quad (8.86)$$

where $E\{\}$ is the mathematical expectation, $L(k)$ is the solution to the deterministic regulator problem (8.69), $\underline{\hat{x}}^*(k/\tau)$ is the conditional mean of the state vector and $E(\underline{x}^*(k)/Y^\tau)$. $Y^\tau = (\underline{y}(\tau), \underline{y}(\tau-1), \dots, \underline{y}(0))$, is the data available for computing the control action at time k . Equation (8.86) is known as the certainty equivalence theorem. This result is valid regardless of the distribution of the $\alpha(k)$'s, Root (1969).

If the disturbances, stochastic or deterministic, can be modelled by the equations of the form (8.29) and (8.30), a controller designed using linear quadratic techniques will depend on the C and D matrices and not the Γ matrix. As we recall, if C and D are a state representation of the disturbance difference equations (8.6) or (8.7), these matrices depend only on the autoregressive terms. Thus the structure of a linear quadratic controller designed to guard against disturbances of the form (8.7)

will be identical to that designed to guard against disturbances modelled by (8.26), where there are no moving average parameters. As we shall see, the moving average parameters are required to compute the conditional mean $\underline{x}^*(k/\tau)$. However, computation of the conditional mean is usually only tractable when the shocks are from a Gaussian distribution. In this case, the equations describing the time evolution of the conditional mean are commonly referred to as the Kalman filter equations.

8.6 Control of Processes Subject to Both Deterministic and Stochastic Disturbances

As has been outlined in the previous sections, the control of processes subject to only stochastic disturbances or to only deterministic disturbances is straightforward and the two problems are the duals of one another in that their structures and solutions are identical. The state estimation problem is also straightforward for both of these cases. The Kalman Filter for the stochastic case and the ^{Luven}Leunburger type observers for the deterministic problem are straightforward both in a theoretical and a practical sense. The key to all approaches is to incorporate into the model of the system an adequate model describing the behaviour of the disturbances.

It will soon become apparent in this section that the solution of the state estimation problem for processes subject to both stochastic and randomly occurring deterministic disturbances

is much more difficult from both a theoretical and practical viewpoint. These difficulties will be seen to result from the fact that with both types of disturbances, the system is no longer Gaussian, nor is it unperturbed for significant lengths of time, conditions which eliminate the linear Kalman filter or the deterministic observers as efficient methods for state estimation.

8.6.1 An Example

Consider the following process

$$x(k+1) = .5x(k) + z(k) + w(k)$$

$$z(k+1) = z(k) + \alpha(k) \quad (8.87)$$

$$y(k) = x(k) + v(k) \quad (8.88)$$

$w(k)$ and $v(k)$ are from a normally distributed population, each with mean 0 and variance 1.0, $N(0,1)$. $z(k)$ has been included to detect occasional step disturbances. We must specify the distribution of $\alpha(k)$ in order to develop a filter. One approach, which might seem reasonable, is to assume that $\alpha(k)$ is normally distributed, with mean 0 and variance proportional to the anticipated magnitude of the step disturbance. To see how effective this technique is, the process (8.87-8.88) was simulated twenty times, for one hundred time steps. To generate the process output, the distribution of $w(k)$ was $N(0,1)$. $\alpha(k)$ was zero for all time steps except at $k=50$ where $\alpha(50)=1$. To estimate the states $(x(k), z(k))$, $w(k)$ and $\alpha(k)$ were each assumed to be distributed as $N(0,1)$.

The ability of the Kalman filter (Equations 3.14-3.18) to reconstruct the states is shown in Figure 8.2. The step is detected very quickly. However, the estimate of this state is very noisy. As the variance of $\alpha(k)$ used in the filter decreases, the estimate of $z(k)$, becomes smoother at the expense of much slower estimation of $z(k)$ and poorer estimation of $x(k)$. Reconstruction of the states for variance of $\alpha(k) = .001$ is shown in Figure 8.3. One might seek a compromise by adjusting the variance of $\alpha(k)$.

The specification of $\alpha(k)$ as having a normal distribution is the cause of these problems. For the approximation, $\text{var} \{ \alpha(k) \} = .001$ most of the mass in the probability density function for $\alpha(k)$ is concentrated about 0. Large excursions in $y(k) - \hat{y}(k)$, (the driving force for the Kalman Filter) do not result in quick changes in the state estimate for $z(k)$, due to the low probability that these departures originate from a non-zero value of $\alpha(k)$. Whereas this leads to good estimates of $x(k)$ and $z(k)$, when no step has occurred ($k < 50$ in Figure 8.3), it results in poor estimates when a step occurs. When we approximate $\text{var} \{ \alpha(k) \} = 1.0$, the mass of the probability density function for $\alpha(k)$ is not as concentrated about 0 as before. Thus, the estimates of $z(k)$ and $x(k)$ are more sensitive to excursions in $y(k) - \hat{y}(k)$, and this explains why the Kalman Filter quickly detects the step change in $z(k)$. The penalty paid is in poor estimates for $z(k)$, when $\alpha(k)$ is zero.

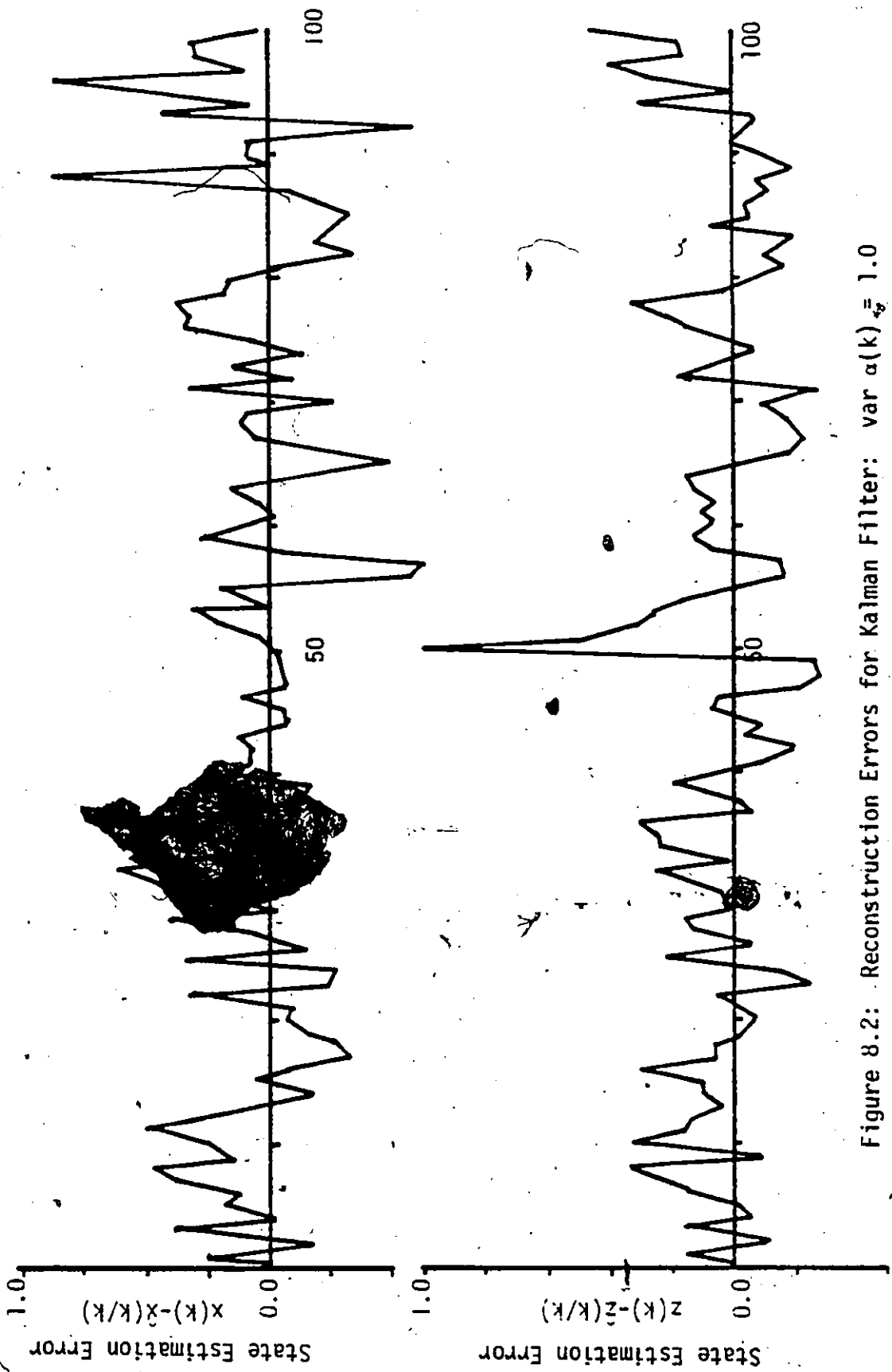


Figure 8.2: Reconstruction Errors for Kalman Filter: $\text{var } \alpha(k) = 1.0$

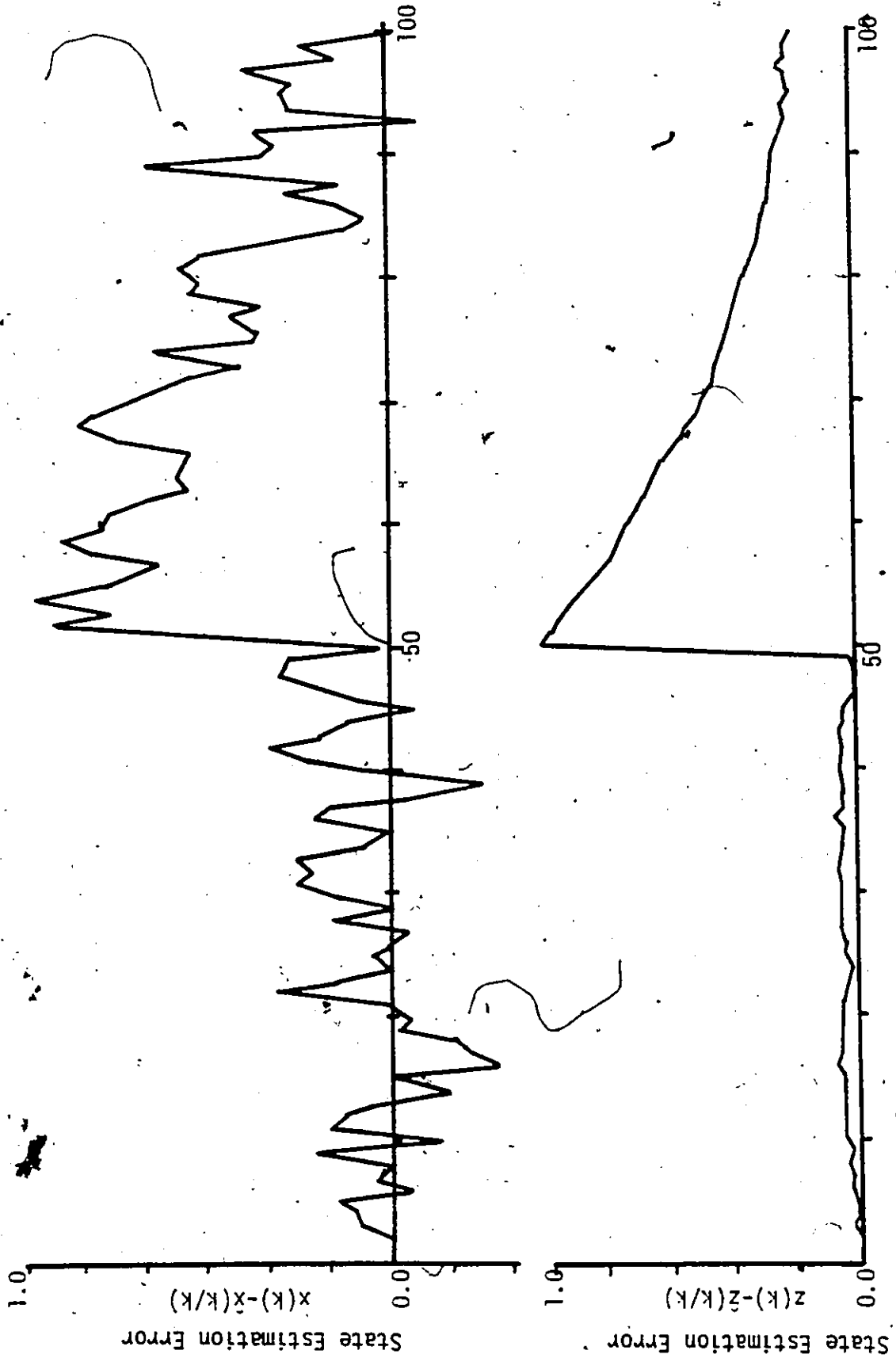


Figure 8.3: Reconstruction Errors for Kalman Filter: $\text{var } \alpha(k) = .001$

This example has been presented to illustrate some of the problems that can be encountered in state estimation of processes subject to both stochastic and deterministic disturbances. In the next few sections, we will examine several approaches that might be used to obtain better state estimates. In the first, a more reasonable probability distribution for $\alpha(k)$ is proposed and this is used to develop a nonlinear filtering algorithm to estimate the states. In the second approach, the occurrence of a deterministic disturbance is treated as a transition between models generating the process output. If we can detect a transition, it is possible to switch to a different Kalman Filter. This technique is known as 'jump' or 'event' detection.

8.6.2 A Distribution for $\alpha(k)$

In the previous section, we saw that the accuracy and variance of the state estimates depended on the assumed normal distribution of $\alpha(k)$ and its variance. In this section, we examine the use of a more reasonable distribution for $\alpha(k)$ and outline a technique for estimating states in the presence of stochastic and deterministic disturbances.

Deterministic disturbances occur when there is a non-zero value of $\alpha(k)$. If we suppose that these disturbances only affect the process occasionally, then $\alpha(k)$ is zero most of the time. However, there is a small probability that $\alpha(k)$ is non-zero, and it is these occasional non-zero values we wish to

detect. The distribution for $\alpha(k)$ is highly concentrated about zero, with long tails, Figure 8.4. This distribution for $\alpha(k)$ might be represented as

$$p(\alpha(k)) = BS(\alpha(k)) + (1-B)\phi^*(\alpha(k), \sigma_2^2) \quad (8.89)$$

where $S(t-a)$ is the delta function, which is zero everywhere except at $t = a$. $\phi^*(t-b, \sigma^2)$ is a shorthand notation for the probability density function of the normal distribution with mean b , and variance σ^2 . In (8.89), σ_2^2 is proportional to the magnitude of an expected step. The parameter B , $0 \leq B \leq 1$, is the weight we assign to each distribution based on the relative frequency of the deterministic disturbances. If we thought that a step entered one time in twenty, B would have the value .95.

The double exponential distribution might adequately describe the distribution of $\alpha(k)$ since it also is very concentrated about its mean with long tail areas. In either case, we note that the distribution for $\alpha(k)$ is highly non-normal.

The reconstruction of the states $\underline{x}(k)$ from observations, $Y(k) = \{y(k), y(k-1), \dots, y(0)\}$, is a parameter estimation problem. The statistical properties of these estimates, i.e., expected values and variances, depend on the distribution of $\underline{x}(k)$ and $Y(k)$. The conditional distribution $p(\underline{x}(k)/Y(k))$ contains all the information about $\underline{x}(k)$ coming from the data set $Y(k)$. A point estimate of $\underline{x}(k)$ can be obtained from this distribution

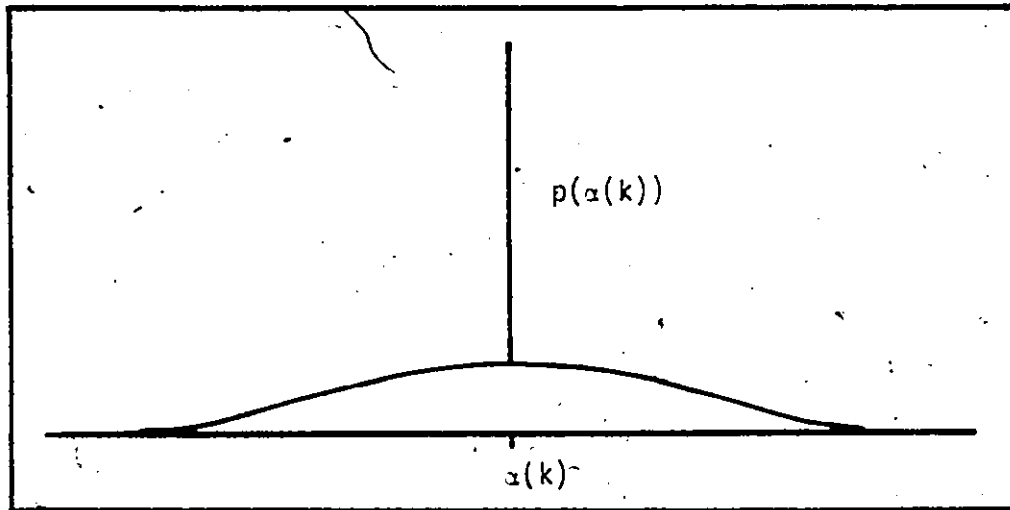


Figure 8.4: A Distribution for $a(k)$

1. Mode of $p(\underline{x}(k)/Y(k))$
2. Mean of $p(\underline{x}(k)/Y(k))$
3. Median of $p(\underline{x}(k)/Y(k))$

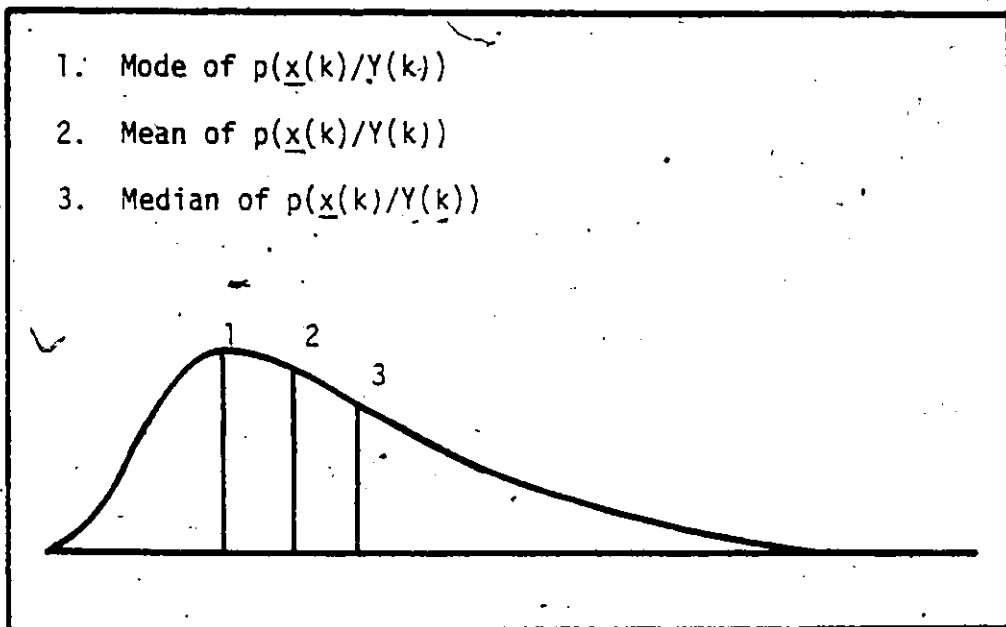


Figure 8.5: Estimates for $\underline{x}(k)$

(for example, its mean or mode). We defer for the moment, how to obtain this distribution.

Consider the distribution $p(\underline{x}(k)/Y(k))$ shown in Figure 8.5.

Three common estimates for $\underline{x}(k)$, shown in Figure 8.5 are:

- i) $\hat{\underline{x}}(k) = \text{mode of } p(\underline{x}(k)/Y(k))$. This maximizes the probability that $\hat{\underline{x}}(k) = \underline{x}(k)$, and is known as the most probable estimate;
- ii) $\hat{\underline{x}}(k) = \text{mean of } p(\underline{x}(k)/Y(k))$. This minimizes the mean square error $E\{(\hat{\underline{x}}(k) - \underline{x}(k))'(\underline{x}(k) - \underline{x}(k))\}$, and is known as the conditional mean;

and

- iii) $\hat{\underline{x}} = \text{median of } p(\underline{x}(k)/Y(k))$. This minimizes the maximum $|\underline{x}(k) - \hat{\underline{x}}(k)|$, and is known as the mini - max estimate.

When the process and measurement statistics are Gaussian, these estimates coincide. The Kalman Filter equations can be derived from an algebraic argument (see Astrom (1970), without having to explicitly compute $p(\underline{x}(k)/Y(k))$, if $\underline{w}(k)$ and $\underline{v}(k)$ have finite second moments. The Kalman Filter is the best linear predictor for $\underline{x}(k)$, in the sense of minimizing the mean square error, Astrom (1970). Even in the case where $\underline{w}(k)$ and $\underline{v}(k)$ are non-normal, we can justify the use of a Kalman Filter simply as a measure of closeness of fit. However, when $\underline{w}(k)$ and $\underline{v}(k)$ are not Gaussian, there exist other estimators which have a smaller mean square prediction error, i.e., the condition mean of the non-normal $p(\underline{x}(k)|Y(k))$. Thus, there is an incentive to

seek other state estimators when the disturbances have highly non-Gaussian distribution.

The proposed density function for $\alpha(k)$ in (8.89) is not a purely continuous or discrete density function. It is, however, a properly defined density function of the general type, Gnedenko (1962). The moments of this distribution are defined in the usual manner

$$u_i = \int_{-\infty}^{\infty} (\alpha(k))^i p(\alpha(k)) d\alpha(k) \quad (8.90)$$

To integrate the delta function, we must interpret the right hand side of (8.90) as a Stiltjes integral, Gnedenko (1962), Neuts (1973). When the integrand is a smooth function, a Stiltjes integral is identical to a Riemann integral, the 'usual' interpretation of integration. Functions or transformations of random variables having such a general pdf are difficult to evaluate, and are most readily computed using the distribution function of the random variable. If it were possible to approximate the delta function by a smooth function, the pdf of $\alpha(k)$ would be purely continuous, for which transformations of variables are easily evaluated.

The pdf of a normally distributed variable approaches the delta function as the variance approaches zero. The distribution shown in Figure 8.4 can be very closely approximated by

$$p(\alpha(k)) = B\phi^*(\alpha(k), \sigma_1^2) + (1-B)\phi^*(\alpha(k), \sigma_2^2) \quad (8.91)$$

where σ_1^2 is small relative to σ_2^2 , but not zero. This distribution for $\alpha(k)$ is known as a Gaussian sum, Sorenson and Alspach (1971), or a mixture of Gaussians. Having described a distribution for $\alpha(k)$, we can now try to evaluate the conditional distribution $p(\underline{x}(k)/Y(k))$.

The conditional pdf of $(\underline{x}'(k)/Y(k))$ can be derived from a Bayesian argument, Ho and Lee (1964). If at time k , $p(\underline{x}(k)/Y(k))$ is known, $p(\underline{x}(k+1)/Y(k+1))$ is given by, Sorenson and Alspach (1971)

$$p(\underline{x}(k+1)/Y(k+1)) = \frac{p(\underline{x}(k+1)/Y(k))p(\underline{y}(k+1)/\underline{x}(k+1))}{p(\underline{y}(k+1)/Y(k))} \quad (8.92)$$

where

$$p(\underline{x}(k+1)/Y(k)) = \int p(\underline{x}(k)/Y(k))p(\underline{x}(k+1)/\underline{x}(k))d\underline{x}(k) \quad (8.93)$$

The normalizing constant $p(\underline{y}(k+1)/Y(k))$ is given by

$$p(\underline{y}(k+1)/Y(k)) = \int p(\underline{x}(k+1)/Y(k))p(\underline{y}(k+1)/\underline{x}(k+1))d\underline{x}(k+1) \quad (8.94)$$

The initial density $p(\underline{x}(0)/\underline{y}(0))$ is

$$p(\underline{x}(0)/\underline{y}(0)) = \frac{p(\underline{y}(0)/\underline{x}(0))p(\underline{x}(0))}{p(\underline{y}(0))} \quad (8.95)$$

All integrations are n -fold over the domain of $\underline{x}(k)$, or $\underline{x}(k+1)$.

$p(\underline{x}(k+1)/Y(k))$ is the distribution for $\underline{x}(k+1)$ given only information to time k , and can be thought of as the a priori,

or prior distribution for $\underline{x}(k+1)$. New information about $\underline{x}(k+1)$ is provided by the measurement $\underline{y}(k+1)$. The posterior distribution for $\underline{x}(k+1)$, $p(\underline{x}(k+1)/Y(k+1))$, is obtained by weighting the prior distribution with this new information via the likelihood function $p(\underline{y}(k+1)/\underline{x}(k+1))$.

For nonlinear processes, or processes with non-Gaussian process and measurement noise, it is difficult, if not impossible to obtain a closed form solution to (8.92-8.94). However, these density functions can be evaluated explicitly for linear processes subject to Gaussian sum disturbances, Sorenson and Alspach (1971). These authors used Gaussian sums to approximate arbitrary density functions such as rectangular distributions, as one means of evaluating (8.92-8.94) for such non-normal distributions. Although they evaluated their equations for a single state and one measurement, evaluation of these density functions for multivariate processes is a straightforward, but tedious extension of their work.

Consider the process

$$\underline{x}(k+1) = A\underline{x}(k) + \underline{w}(k) \quad (8.96)$$

and

$$\underline{y}(k) = H\underline{x}(k) + \underline{v}(k) \quad (8.97)$$

$\underline{w}(k)$ has the Gaussian sum pdf

$$p(\underline{w}(k)) = \sum_{j=1}^2 B_j \phi^*(\underline{w}(k), \underline{\Sigma}_j) \quad (8.98)$$

where $\phi(\underline{z}-\underline{\mu}, \underline{\Sigma})$ is an abbreviation for the pdf of an n-variate normal random variable \underline{z} , that is

$$\phi^*(\underline{z}-\underline{\mu}, \underline{\Sigma}) = \frac{1}{(2\pi)^{n/2} |\underline{\Sigma}|^{1/2}} \exp. \{ -\frac{1}{2} (\underline{z}-\underline{\mu})' \underline{\Sigma}^{-1} (\underline{z}-\underline{\mu}) \} \quad (8.99)$$

The measurement noise $\underline{v}(k)$ has the pdf

$$p(\underline{v}(k)) = \phi^*(\underline{v}(k), \underline{\Sigma}_v) \quad (8.100)$$

The controls $\underline{u}(k)$ have been temporarily omitted from the state description (8.96) without loss of generality.

Let us assume for the moment that at time k , $p(\underline{x}(k)/Y(k))$ can be represented by the Gaussian sum

$$p(\underline{x}(k)/Y(k)) = \sum_{i=1}^{n(k)} \varepsilon_i(k) \phi^*(\underline{x}(k) - \underline{\mu}_i(k), \underline{\Sigma}_i(k)) \quad (8.101)$$

where $\varepsilon_i(k)$ is the weight assigned the i th normal distribution at time k . These weights must sum to 1.0. The mean $\underline{\mu}_i(k)$ and $\underline{\Sigma}_i(k)$ are as yet unspecified. The bracketed term (k) emphasizes that these quantities may change with time. The prior distribution for $\underline{x}(k+1)$ is computed from (8.93). Evaluating this expression we find that $p(\underline{x}(k+1)/Y(k))$ is also a Gaussian sum with the pdf

$$p(\underline{x}(k+1)/Y(k)) = \sum_{i=1}^{n(k)} \sum_{j=1}^2 \epsilon_i(k) B_j \phi^*(\underline{x}(k+1) - \underline{\mu}_{i,j}, \hat{\Sigma}_{i,j}) \quad (8.102)$$

where

$$\underline{\mu}_{i,j} = A \underline{\mu}_i(k) \quad (8.103)$$

and

$$\hat{\Sigma}_{i,j} = A \hat{\Sigma}_i(k) A' + \hat{\Sigma}_{w,j} \quad (8.104)$$

The posterior density function $p(\underline{x}(k+1)/Y(k+1))$ is computed from (8.92) and (8.94). We find that it is also a Gaussian sum with pdf

$$p(\underline{x}(k+1)/Y(k+1)) = \sum_{i=1}^{n(k)} \sum_{j=1}^2 \hat{\epsilon}_{i,j} \phi^*(\underline{x}(k+1) - \hat{\underline{\mu}}_{i,j}, \hat{\Sigma}_{i,j}) \quad (8.105)$$

with

$$\hat{\underline{\mu}}_{i,j} = \underline{\mu}_{i,j} + \hat{\Sigma}_{i,j} H' (H \hat{\Sigma}_{i,j} H' + \hat{\Sigma}_v)^{-1} (y(k+1) - H \underline{\mu}_{i,j}) \quad (8.106)$$

and

$$\hat{\Sigma}_{i,j}^{-1} = \hat{\Sigma}_{i,j}^{-1} + H' \hat{\Sigma}_v^{-1} H \quad (8.107)$$

The weighting coefficient $\hat{\epsilon}_{i,j}$ is determined from

$$\hat{\epsilon}_{i,j} = \frac{\hat{\epsilon}_{i,j}}{\sum_{i=1}^{n(k)} \sum_{j=1}^2 \hat{\epsilon}_{i,j}} \quad (8.108)$$

with $\hat{\epsilon}_{i,j}$ defined as

$$\hat{\epsilon}_{i,j} = \epsilon_i(k) B_j \frac{|\hat{\Sigma}_{i,j}|^{1/2}}{|\hat{\Sigma}_{i,j}|^{1/2}} \exp\left(-\frac{1}{2}(\underline{y}(k+1) - H\hat{\underline{x}}_{i,j})' (\hat{\Sigma}_V + H\hat{\Sigma}_{i,j}H')^{-1} (\underline{y}(k+1) - H\hat{\underline{x}}_{i,j})\right) \quad (8.109)$$

The conditional mean, $\hat{\underline{x}}(k+1)/k+1$ is given by

$$\hat{\underline{x}}(k+1/k+1) = \sum_{i=1}^{n(k)} \sum_{j=1}^2 \hat{\epsilon}_{i,j} \hat{\underline{x}}_{i,j} \quad (8.110)$$

In spite of the apparent complexity of these equations, the conditional mean, (8.110) has a nice interpretation. It is the weighted sum of a growing number of individual estimates for the conditional mean of each term in the Gaussian sum. Each of these individual estimates evolves in the same fashion as the Kalman filter estimate, Equations (3.14-3.18). The weight attached to each estimate, (8.108) and (8.109) depends on the data. The exponential in (8.109) is proportional to the likelihood that the (k+1)st observation, $\underline{y}(k+1)$ is generated from a state space model with mean $\underline{x}(k+1) = H\hat{\underline{x}}_{i,j}$ and covariance $\hat{\Sigma}_V + H\hat{\Sigma}_{i,j}H'$.

Due to the growing number of terms, it is impossible to use this approach directly. Sorenson and Alspach (1971) have proposed that the conditional density function $p(\underline{x}(k+1)/Y(k+1))$, in (8.105), be approximated by a Gaussian sum involving fewer terms.

The parameters of this approximate distribution are obtained by minimizing some measure of closeness between the two distributions. For multivariate distributions these techniques involve adjusting for too many parameters to be of practical use unless most of the parameters describing the approximated distribution are preselected.

A simple technique would be to approximate the posterior by a single, normal distribution of each stage. The posterior at the next time step would be described as a sum of two Gaussians. This again would be approximated by a single Gaussian whose mean and covariance would be the weighted average of the individual means and covariances. The state estimate would then evolve from a set of equations whose mathematical structure was identical to the Kalman filter equations. The difference is that the Kalman gain $K(k)$ would now be a function of the data. This approach was tried and the results were very disappointing. Excellent state estimates were obtained when there was no step. When a step occurred, the behaviour of the state estimates was very similar to that shown in Figure 8.3. It was felt that this poor performance was due to approximating the posterior by a single normal distribution.

As a last remark in this section, we note that had we retained the original description for the distribution of $p(\alpha(k))$ in (8.89), we would have found that the posterior

distribution was also described by a Gaussian sum. Thus, the approximation of (8.89) by (8.91) does not change the essential features of the time evolution of the posterior distribution.

8.6.3 Robust Estimation of the Conditional Mean

Masreliez (1975), and Masreliez and Martin (1977) considered 'robustifying' the Kalman filter. Their intentions were to improve estimation of the process states, when $w(k)$, was more heavily tailed than the normal distribution. In this way, they hoped to obtain better state estimates when the process was subject to occasional step or impulse disturbances. The process disturbance was modelled as a Gaussian sum. A stochastic approximation algorithm involving a nonlinear gain was used to estimate the states. The nonlinear gain was determined from the influence function, which plays a role in robust estimation theory. Unfortunately, the algorithm requires that $H'R_vH$ be of full rank and this is not the case when the disturbances affecting the process are not measured. If, however, a measured state is subject to an occasional step change (and $(H'R_vH)^{-1}$ exists), their algorithm provides an alternative means of estimating the states.

8.6.4 An "Event" Detection Approach

In a previous section, the pdf of the $\alpha(k)$'s was modelled as the sum of two normals. Alternatively, the occurrence of a deterministic disturbance could be thought of as a change

in the probability density function of the process disturbance. From an analysis of the observations, one would hope to quickly detect this transition, and then switch to an estimation algorithm more appropriate for the new environment. When no deterministic disturbances are present, $\underline{a}(k)$ is presumed to be distributed as $N(\underline{0}, \underline{\Sigma}_0)$. The covariance matrix $\underline{\Sigma}_0$ is 'small' enough, that the pdf of $\underline{a}(k)$ is nearly a delta function. When a deterministic disturbance occurs, the distribution of $\underline{a}(k)$ is assumed to change. The direction of the disturbance is usually not known in advance. A reasonable assumption, therefore, is that $\underline{a}(k)$ is normally distributed with mean zero and covariance $\underline{\Sigma}_1$. The covariance matrix $\underline{\Sigma}_1$ is now 'large' enough so as to admit the possibility of non-zero $\underline{a}(k)$'s of a magnitude that might reasonably be expected in the process.

Ackerson and Fu (1970) have investigated state estimation in switching environments. The transition between Gaussian noise models is assumed to be a Markov jump process. To each noise model a probability is assigned, based on one's degree of belief in its occurrence. The conditional mean of the state is computed, realizing that the process output could have been generated from a number of models. The result is that the conditional mean is the weighted sum of an expanding number of Kalman Filters. The recursion equations are almost identical to those developed in section 8.6.2. However, the motivation for the two approaches is much different. The growing number of terms precludes practical implementation.

A different approach was taken by Willisky and Jones (1976). Rather than model the occurrence of a large non-zero value of $\underline{a}(k)$ as a change in its distribution, the process is modelled as

$$\underline{x}(k+1) = A\underline{x}(k) + S(\theta, k+1)\underline{\gamma} + \underline{w}(k) \quad (8.111)$$

and

$$\underline{y}(k) = H\underline{x}(k) + \underline{v}(k) \quad (8.112)$$

$S(\theta, k+1)$ is an impulse which is non-zero for all time except $k=\theta$, when it has the value 1. $\underline{\gamma}$ is the magnitude of the jump. A likelihood ratio is formulated to detect the occurrence of an impulse. When one has been detected, the states are estimated by modifying the original Kalman Filter equations.

How does one distinguish between an impulse and the shocks $\underline{w}(k)$ occurring at every time step? The approach assumes that $\underline{\gamma}(k)$ is much larger than the variance of the process noise. Willisky and Jones (1976) claim that their method is suitable for cases where the magnitude of $\underline{\gamma}(k)$ is only ten times the average variance of the process noise. Their method was developed for aerospace applications, where large impulses represent a dramatic shift in altitude of an aircraft or satellite. In the reactor, large changes in the major load variables such as the oil temperature would not be this large and could also result in significant changes in the matrices (A,G), since we only have a linearized state model.

In this section, a number of different approaches for estimation of randomly occurring deterministic disturbances have been examined. The theoretical solutions preclude practical implementation. To use a filter to reconstruct the states when there are both stochastic and deterministic disturbances present, it is necessary to specify a distribution for the deterministic disturbances. The only tractable approach is to model these as from a multivariate normal with mean zero and covariance Σ_0 . In this case, the states can be estimated using the Kalman Filter equations (3.14-3.18). The covariance matrix would have to be chosen so as to represent a compromise between the speed of detection of the deterministic disturbance, and the degradation in state estimation that occurs when no deterministic disturbances are present.

8.7 Summary

In this chapter, the modelling and control of processes subject to randomly occurring deterministic disturbances have been examined. As observed, stochastic disturbances can be modelled as deterministic disturbances with adaptive coefficients. The duality between the control of deterministic and stochastic disturbances was outlined. A number of solutions to detection of deterministic changes in the presence of stochastic disturbances were also studied.

An important conclusion from this chapter, is that if one wants to design a controller to compensate for a particular process disturbance, the mathematical description of this disturbance, stochastic or deterministic, must be included in the process description. If this is done, the estimation and control of the process is readily accomplished in the framework of existing theory. However, when both types of disturbances are present, the state estimation problem is more difficult from both a practical and theoretical viewpoint.

CHAPTER 9

CONCLUSIONS

The focus in this thesis has been on the use of time series modelling and stochastic control as a means of analyzing, and designing control systems for a complex multivariate process. Specific topics were studied and these are briefly summarized.

The concept of independence of reaction was studied in Chapter 2. An attempt was made to distinguish between stoichiometric and kinetic independence of reaction. An expression was developed to explain the reaction stoichiometry in a tubular reactor. Depending on the choice of the reactor model, the number of species included in the material balance for the process may be reduced by the use of this result. The use of reactor invariants in stability analysis of tubular reactors is an area for which future investigation is required. Previous difficulties encountered in controlling a pilot scale tubular reactor can be attributed, in part, to the extreme sensitivity of the temperature profile to perturbations in the reactant flow-rates. This sensitivity emphasizes the need for powerful control strategies to regulate this process.

The material and energy balances describing the dynamics of a tubular reactor are a set of coupled nonlinear partial differential equations. The manner in which these equations are

put in a form suitable for control purposes was reviewed in Chapter 3. The control theory used in subsequent chapters was also reviewed.

On processes subject to stochastic disturbances, the location and number of measurement sensors influences the degree of control that can be exerted over the variables of interest. The optimal location of sensors was studied in Chapter 4. The theory for locating sensors in some optimal fashion closely parallels the design of experiments in the statistical literature. An application of this theory to locate thermocouples in a packed bed tubular reactor indicated that good state estimation and control could be obtained with one or two strategically located thermocouples. An essential requirement to use this theory is that a theoretical model of the process dynamics and disturbances be available.

When such a model cannot be obtained from mechanistic arguments, an empirical model can be estimated from input/output data collected from a designed experiment. For multivariate processes, a canonical structure for the dynamic/stochastic model must be specified. Although not a parsimonious representation, a joint auto-regressive structure for the input and output variables is a quick means of obtaining a dynamic/stochastic model of the process. Due to the large number of parameters that must be identified when an empirical model is fit to data, relationships among the variables may be obscured.

The use of principal component analysis and canonical variate analysis to time series were shown to be powerful methods for analyzing complex linear(ized) dynamic systems. Application of these techniques to experimental data revealed that the variability in the reactor temperature profile (as reconstructed from seven independently measured axial temperatures) could be attributed to two orthogonal linear combinations of the axial temperatures. Although these two linear combinations lacked physical meaning, it was demonstrated that the hot spot temperature and average temperature in the reactor were variables that had predictable variation. Some preliminary analysis on the matching of inputs and output variables for stochastic processes was proposed. A number of potential problems with this technique were noted. Further analysis into the feasibility of this method is required.

The frequency response of some univariate stochastic controllers was examined in Chapter 6. For processes with transport delay, the frequency characteristics of these regulators are primarily dictated by the structure of the disturbances affecting the processes. The manner in which stochastic controllers compensate for process deadtime was compared to a more traditional technique for deadtime compensation. The Smith regulator was seen to correspond to a stochastic controller for a particular choice of disturbance model.

The application of self-tuning regulators to control a pilot scale tubular reactor was investigated in Chapter 7. The self-tuning regulator gave good control over the hot spot temperature. Although it is usually desired to control the effluent concentrations, this requires an inferential estimator, usually obtained from a complex model of the process. The applicability of self-tuning regulators as a means for reactor control must be reconciled against the difficulties involved in obtaining such a model.

Aside from the question of applicability of using self-tuning regulators for reactor control, these were shown to give good control of a process which has proven difficult to control by conventional means. The self-tuning algorithm was robust to the assumptions in its development. The controller readily retuned itself when changes were made in the operating conditions, or when changes in the process occurred due to catalyst decay. Self-tuning regulators are most applicable to multiple input/single output process. Although the theory for multivariate processes is not difficult to derive, complications arise from the large number of parameters that need to be estimated, and this precludes most multivariate applications.

The duality between the control of deterministic and stochastic processes was explored in Chapter 8. It was shown that stochastic difference equations are capable of modelling random disturbances, and randomly occurring deterministic

disturbances. As a consequence, linear quadratic controllers designed to guard against a particular class of stochastic disturbances, have the same structure as controllers designed to guard against an 'equivalent class' of deterministic disturbances. A number of practical and theoretical problems encountered in controlling processes subject to both types of disturbances were observed. Proposed nonlinear filtering solutions to alleviate these difficulties were too complex to be of practical use.

The nature of the process disturbances play a large role in the performance of a control system. Modelling of a process is never complete when a dynamic model is only available. It is important that the disturbances be modelled and included in the description of the process. When a dynamic model of the process cannot be obtained from fundamental arguments, a model can still be obtained by fitting input/output data to an empirical model. Although this model may not give the insight and understanding into the operation of the process it still provides a route to designing a controller. The use of empirical models in conjunction with multivariate statistical techniques, such as principal components and canonical variates analysis of time series allows one to study the structure of the process. In instances where the design of a controller is only of interest, adaptive controllers are a convenient means of proceeding to this stage directly.

REFERENCES

- Abraham, B., Box, G.E.P. (1978), "Deterministic and Forecast-adaptive Time-dependent Models", Appl. Statist. 27, 120.
- Ackerson, G.A., Fu, K. (1970), "On State Estimation in Switching Environments", IEEE AC-15, 10.
- Aidarous, S.E., Gevers, M.R., Installe, M.J. (1975), "Optimal Sensors' Allocation for a Class of Stochastic Distributed Systems", Int. J. Control, 22, 197.
- Akaike, H. (1967), "Some Problems in the Application of the Cross Spectral Method", Proceedings of an Advanced Seminar on Spectral Analysis of Time Series, B. Harris, editor, Wiley, New York.
- Akaike, H. (1971), "Autoregressive Model-Fitting for Control", Ann. Inst. Stat. Math., 23, 163.
- Akaike, H. (1974), "A New Look at the Statistical Model Identification", IEEE AC-19, 716.
- Akaike, H. (1978a), "On Newer Statistical Approaches to Parameter Estimation and Structure Determination", IFAC World Congress, Helsinki.
- Akaike, H. (1978b), "On the Identification of State Space Models and Their Use in Control", Special Topics Meeting on Time Series Analysis, Ames, Iowa.
- Alspach, D.L. (1972), "Comments on 'On the Identification of Variances and Adaptive Kalman Filtering'", IEEE AC-17, 843.
- Anderson, T.W. (1958), Introduction to Multivariate Statistical Analysis, Wiley, New York.
- Aris, R., Mah, R.H.S. (1963), "Independence of Chemical Reactions", Ind. Eng. Chem. Fundamentals, 2, 90.
- Asbjornsen, O.A., Fjeld, M. (1970), "Response Modes of Continuous Stirred Tank Reactors", CES 25, 1627.

- Astrom, K.J. (1967), "Computer Control of a Paper Machine - An Application of Linear Stochastic Control Theory", IBM Journal, 389.
- Astrom, K.J. (1970), Introduction to Stochastic Control Theory, Academic Press.
- Astrom, K.J. (1974), "A Self-tuning Regulator for Nonminimum Phase Systems", Department of Automatic Control, Lund Institute of Technology, Report 7411(C).
- Astrom, K.J. (1977), "Frequency Domain Properties of Otto Smith Regulators", Int. J. Control, 26, 307.
- Astrom, K.J., Borrisson, U., Ljung, L., Wittenmark, B., (1977), "Theory and Applications of Self-tuning Regulators", Automatica, 13, 457.
- Astrom, K.J., Wittenmark, B. (1973), "On Self-Tuning Regulators", Automatica 9, 185.
- Astrom, K.J., Wittenmark, B. (1974), "Analysis of a Self-tuning Regulator for Nonminimum Phase Systems", IFAC Symposium on Stochastic Control, Budapest.
- Athans, M. (1971), "The Role and Use of the Stochastic Linear-Quadratic-Gaussian Problem in Control System Design", IEEE AC 16, 529.
- Athans, M. (1972), "On the Determination of Optimal Costly Measurement Strategies for Linear Stochastic Systems", Automatica 8, 397.
- Barkley, C.H. (1959), "Stability of Chemical Reactors", Chem. Engng. Prof. Symp. Ser. 55, 37.
- Bartlett, M.S. (1946), "On the Theoretical Specifications of Sampling Properties of Auto-correlated Time Series", J. Roy. Statis. Soc., 88, No. 27.
- Bartlett, M.S. (1955), Stochastic Processes, Cambridge University Press.
- Berezin, I.S., Zhidkow, N.P. (1965), Computing Methods, Pergamon Press.
- Bohlin, T. (1971), "On the Problems of Ambiguities in Maximum Likelihood Identification", Automatica 7, 199.

- Borisson, U., Syding, R. (1976), "Self-Tuning Control of an Ore Crusher", *Automatica* 12, 1.
- Box, G.E.P., Hunter, W.G., MacGregor, J.F., Erjavec, J. (1973), "Some Problems Associated with the Analysis of Multiresponse Data", *Technometrics* 15, 33.
- Box, G.E.P., Jenkins, G.M. (1970), Time Series Analysis Forecasting and Control, Holden-Day.
- Box, G.E.P., MacGregor, J.F. (1976), "Parameter Estimation with Closed-Loop Operation Data", *Technometrics* 18, 371.
- Box, G.E.P., Tiao, G.C. (1977), "A Canonical Analysis of Multiple Time Series", *Biometrika*, 64, 355.
- Bradshaw, A., Porter, B. (1978), "Design of Linear Multivariable Discrete-Time Tracking Systems Incorporating Error-Activated Controllers", *Int. J. Systems Sci.* 9, 185.
- Brinkley, J.R. (1946), "A Note on the Conditions of Equilibrium for Systems of Many Constituents", *J. Chem. Phys.* 14, 563.
- Bristol, E.H., (1966), "On a New Measure of Interaction for Multivariable Process Control", *IEEE AC-11*, 133.
- Bucy, R.S., Joseph, P.D. (1968), Filtering for Stochastic Processes with Applications to Guidance and Control. Intersciences, New York.
- Carew, B., Belanger, P.R. (1973), "Identification of Optimum Filter Steady State Gain for Systems with Unknown Noise Covariances", *IEEE AC-18*, 582.
- Cegrell, T., Hedqvist, T. (1975), "Successful Adaptive Control of Paper Machines", *Automatica* 11, 53.
- Clarke, D.W., Gawthrop, P.J. (1975), "Self-Tuning Controller", *IEE* 122, 929.
- Clarke, D.W., Hasting-James, R. (1971). "Design of Digital Controllers for Randomly Disturbed Systems", *IEE* 118, 1503.
- Colanfuoni, G., Padmanabhan, L. (1977), "Optimal Sensor Locations for Tubular-Flow Reactor Systems", *CES*, 32, 1035.

- Davison, E.J. (1972), "The Output Control of Linear Time-Invariant Multivariable Systems with Unmeasurable Arbitrary Disturbances", IEEE AC-17, 621.
- Davison, E.J., Smith, H.W. (1971), "Pole Assignment in Linear Time-Invariant Multivariable Systems with Constant Disturbances", Automatica, 7, 489.
- Denbigh, K. (1971), The Principles of Chemical Equilibrium, Cambridge University Press.
- Dumont, G.A., Belanger, P.R. (1978), "Self-Tuning Control of a Titanium Dioxide Kilm", IEEE AC-23, 532.
- Finlayson, B. (1971), "Packed Bed Reactor Analysis by Orthogonal Collocation", CES, 26, 1081.
- Fjeld, M., Asbjornsen, O.A., Astrom, K.J. (1974), "Reaction Invariants and Their Importance in the Analysis of Eigenvectors, State Observability and Controllability of the Continuous Stirred Tank Reactor", CES, 29, 1917.
- Froment, G.F. (1967), "Fixed Bed Catalytic Reactors", IEC, 59 #2, 18.
- Georgakis, C., Aris, R., Amundson, N.R. (1977), "Studies in the Control of Tubular Reactors - I: General Considerations", CES 32, 1359.
- Georgakis, C., Aris, R., Amundson, N.R. (1977), "Studies in the Control of Tubular Reactors - II: Stabilization by Modal Control", CES 32, 1371.
- Georgakis, C., Aris, R., Amundson, N.R. (1977), "Studies in the Control of Tubular Reactors - III: Stabilization by Observer Design", CES 32, 1381.
- Gnedenko, B.V. (1967), The Theory of Probability, Chelsea, New York.
- Gould, L.A. (1969), Chemical Process Control: Theory and Applications. Addison Wesley, Massachusetts.
- Gustavsson, I., Ljung, L., Soderstrom, T. (1977), "Identification of Processes in Closed Loop - Identifiability and Accuracy Aspects", Automatica 13, 59.

- Haggan, V. (1975), "Dimensionality Reduction in Multivariable Stochastic Systems: Application to a Chemical Engineering Plant", Int. J. Control, 22, 763.
- Haggan, V., Priestley, M.B. (1975), "Dimensionality Reduction in Multivariable Stochastic Systems", Int. J. Control, 22, 249.
- Hammarstrom, L.G. (1979), "Control of Chemical Reactors in the Subspace of Reaction and Control Variants", CES, 34, 891.
- Herring, K.D., Melsa, J.L. (1972), "Optimum Measurements for Estimation", IEEE AC 19, 264.
- Hlavacek, V., Votruba, J. (1977), "Steady-State Operation of Fixed-Bed Reactors and Monolithic Structures", in Chemical Reactor Theory - A Review, ed. Lapidus, L., Amundson, N.R., Prentice-Hall.
- Ho, Y., Lee, R.C.K. (1964), "A Bayesian Approach to Problems in Stochastic Estimation and Control", IEEE AC-9, 333.
- Jenson, V.G., Jeffreys, G.V. (1963), Mathematical Methods in Chemical Engineering, Academic Press, New York.
- Johnson, C.D. (1968), "Optimal Control of the Linear Regulator with Constant Disturbances", IEEE AC 13, 416.
- Johnson, C.D. (1969), "Optimization of a Certain Quality of Complete Controllability and Observability for Linear Dynamical Systems", J. Basic Eng., Trans. ASME, Series D, 229.
- Johnson, C.D. (1970a), "Further Study of the Linear Regulator with Disturbances - The Case of Vector Disturbances Satisfying a Linear Differential Equation", IEEE AC, 15, 222.
- Johnson, C.D. (1970b), "Further Comments on 'Optimal Control of the Linear Regulator with Constant Disturbances'", IEEE AC, 15, 516.
- Johnson, C.D. (1971), "Accommodation of External Disturbances in Linear Regulator and Servomechanism Problems", IEEE AC, 16, 636.
- Johnson, C.D. (1972a), "Accommodation of Disturbances in Optimal Control Problems", Int. J. Control 15, 209.

- Johnson, C.D. (1972b), "Comments on 'The Output Control of Linear Time-Invariant Multivariable Systems with Unmeasurable Arbitrary Disturbances'", IEEE AC-17, 836.
- Johnson, C.D. (1973), "Control of Dynamic Systems", Symposium on Stochastic Control in Mechanics, University of Waterloo.
- Johnson, C.D. (1974a), "On Observers for Systems with Unknown and Inaccessible Inputs", Int. J. Control, 825.
- Johnson, C.D. (1974b), "Algebraic Solution of the Servomechanism Problem with External Disturbances", ASME Journal of Dynamic Systems, Measurement and Control", 96, 25.
- Johnson, C.D. (1975), "Improved Computational Procedures for 'Algebraic Solution of the Servomechanism Problem with External Disturbances'", ASME Journal of Dynamic Systems, Measurement and Control", 97, 161.
- Jorgensen, S.B., Clement, K. (1977), "Experimental Investigation of Reaction Zone Control in a Fixed Bed Reactor", in Chemical Engineering with Per Soltoft, Danish Technical Press, Copenhagen, 69.
- Jutan, A. (1976), "State Space Modelling and Multivariable Stochastic Control of a Pilot Plant Packed-Bed Reactor", Ph.D. Thesis, McMaster University, Hamilton, Canada.
- Jutan, A., Tremblay, J.P., MacGregor, J.F., Wright, J.D. (1977a) "Multivariable Computer Control of a Butane Hydrogenolysis Reactor, Part I: State Space Reactor Modelling", AIChE J. 23, 732-742.
- Jutan, A., MacGregor, J.F., Wright, J.D. (1977b), "Multivariable Computer Control of a Butane Hydrogenolysis Reactor, Part II: Data Collection, Parameter Estimation and Stochastic Disturbance Identification", AIChE J. 23, 742.
- Jutan, A., Wright, J.D., MacGregor, J.F. (1977c), "Multivariable Computer Control of a Butane Hydrogenolysis Reactor, Part III: On-line Linear Quadratic Stochastic Control Studies", AIChE J. 23, 751.
- Kalman, R.E. (1960), "A New Approach to Linear Filtering and Prediction Problems", J. Basic Eng., 82, 34.
- Kashyap, R.L., Rao, A.R. (1976), Dyanamic Stochastic Models from Empirical Data, Academic Press, New York.

- Kumar, S., Seinfeld, J. (1978a), "Optimal Location of Measurements for Distributed Parameter Estimation", IEEE AC 23, 690.
- Kumar, S., Seinfeld, J.H. (1978b), "Optimal Location of Measurements in Tubular Reactors", Chem. Eng. Sci., 33, 1507.
- Kuo, B.C. (1970), Discrete Data Control Systems, Prentice-Hall.
- Kushner, H. (1967), Introduction to Stochastic Control, Holt, Rinehart and Winston, Toronto.
- Kwakernaak, H., Sivan, R. (1972), Linear Optimal Control Systems, Wiley-Interscience, Toronto.
- Liou, C.T., Weigand, W.A., Lim, H.C. (1973), "Integral Action in the Optimal Control of Linear Systems with Some Inaccessible State Variables", Int. J. Control 17, 1029.
- Ljung, L., Wittenmark, B. (1974a), "Asymptotic Properties of Self-Tuning Regulators", Department of Automatic Control, Lund Institute of Technology, Report 7404.
- Ljung, L., Wittenmark, B. (1974b), "Analysis of a Class of Adaptive Regulators", Department of Automatic Control, Lund Institute of Technology, Report 7414(c).
- Luenberger, D.G. (1964), "Observing the State of a Linear System", IEEE Trans. Milit. Electron., 8, 74.
- MacGregor, J.F. (1973), "Optimal Discrete Stochastic Control Theory for Process Applications", Can. J. Chem. Eng. 51, 468.
- MacGregor, J.F. (1976), "Optimal Choice of the Sampling Interval for Discrete Process Control", Technometrics 18, 151.
- MacGregor, J.F. (1977), "Eratum: Optimal Choice of the Sampling Interval for Discrete Process Control", Technometrics 19, 224.
- MacGregor, J.F., Tidwell, P. (1977), "Discrete Stochastic Control with Input Constraints", Proc. IEE 124, 732.
- Masreliez, C.V. (1975), "Approximate Non-Gaussian Filtering with Linear and Observation Relations", IEEE AC-20, 107.
- Masreliez, C.J., Martin, R.D. (1977), "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter", IEEE AC 22, 361.

- Mehra, R.K. (1970), "On the Identification of Variances and Adaptive Kalman Filtering", IEEE AC 15, 175.
- Mehra, R.K. (1976), "Optimization of Measurement Schedules and Sensor Designs for Linear Dynamic Systems", IEEE AC 21, 55.
- Meier, L., Peschon, J., Dressler, R.M. (1967), "Optimal Control of Measurement Subsystems", IEEE, AC-12, 528.
- Mellefont, D.J., Sargent, R.W.H. (1977), "Optimal Measurement Policies for Control Purposes", Int. J. Control 26, 595.
- Mellefont, D.J., Sargent, R.W. (1978), "Selection of Measurements for Optimal Feedback Control", Ind. Eng. Chem. Process Des. Dev., 17, 549.
- Michelsen, M.L., Vakil, H.P., Foss, A.S. (1973), "State-Space Formulation of Fixed-Bed Reactor Dynamics", IEC Fundamentals, 12, 323.
- Mitchell, T.J. (1974), "An Algorithm for the Construction of 'D-Optimal' Experimental Designs", Technometrics, 16, 203.
- Muller, P.C., Weber, H.I. (1972), "Analysis and Optimization of Certain Qualities of Controllability and Observability for Linear Dynamic Systems", Automatica 8, 237.
- Neethling, C., Young, D.C. (1974), "Comments on 'Identification of Optimum Filter Steady State Gain for Systems with Unknown Noise Covariances'", IEEE AC-19, 623.
- Neuts, M.F. (1973), Probability, Allyn and Bacon.
- Noble, B. (1969), Applied Linear Algebra, Prentice-Hall.
- Omatu, S., Koide, S., Soeda, T. (1978), "Optimal Sensor Location Problem for a Linear Distributed Parameter System", IEEE-AC, 23, 665.
- Oppenheim, A.V., Schaffer, R.W. (1975), Digital Signal Processing, Prentice-Hall.
- O'Reilly, J. (1979), "Minimal-order Observers for Discrete-Time Linear Invariant Systems with Unmeasurable Disturbances", Int. J. Control, 29, 429.
- Parzan, E. (1969), "Multiple Time Series Modelling" in Multi-variate Analysis II, ed. Krishnaiah, P.R., Academic Press, New York, 389.

- Pell, T.M., Aris, R. (1974), "Problems in Chemical Reactor Analysis with Stochastic Features. Control of Linearized Distributed Systems on Discrete and Corrupted Observations", IEC Fund. 9, 15.
- Peterka, V., (1972), "On Steady State Minimum Variance Control Strategy", Kybernetika 8, 219.
- Priestley, T., Rao, T.S., Tong, H. (1974a), "Identification of the Structure of Multivariable Stochastic Systems" in Multivariate Analysis III, ed. Krishnaiah, P.R., Academic Press, New York, 351.
- Priestley, M.B., Rao, T.S., Tong, H. (1974b), "Applications of Principal Component Analysis and Factor Analysis in the Identification of Multivariable Systems", IEEE AC 19, 730.
- Orlickas, A. (1970), Kinetic Study of the Hydrogenolysis of n-Butane on Nickel Catalyst, M. Eng. Thesis, McMaster University, Hamilton, Ontario.
- Orlickas, A., Hoffman, T.W., Shaw, I.D., Reilly, P.M. (1972), "The Hydrogenolysis of n-Butane on a Nickel on Silica Catalyst: I - A Kinetic Model", Can. J. Chem. Eng., 50, 628.
- Quenouille, M.H. (1957), The Analysis of Multiple Time-Series, Hafner.
- Rader, C.M., Gold, B. (1967), "Digital Filter Design Techniques in the Frequency Domain", Proc. IEEE, 55, 149.
- Ragazzini, J.R., Franklin, G.F. (1958), Sampled Data Control Systems, McGraw-Hill.
- Rao, R.S. (1975), "An Innovation Approach to the Reductions of the Dimensions in a Multivariable Stochastic System", Int. J. Control 21, 673.
- Rissanen, J. (1976), "Minimax Entropy Estimation of Models for Vector Processes" in System Identification: Advances and Case Studies, ed. Mehra, R.K. Lainiotis, D.G., Academic Press, New York (1970).
- Root, J.G. (1969), "Optimum Control of Non-Gaussian Linear Stochastic Systems with Inaccessible State Variables", SIAM J. Control, 7, 317.
- Sage, A.P., White, C.C. (1977), Optimum Systems Control, 2nd edition, Prentice-Hall, Toronto.

- Sastry, V.A., Seborg, D.E., Wood, R.K. (1977), "An Application of a Self-Tuning Regulator to a Binary Distillation Column", *Automatica* 13, 417.
- Schneider, D.R., Reklaitis, G.V. (1975), "On Material Balances for Chemically Reacting Systems", *CES* 30, 243.
- Schubert, E. (1974), Letter to the Editor, *AIChE*, 20, 1232.
- Shaw, I.D., Hoffman, I.W., Orlickas, A., Reilly, P.M. (1972), "The Hydrogenolysis of n-Butane on a Nickel on Silica Catalyst: II Fluidized Bed Studies", *Can. J. Chem. Eng.*, 50, 637.
- Shaw, I.D. (1974), Modelling and Discrimination Studies in a Catalytic Fluidised Bed Reactor, Ph.D. Thesis, McMaster University, Hamilton, Ontario.
- Sinha, N.K., Tom, A.F. (1978), "Adaptive State Estimation for Systems with Coloured Observation Noise", *Int. J. Systems Sci.*, 9, 31.
- Smith, C.L. (1972), Digital Computer Process Control, In text, Toronto.
- Smith, O.J.M. (1957), "Closer Control of Loops with Deadtime", *Chem. Engng. Prog.*, 53, 217.
- Soderstrom, T., Ljung, L., Gustavsson, I. (1978), "A Theoretical Analysis of Recursive Identification Methods", *Automatica*, 14, 231.
- Sorenson, H.W. (1968), "Controllability and Observability of Linear, Stochastic, Time-Discrete Control Systems", Advances in Control Systems, Academic Press, New York.
- Sorenson, H.W., Alspach, D.L. (1971), "Recursive Bayesian Estimation Using Gaussian Sums", *Automatica* 7, 465.
- Sorenson, J.P. (1976), "Experimental Investigation of the Dynamics of a Fixed-Bed Reactor", *CES*, 31, 719.
- Sorenson, J.P. (1977), "Experimental Investigation of the Optimal Control of a Fixed-Bed Reactor", *CES*, 32, 763.

- Strangeland, B.E., Foss, A.S. (1970), "Control of a Fixed-Bed Chemical Reactor", IEC Fundamentals, 9, 38.
- Tomizuka, M., Rosenthal, D.E. (1979), "On the Optimal Digital State Vector Feedback Controller with Integral and Preview Actions"; ASME Journal of Dynamic Systems, Measurement and Control, 101, 172.
- Tou, J.T. (1967), Optimum Design of Digital Control Systems, Academic Press, New York.
- Tremblay, J.P. (1977), Real Time Computer Data Acquisition and Control Systems: An Application to Model Reference Control of a Pilot Scale, Packed Bed Tubular Reactor, Ph.D. Thesis, McMaster University.
- Tremblay, J.P., Wright, J.D. (1977), "Multivariable Model Reference Adaptive Control of a Pilot Scale Packed Bed Tubular Reactor", 5th IFAC/IFIP International Conference on Digital Computer Applications to Process Control, The Hague, 513.
- Vakil, H.B., Michelsen, M.L., Foss, A.S. (1973), "Fixed-Bed Reactor Control with State Estimation", IEC Fundamentals, 12, 328.
- Van Welsenaere, R.J., Froment, G.F. (1970), "Parametric Sensitivity and Runaway in Fixed Bed Catalytic Reactors", CES, 25, 1503.
- Villadsen, J., Michelsen, M. (1977), Solution of Differential Equation Models by Polynomial Approximation, Prentice-Hall, Toronto.
- Villadsen, J.V., Stewart, W.E. (1967), "Solution of Boundary-value Problems by Orthogonal Collocation", CES, 22, 1483.
- Whittle, P. (1963), Prediction and Regulation by Linear Least-Squares Methods, English University Press.
- Whittle, P. (1963b), "On the Fitting of Multivariate Auto-regressions and the Approximate Canonical Factorization of a Spectral Density Matrix", Biometrika 50, 129.
- Wiener, N. (1949), The Extrapolation, Interpolation and Smoothing of Stationary Time Series, Wiley, N.Y.

- Willsky, A.S., Jones, H.L. (1976), "A Generalized Approach to the Detection and Estimation of Jumps in Linear Systems", IEEE AC '21, 108.
- Wilson, G.T. (1970), Modelling Linear Systems for Multivariable Control, Ph.D. Thesis, University of Lancaster.
- Wittenmark, B. (1973), "A Self-Tuning Regulator", Department of Automatic Control, Lund Institute of Technology, Report 7311.
- Wittenmark, B. (1974), "Adaptive Control of a Paper Machine", Proc. of the Symposium on Computers, Electronics and Control, Calgary.
- Wong, L.K. (1977), "Multivariate Identification and Stochastic Control of a Pilot Plant Packed Bed Reactor", M. Eng. Thesis, McMaster University, Hamilton, Ontario.
- Wonham, W.M. (1974), Linear Multivariable Control - A Geometric Approach, Springer-Verlag, New York,
- Wright, J.D., MacGregor, J.F., Tremblay, J.P., Wong, L.K. (1977), "Inferential Control of An Exothermic Packed Bed Tubular Reactor", JACC, San Francisco, 1516.
- Wynn, H.P. (1970), "The Sequential Generation of D-optimum Designs", Ann. Math. Statist., B35, 1655.
- Yu, T.K., Seinfeld, J.H. (1973), "Observability and Optimal Measurement Location in Linear Distributed Parameter Systems", Int. J. Control, 18, 785.

APPENDIX 1

RATE EXPRESSIONS FOR THE HYDROGENOLYSIS OF N-BUTANE

A1.1 Reaction Scheme

The following reactions have been proposed (Orlickas (1970, 1972) and Shaw (1972, 1974)) to represent the hydrogenolysis of n-butane.



A1.2 Reaction Rates of the Reaction Species

Net rate of disappearance of butane

$$R_{\text{C}_4\text{H}_{10}} = k/k_0 \cdot P_{\text{C}_4\text{H}_{10}}^{m'} \cdot P_{\text{H}_2}^{n'} \cdot A \quad (\text{A1.5})$$

where

$$A = k_B \cdot \exp(-\Delta E_B/RT) \quad (\text{A1.6})$$

Net rate of appearance of propane

$$R_{C_3H_8} = \frac{F \cdot R_{C_4H_{10}} - k/k_0 \cdot B \cdot P_{C_3H_8}^{m''} \cdot P_{H_2}^{n''}}{1 + C} \quad (A1.7)$$

where

$$B = k_{p1} \cdot \exp(-\Delta E_{p1}/RT) \quad (A1.8)$$

$$C = k_{p2} \cdot \exp(-\Delta E_{p2}/RT) \quad (A1.9)$$

Net rate of appearance of ethane

$$R_{C_2H_6} = \frac{(2-F) \cdot R_{C_4H_{10}} - R_{C_3H_8} - k/k_0 \cdot D \cdot P_{C_2H_6}^{m''} \cdot P_{H_2}^{n''}}{1 + G} \quad (A1.10)$$

where

$$D = k_{E1} \cdot \exp(-\Delta E_{E1}/RT) \quad (A1.11)$$

$$G = k_{E2} \cdot \exp(-\Delta E_{E2}/RT) \quad (A1.12)$$

Net rate of appearance of methane

$$R_{CH_4} = 4 \cdot R_{C_4H_{10}} - 3 \cdot R_{C_3H_8} - 2 \cdot R_{C_2H_6} \quad (A1.13)$$

Net rate of disappearance of hydrogen

$$R_{H_2} = 3 \cdot R_{C_4H_{10}} - 2 \cdot R_{C_3H_8} - R_{C_2H_6} \quad (A1.14)$$

where

- F = fraction of butane which reacts to propane
- k/k_0 = catalyst activity (dimensionless).
- k_B = frequency factor for butane ($\text{moles-sec}^{-1}\text{-gm catalyst}^{-1}\text{-atm}^{(m'+n')}$)
- ΔE_B = activation energy for rate of butane cracking (cal-gm mole^{-1})
- m' = exponent on butane partial pressure
- n' = exponent on hydrogen partial pressure in the butane rate expression
- $P_{C_4H_{10}}$ = partial pressure of butane (atm)
- P_{H_2} = partial pressure of hydrogen (atm)
- k_{p1} = pre-exponential factor in propane rate expression ($\text{moles-sec}^{-1}\text{-gm catalyst}^{-1}\text{-atm}^{-(m''+n'')}$)
- k_{p2} = pre-exponential factor in propane rate expression (dimensionless)
- $\Delta E_{p1}, \Delta E_{p2}$ = activation energies in propane rate expression (cal-gm mole^{-1})
- m'' = exponent on propane partial pressure
- n'' = exponent on hydrogen partial pressure in the propane rate expression
- $P_{C_3H_8}$ = partial pressure of propane (atm)

- k_{E1} = pre-exponential factor in ethane rate expression (moles-sec⁻¹-gm catalyst⁻¹-atm^(m'''+n'''))
 k_{E2} = pre-exponential factor in ethane rate expression (dimensionless)
 $\Delta E_{E1}, \Delta E_{E2}$ = activation energies in ethane rate expression (cal-gm mole⁻¹)
 m''' = exponent on ethane partial pressure
 n''' = exponent on hydrogen partial pressure in the ethane rate expression
 $P_{C_2H_6}$ = partial pressure of ethane (atm)
 R_i = rate of disappearance or appearance of component i (gm moles-sec⁻¹-gm catalyst⁻¹)
 T = reacting temperature (°K)
 R = universal gas law constant (cal-gm mole⁻¹-°K⁻¹)

A1.3 Values of Kinetic Parameters Shaw (1974)

(a) Butane rate

$$k_B = 10^{15.6}$$

$$m' = 1$$

$$\Delta E_B = 5.1 \times 10^4$$

$$n' = -2.34 \text{ (or } -2.15)$$

(b) Propane rate expression

$$k_{p1} = 10^{10.6}$$

$$E_{p2} = 3.0 \times 10^4$$

$$k_{p2} = 10^{12.12}$$

$$m'' = 1.0$$

$$\Delta E_{p1} = 4.0 \times 10^4$$

$$n'' = -2.15 \text{ (or } -2.08)$$

(c) Ethane rate expression

$$k_{E1} = 10^{4.52}$$

$$\Delta E_{E2} = 1.6 \times 10^4$$

$$k_{E2} = 10^{6.81}$$

$$m''' = 1.0$$

$$\Delta E_{E1} = 2.6 \times 10^4$$

$$n''' = -2.2$$

A1.4 Heats of Reaction (Units cal/gmole) Jutan (1976)

Reaction 1: $\Delta h_1 = -12,560 - 5.0 (T-298)$

Reaction 2: $\Delta h_2 = -10,322 - 6.3 (T-298)$

Reaction 3: $\Delta h_3 = -13,305 - 3.28 (T-298)$

Reaction 4: $\Delta h_4 = -15,542 - 2.52 (T-298)$

APPENDIX 2

REACTION STOICHIOMETRY IN TUBULAR REACTORS

A2.1 Introduction

In this Appendix an expression for the reaction stoichiometry in tubular reactors with no axial dispersion, and with axial dispersion are derived. The basic assumptions in each model are that the velocity profile is constant, there is negligible pressure drop across the reactor and the radial and axial diffusivities are identical for all components.

A2.2 Tubular Reactor with no Axial Dispersion

The mass balance in a tubular reactor, with radial gradients, is

$$\frac{\partial \underline{C}}{\partial t} = -v_c \cdot \frac{\partial \underline{C}}{\partial z} + \frac{D_{er}}{\epsilon R^2} \cdot \frac{\partial}{\partial r} \left(r \frac{\partial \underline{C}}{\partial r} \right) + \frac{\rho_B S' \underline{\xi}}{\epsilon} \quad (\text{A2.1})$$

The boundary conditions are:

$$r = 0: \quad \frac{\partial \underline{C}}{\partial r} = \underline{0} \quad (\text{symmetry}) \quad (\text{A2.2})$$

$$r = 1: \quad \frac{\partial \underline{C}}{\partial r} = \underline{0} \quad (\text{A2.3})$$

$$z = 0: \quad \underline{C} = \underline{C}(t, r, 0) \quad (\text{A2.4})$$

$$t = 0: \quad \underline{c} = \underline{c}(0, r, z) \quad (\text{A2.5})$$

Multiplying (A2.1) by the element by species matrix E, and using the fact that

$$ES' = 0 \quad (\text{A2.6})$$

we obtain

$$\frac{\partial \underline{EC}}{\partial t} = -v_c \frac{\partial \underline{EC}}{\partial z} + \frac{D_{er}}{R^2} \cdot \frac{\partial}{\partial r} \left(r \frac{\partial \underline{EC}}{\partial r} \right) \quad (\text{A2.7})$$

The boundary conditions are obtained by multiplying (A2.2 - A2.5) by E. Let us define a variable $\underline{EC}^* = \underline{EC}(0, r, z)$ which satisfies the differential equation (A2.7), and the associated boundary conditions (A2.2) and (A2.3). There are several possibilities for C^* , but selection of a particular choice will be deferred for the moment. Define

$$\underline{w} = \underline{EC} - \underline{EC}^* \quad (\text{A2.8})$$

\underline{w} satisfies the differential equation

$$\frac{\partial \underline{w}}{\partial t} = -v_c \frac{\partial \underline{w}}{\partial z} + \frac{D_{er}}{R^2} \cdot \frac{\partial}{\partial r} \left(r \frac{\partial \underline{w}}{\partial r} \right) \quad (\text{A2.9})$$

The appropriate boundary conditions are

$$r = 0: \frac{\partial w}{\partial r} = 0 \quad (\text{symmetry}) \quad (\text{A2.10})$$

$$r = 1: \frac{\partial w}{\partial r} = 0 \quad (\text{A2.11})$$

$$z = 0: \underline{w} = \underline{w}(t, r, 0) \quad (\text{A2.12})$$

$$t = 0: \underline{\dot{w}} = 0 \quad (\text{A2.13})$$

Equation (A2.9) is a set of linear partial differential equations. Since the equations are uncoupled, we need only consider the solution of the partial differential equation

$$\frac{\partial w}{\partial t} = -v_c \frac{\partial w}{\partial z} + \frac{D_{er}}{R^2 e_r} \cdot \frac{\partial}{\partial r} \left(r \frac{\partial w}{\partial r} \right) \quad (\text{A2.14})$$

Taking the time domain Laplace transform of (A2.14), and using the initial condition (A2.13), we obtain

$$s\bar{w}(r, z) = -v_c \frac{\partial \bar{w}(r, z)}{\partial z} + \frac{D_{er}}{R^2 e_r} \cdot \frac{\partial}{\partial r} \left(r \frac{\partial \bar{w}(r, z)}{\partial r} \right) \quad (\text{A2.15})$$

The time domain Laplace transform of $w(t, r, z)$ has been denoted by $\bar{w}(r, z)$. We now contemplate a separation of variables solution to (A2.15). Let

$$\bar{w}(z, r) = \bar{f}(z)\bar{g}(r) \quad (\text{A2.16})$$

Substituting this into (A2.15) and separating variables we obtain

$$-v_c \frac{\bar{f}'(z)}{\bar{f}(z)} = \frac{s\bar{g}(r) - \frac{D_{er}}{R^2\epsilon} \left(\frac{1}{r} \cdot \frac{\partial}{\partial r} (r\bar{g}'(r)) \right)}{\bar{g}} = -\lambda \quad (\text{A2.17})$$

where λ is the separation constant. Following the usual method for separation of variables, Jenson and Jeffreys (1963), we obtain

$$\bar{g}''(r) + \frac{1}{r} \bar{g}'(r) - \frac{R^2\epsilon}{D_{er}} (s+\lambda)\bar{g}(r) = 0 \quad (\text{A2.18})$$

and

$$v_c \bar{f}'(z) - \lambda \bar{f}(z) = 0 \quad (\text{A2.19})$$

Equation (A2.18) is complex Bessel equation of order zero, and has the solution, Jenson and Jeffreys (1963),

$$\bar{g}(r) = AJ_0(jr\sqrt{(s+\lambda)R^2\epsilon/D_{er}}) + BY_0(jr\sqrt{(s+\lambda)R^2\epsilon/D_{er}}) \quad (\text{A2.20})$$

J_0 is a Bessel of the first kind of order zero, and Y_0 is the 'second solution'. A and B are constants of integration. The solution of $\bar{g}(r)$ must be bounded at $r = 0$. Since $Y_0(0) \rightarrow -\infty$, and $J_0(0) = 1$, B must be equal to zero. The Laplace transform of the boundary conditions (A2.10) and (A2.11) require that

$$\bar{f}(z)\bar{g}'(r) = 0, \quad r = 0, \quad r = 1 \quad (\text{A2.21})$$

Suppose $\bar{f}(z)$ does not vanish at $r = 0$, or $r = 1$. Substituting for the derivative of a zero order Bessel function (A2.21) implies that

$$A_j \sqrt{(s+\lambda)R^2\epsilon/D_{er}} \cdot J_1(jr\sqrt{(s+\lambda)R^2\epsilon/D_{er}}) = 0 \quad (\text{A2.22})$$

The solutions to this equation are the zeroes of J_1 , denoted by B_j , $i = 1, 2, \dots$ and $s = -\lambda$. For each of the zeroes B_j , there is a corresponding value of λ , λ_j . For each value of λ_j and $\lambda = s$, (A2.19) has the solution

$$\bar{f}(z) = c_i e^{z\lambda_i/v_c} \quad (\text{A2.23})$$

where c_i is a constant of integration. If $c_i \neq 0$, our assumption that $\bar{f}(z)$ is not zero is valid. The total solution to (A2.15) is given by a superposition of individual solutions, that is

$$\bar{w}(r,z) = a_0 e^{zs/v_c} + \sum_{i=1}^{\infty} a_i e^{\lambda_i z/v_c} J_0(B_i) \quad (\text{A2.24})$$

The a_i , $i = 0, 1, \dots$ are combinations of the integration constants from (A2.23) and (A2.20). The solution (A2.24) must hold for all values of (r,z) , and so it certainly holds for $z = 0$. Letting $z = 0$ in (A2.24) we obtain

$$\bar{w}(r,0) = a_0 + \sum_{i=1}^{\infty} a_i J_0(B_i) \quad (\text{A2.25})$$

Using the integral properties of the Bessel function, it is straightforward to verify that

$$a_0 = 2 \int_0^1 \bar{w}(r,0) r dr \quad (\text{A2.26})$$

and

$$a_i = \frac{2 \int_0^1 \bar{w}(r,0) J_0'(B_i) r dr}{J_0^2(B_i)}, \quad i = 1, 2, \dots \quad (\text{A2.27})$$

We can never hope to evaluate the a_i 's for an arbitrary $\bar{w}(r,0)$, as we would then need to know the complete set of zeroes for J_1 , of which there are an infinite number. Suppose though that $\bar{w}(r,0)$ is not a function of r , that is

$$\frac{\partial \bar{w}(r,0)}{\partial r} = 0 \quad (\text{A2.28})$$

The a_i 's are then given by

$$a_0 = \bar{w}(r,0) \quad (\text{A2.29})$$

and

$$a_i = 0, \quad i = 1, 2, \dots \quad (A2.30)$$

Using (A2.29) and (A2.30), and taking the inverse Laplace transform of (A2.24) we have

$$w(t, r, z) = w(t - z/v_c, r, 0) \quad (A2.31)$$

w was an arbitrary element of the vector

$$\underline{w} = E(\underline{C}(t, r, z) - \underline{C}^*(t, r, z)).$$

Thus

$$E(\underline{C}(t, r, z) - \underline{C}^*(0, r, z)) = E(\underline{C}(t - z/v_c, r, 0) - \underline{C}^*(0, r, 0)) \quad (A2.32)$$

$\underline{C}^*(0, r, z)$ is the concentration of the components in the reactor initially and $\underline{C}^*(0, r, 0)$ is the concentration of the components initially at the reactor entrance. The restrictions on $\underline{C}^*(0, r, z)$ are that it satisfy the differential equation (A2.2) and the associated boundary conditions. If $\underline{C}^*(0, r, z)$ is taken as the steady state solution of the mass balance (A2.1), it certainly satisfies the partial differential equation (A2.7). Initially, the reactor may be filled with feed mixture. If the initial feed is such that

$$\frac{\partial \underline{C}^*(0, r, 0)}{\partial r} = \underline{0} \quad (A2.33)$$

then $\underline{C}^*(0, r, 0)$ satisfies the differential equation (A2.7).

As well, we require, from the restriction (A2.28) that

$$\frac{\partial E(\underline{C}^*(t,r,0) - \underline{C}^*(0,r,0))}{\partial r} = 0 \quad (\text{A2.34})$$

Both these conditions are satisfied if the feed is distributed uniformly across the entrance of the reactor. This would be the case in most reactor installations.

To summarize these results, we find that:

Theorem 1: The reaction stoichiometry is given by (A2.32) where $\underline{C}^*(0,r,z)$ is chosen as i) the steady state solution to the mass balance (A2.1), or ii) the concentration of reactants in the feed at $t = 0$. In both cases, the feed must be such that at $z = 0$, (A2.34) holds. Furthermore, if the reference value $\underline{C}^*(0,r,z)$ is chosen as the concentrations of reactants at $t = 0$, (A2.33) must also hold.

Corollary 1: A sufficient condition for (A2.33) and (A2.34) to hold is that the feed be distributed uniformly across the entrance.

Corollary 2: The results do not depend on the dispersion coefficient D_{er} . This arises from the assumption of a uniform velocity profile. Consequently, the main result also holds for

adiabatic and isothermal tubular reactors with constant velocity profiles. It does not apply to reactors with axial dispersion.

Corollary 3: To determine the steady state solution, $\underline{C}^*(0,r,z)$ is taken as the concentration of the reactants at $t = 0$. In this instance (A2.31) reduces to

$$\underline{EC}(\infty,r,z) = \underline{EC}(0,r,0) \quad (\text{A2.34a})$$

A2.3 Tubular Reactors with Axial Dispersion

The mass balance in a tubular reactor with a constant velocity profile, no radial gradients and axial dispersion is

$$\frac{\partial \underline{C}}{\partial t} = -v_c \frac{\partial \underline{C}}{\partial z} + \frac{D_{ea}}{\epsilon} \frac{\partial^2 \underline{C}}{\partial z^2} + \frac{\rho_B}{\epsilon} S' \underline{\xi} \quad (\text{A2.35})$$

The boundary conditions are

$$z = 0: \quad \left. \frac{-D_{ea}}{\epsilon} \frac{\partial \underline{C}}{\partial z} \right|_{z=0+} = v_c (\underline{C}(t,0) - \underline{C}(t,0+)) \quad (\text{A2.36})$$

$$z = L: \quad \frac{\partial \underline{C}}{\partial z} = 0 \quad (\text{A2.37})$$

$$t = 0: \quad \underline{C} = \underline{C}(0,z) \quad (\text{A2.38})$$

The boundary condition (A2.36) is a statement of the mass balance at the reactor entrance. The boundary condition (A2.37) is only true at steady state, although it is used for transient analysis

because an expression is not available; see Georgakis et al (1977a).

Multiplying (A2.36) by E , we obtain

$$\frac{\partial EC}{\partial t} = -v_c \frac{\partial EC}{\partial z} + \frac{D_{ea}}{\epsilon} \frac{\partial^2 EC}{\partial z^2} \quad (A2.39)$$

The boundary conditions are obtained by multiplying (A2.36 - A2.38) by E . Introduce a reference variable $\underline{C}^*(0,z) = \underline{C}(0,z)$ such that $\underline{EC}^*(0,z)$ satisfies the boundary conditions (A2.36), (A2.37) and the differential equation (A2.39).

Define

$$\underline{w} = E(\underline{C}(t,z) - \underline{C}^*(0,z)) \quad (A2.40)$$

Then \underline{w} satisfies the linear partial differential equation

$$\frac{\partial \underline{w}}{\partial t} = -v_c \frac{\partial \underline{w}}{\partial z} + \frac{D_{ea}}{\epsilon} \frac{\partial^2 \underline{w}}{\partial z^2} \quad (A2.41)$$

with boundary conditions

$$z = 0: \quad \left. \frac{-D_{ea}}{\epsilon} \frac{\partial \underline{w}}{\partial z} \right|_{z=0+} = v_c (\underline{w}(t,0) - \underline{w}(t,0+)) \quad (A2.42)$$

$$z = L: \quad \frac{\partial \underline{w}}{\partial z} = 0 \quad (A2.43)$$

$$t = 0: \quad \underline{w}(0,z) = 0 \quad (A2.44)$$

Taking the Laplace transform of (A2.42) and using (A2.44) we obtain

$$s\underline{w}^*(z) = -v_c \frac{\partial \underline{w}^*(z)}{\partial z} + \frac{D_{ea}}{\epsilon} \frac{\partial^2 \underline{w}^*(z)}{\partial z^2} \quad (\text{A2.45})$$

where \underline{w}^* denotes the Laplace transform of $w(t,z)$. Equation (A2.45) is a second order linear differential having the solution

$$\underline{w}^*(z) = \underline{a}_1 e^{\alpha_1 z} + \underline{a}_2 e^{\alpha_2 z} \quad (\text{A2.46})$$

with

$$\alpha_{1,2} = \frac{v_c \pm \sqrt{v_c^2 + 4sD_{ea}/\epsilon}}{2D_{ea}/\epsilon} \quad (\text{A2.47})$$

The constants \underline{a}_1 and \underline{a}_2 are determined from the Laplace transform of the boundary conditions (A2.42) and (A2.43). Substituting for $\underline{w}^*(z)$, we obtain the following equations for the constants of integration

$$\begin{bmatrix} \frac{D_{ea} \alpha_1 - v_c}{\epsilon} & \frac{D_{ea} \alpha_2 - v_c}{\epsilon} \\ \alpha_1 e^{\alpha_1 L} & \alpha_2 e^{\alpha_2 L} \end{bmatrix} \begin{bmatrix} \underline{a}_1 \\ \underline{a}_2 \end{bmatrix} = \begin{bmatrix} -v_c \underline{w}^{*'}(0) \\ 0 \end{bmatrix} \quad (\text{A2.48})$$

The coefficients of integration are readily obtained. However, they are complicated functions of the variable s , and the inverse

Laplace transform of (A2.46) does not have an analytic expression. For reactors with axial dispersion then, it is necessary to include all components in the material balance as a simple expression for the reaction stoichiometry does not exist.

An expression for the steady state reaction stoichiometry is obtained by solving the steady state version of (A2.41). The reaction stoichiometry is given by (A2.34a) with the radial dependence omitted.

APPENDIX 3

PARAMETERS FOR LINEARIZED FINITE DIMENSIONAL REACTOR MODEL

A3.1 Parameters for the Continuous Model

The linearized lumped model of the dynamics of the butane hydrogenolysis reactor is

$$\dot{\underline{x}} = A_C \underline{x}(t) + G_C \underline{u}(t) + F_C z(t) \quad (A3.1)$$

where

$\underline{x}(t)$ - 7 x 1 vector of mean corrected temperatures at the interior collocation points;

$\underline{u}(t)$ - 2 x 1 vector of mean corrected flowrates (butane and hydrogen);

$z(t)$ - deviation of the wall temperature from its steady state value.

The radially averaged concentration of the reaction species at the exit of the reactor is given by

$$\underline{c}(t) = H_I \underline{x}(t) + D_I \underline{u}(t) + F_I z(t) \quad (A3.2)$$

where

$\underline{c}(t)$ - 3 x 1 vector of mean corrected concentrations at the reactor exit (butane, hydrogen and propane).

From the September 23 data of Jutan (1976), the following average values were used in the linearization.

Average temperature profile, $^{\circ}\text{K}$:

$$\underline{x}'_{\text{av}} = (522.9, 526.6, 533.5, 541.7, 543.6, 538.7, 536.4)$$

The mean flow rates of butane and hydrogen in cc/sec at STP:

$$\underline{u}'_{\text{av}} = (16.1, 87.4)$$

To compute the controls $\underline{u}(t)$ the flow rates must be converted to reactor inlet conditions where the pressure equals 167 kPa (1.65 atmospheres) and the temperature equals the steady wall temperature.

The steady value of the wall temperature, $^{\circ}\text{K}$:

$$z_{\text{av}} = 522^{\circ}\text{K}$$

The radially averaged steady state values of the effluent concentrations (obtained by integrating the steady state mass and energy balances) in $\text{g}\cdot\text{mole}/\text{cc} \times 10^{-6}$.

$$\underline{c}'_{\text{av}} = (1.237, 2.260, 0.592)$$

The matrices ($A_c, G_c, F_c, H_I, D_I, F_I$) are given on the following pages.

A3.2 Parameters for the Discrete Model

The continuous model was integrated over a sixty second control interval, to obtain

$$\underline{x}(k+1) = A\underline{x}(k) + G\underline{u}(k) + Fd(k) \quad (A3.3)$$

The effluent concentrations are still given by (A3.2) but the time increment is changed to

$$\underline{c}(k) = H_I\underline{x}(k) + D_I\underline{u}(k-1) + F_I d(k-1) \quad (A3.4)$$

In Chapter 8, a steady state controller of the form

$$u(k) = -Lx(k) \quad (A3.5)$$

was used. The matrix L, from Jutan (1976) is also reported here.

Continuous Dynamics Matrix - A_C

-.2433E-01	-.5637E-02	.1938E-02	-.1149E-02	.9582E-03	-.1171E+02	.6977E-03
.1154E-01	-.1102E-01	-.4553E-02	.2138E-02	-.1650E-02	.1959E-02	-.1160E-02
-.5144E-02	.5843E-02	-.7459E-02	-.4578E-02	.2753E-02	-.3027E-02	.1767E-02
.2992E-02	-.2924E-02	.4312E-02	-.1466E-02	-.5793E-02	.5053E-02	-.2846E-02
-.2045E-02	.1331E-02	-.2729E-02	.3645E-02	.1878E-02	-.1155E-01	.5709E-02
.1116E-02	-.1205E-02	.6865E-03	-.2715E-02	.5279E-02	.1051E-01	-.1995E-01
-.1882E-02	.1290E-02	-.2142E-02	.2216E-02	-.7622E-02	.5237E-01	-.5329E-01

Continuous Control Matrix - G_C

.7398E-03	-.3838E-03
.1191E-02	-.5356E-03
.3077E-02	-.9454E-03
.8608E-02	-.1797E-02
.1160E-01	-.1704E-02
.8665E-02	-.7090E-03
.7758E-02	-.4335E-03

Continuous Disturbance Matrix - F_C

.110E-01
.118E-01
.132E-01
.149E-01
.152E-01
.142E-01
.138E-01

dr

Discrete Dynamic Matrix A

.187100	-.097390	.051390	-.041270	.038840	-.043250	.025040
.255900	.433000	-.111000	.075390	-.066590	.072360	-.041690
-.079250	.234600	.563900	-.152800	.112100	-.114600	.065270
.051040	-.112000	.241200	.815000	-.238000	.207200	-.114900
-.038240	.046800	-.131300	.285700	.914400	-.389700	.200200
.005250	.035350	.013420	-.135700	.463900	.647000	-.254200
.016040	-.031820	.008155	-.090170	.172300	.833380	-.246400

Discrete Control Matrix G

.019570	-.009980
.051330	-.024400
.129200	-.041450
.432100	-.094680
.675500	-.104500
.490100	-.048140
.421200	-.315100

Discrete Disturbance Matrix F

.296600
.573800
.625800
.846100
.834800
.726000
.688700

Effluent Concentration Matrix - H_I

.001638	.005577	.012020	.029900	.017280	.005750	-.000262
.002186	.005048	.014680	.033750	.036560	.010390	.000495
.000654	.001625	.002950	.004184	.003740	.001411	.000023

Effluent Control Matrix - D_I

.068730	-.023590
.098640	-.033740
.028500	-.009767

Effluent Disturbance Matrix - F_I

.015600
.018000
.013600

Steady State Controller - L_I

.01268	.03750	.11750	.34190	.35380	.05323	.01569
-.02804	-.06498	-.10440	-.0307	-.08942	-.11830	-.03623

APPENDIX 4

SAMPLING PROPERTIES OF AUTOCORRELATION FUNCTIONS

A4.1 Introduction

The theoretical sampling properties of a 'discounted' autocorrelation function are derived in this appendix. The theoretical variances of the sampled estimate of the 'discounted' autocorrelation function are found to be larger than those calculated assuming no discounting. The discounted autocorrelation function can be used to test for parameter convergence in self-tuning regulators.

A4.2 Weighted Least Squares Estimation

To implement a self-tuning regulator (STR), one estimates the parameters of the model:

$$Y(t+b) = \alpha_0 Y(t) + \dots + \alpha_m Y(t-m) + \beta_0 U(t) + \dots + \beta_2 U(t-2) + \epsilon(t+b) \quad (\text{A4.1})$$

by minimizing

$$V_1(\underline{\theta}) = \sum_{t=1}^N \epsilon^2(t+b) \quad (\text{A4.2})$$

by a recursive least squares technique. $\underline{\theta}$ is the vector of parameters (α_i, β_i) . These estimates are then used to compute the

control signal

$$U(t) = -\frac{\hat{\alpha}_0}{\hat{\beta}_0} Y(t) - \dots - \frac{\hat{\alpha}_m}{\hat{\beta}_0} Y(t-m) - \frac{\hat{\beta}_1}{\hat{\beta}_0} U(t-1) - \dots - \frac{\hat{\beta}_2}{\hat{\beta}_0} U(t-2) \quad (A4.3)$$

If the parameter estimates of (A4.1) converge, and if the process dynamics may be described by a linear transfer function, and the effect of the disturbances acting on the output by a linear stochastic difference equation, Astrom and Wittenmark (1973) have shown that the closed loop has the following properties:

$$EY(t+\tau)Y(t) = r_{yy}(\tau) = 0, \quad \tau = b, b+1, \dots, b+m$$

and

$$EY(t+\tau)U(t) = r_{yu}(\tau) = 0, \quad \tau = b, b+1, \dots, b+2 \quad (A4.4)$$

Furthermore, if the orders l & m are the same as those of the minimum variance controller then

$$r_{yy}(\tau) = 0, \quad \tau \geq b \quad (A4.5)$$

$$r_{yu}(\tau) = 0, \quad \tau \geq b$$

It is seen that equation (A4.4) provides a convenient means of testing for parameter convergence. Box and Jenkins (1970) suggest

one replace $r_{yy}(\tau)$ by its sampled value, and compare this against its two standard deviation limits to see if these values are zero. Barlett (1946) has shown that the sample autocorrelation of a stationary normal process is asymptotically normally distributed with variance given by

$$\text{var}(\hat{r}_{yy}(k)) = \frac{1}{N} \sum_{j=-\infty}^{\infty} (\rho_j^2 + \rho_{j+k} \rho_{j-k} + 4\rho_k \rho_j \rho_{j+k} + 2\rho_j^2 \rho_k^2) \quad (A4.6)$$

If the theoretical autocorrelation is deemed to have died out beyond lag k then (A4.6) reduces to

$$\text{var}(\hat{r}_{yy}(\tau)) = \frac{1}{N} \left(1 + \sum_{j=1}^k \rho_j^2 \right), \quad \tau > k \quad (A4.7)$$

If the cross correlation between Y and U is zero beyond a certain lag, then a similar expression to (A4.7) is obtained, Bartlett (1956).

The use of discounted least squares is sometimes preferred when implementing a STR. One estimates the parameters of the model (A4.1) by minimizing

$$V_2(\underline{\theta}) = \sum_{t=1}^N \lambda^{N-t} \varepsilon^2(t+b) \quad (A4.8)$$

The discounting factor λ , is usually chosen to lie in the range $.9 < \lambda \leq 1$.

The model (A4.1) may be expressed as

$$Y(t+b) = \underline{X}'(t)\underline{\theta} + \varepsilon(t+b) \quad (A4.9)$$

where

$$\underline{X}'(t) = (Y(t), Y(t-1), \dots, Y(t-m), U(t), U(t-1), \dots, U(t-l)) \quad (A4.10)$$

and $\underline{\theta}$ is the vector of unknown parameters. Introducing the diagonal matrix

$$S = \text{diag} (1, \lambda^1, \lambda^{N-1}) \quad (A4.11)$$

we see that (A4.8) may be expressed as

$$V_2(\theta) = \underline{\varepsilon}'(t+b) S \underline{\varepsilon}(t+b) \quad (A4.12)$$

A necessary condition for a minimum is that

$$\frac{\partial V_2(\theta)}{\partial \theta} = \underline{0} \quad (A4.13)$$

and this leads to the equation

$$\underline{X}' S \underline{X} \underline{\theta} = \underline{X}' S \underline{Y}(t+b) \quad (A4.14)$$

where

$$\underline{X} = \begin{bmatrix} \underline{X}'(t) \\ \underline{X}'(t-1) \\ \vdots \\ \underline{X}'(1) \end{bmatrix} \quad (A4.15)$$

and

$$\underline{Y}'(t+b) = (Y(t+b), Y(t+b-1), \dots, Y(b)) \quad (\text{A4.15a})$$

Suppose that the control law (A4.3) converges to constant values, implying $X_{\underline{\theta}} = \underline{0}$. Substituting this into (A4.14) we see

$$X'SY(t+b) = \underline{0} \quad (\text{A4.16})$$

The elements of this vector are given by

$$\sum_{t=1}^N \lambda^{N-t} Y(t+k)Y(t) = 0, \quad k = b, b+1, \dots, b+m$$

$$\sum_{t=1}^N \lambda^{N-t} U(t+k)Y(t) = 0, \quad k = b, b+1, \dots, b+l \quad (\text{A4.17})$$

If the control has been implemented for a sufficient length of time for the process output to be covariance stationary, the sums in (A4.17) will approach constant values. Taking expectations of (A4.17) we obtain

$$\frac{1 - \lambda^N}{1 - \lambda} r_{yy}(\tau) = 0, \quad \tau = b, b+1, \dots, b+m$$

and

$$\frac{1 - \lambda^N}{1 - \lambda} r_{yu}(\tau) = 0, \quad \tau = b, b+1, \dots, b+l \quad (\text{A4.18})$$

Even though the expected value of (A4.17) is the same as the expected value when $\lambda = 1.0$, the variances of the two estimates will be different. In the next section, we show how one obtains the approximate variances of (A4.17). This work closely follows the original work of Bartlett (1946).

A4.3 Approximate Variances of the Weighted Autocorrelation Function

Define

$$\hat{r}_k = \frac{\frac{1}{n} \sum_{j=1}^{N-k} \lambda^{N-j} z_j z_{j+k}}{\frac{1}{n} \sum_{j=1}^N \lambda^{N-j} z_j z_j} \quad (\text{A4.19})$$

when $N \gg k$, we may ignore initial effects and define

$$\hat{r}_k = \frac{\frac{1}{n} \sum_{j=1}^N \lambda^{N-j} z_j z_{j+k}}{\frac{1}{n} \sum_{j=1}^N \lambda^{N-j} z_j z_j} = \frac{C_k}{V} \quad (\text{A4.20})$$

If we choose

$$n = \frac{1 - \lambda^N}{1 - \lambda} \quad (\text{A4.21})$$

we find that C_k and V are unbiased estimates of γ_k and σ_z^2 .

Expanding \hat{r}_k in a truncated Taylor series about the expected values of C_k and V , γ_k and σ_z^2 we obtain

$$\hat{r}_k - \frac{\rho_k \sigma_z^2}{\sigma_z^2} = \frac{(C_k - \rho_k \sigma_z^2)}{\sigma_z^2} - (V - \sigma_z^2) \frac{\rho_k \sigma_z^2}{\sigma_z^4} \quad (\text{A4.22})$$

Squaring (A4.22) and taking expectations we obtain

$$\text{var}(\hat{r}_k) = \frac{\text{var}(C_k)}{\sigma_z^4} - \frac{2\rho_k}{\sigma_z^4} \text{cov}(V, C_k) + \frac{\rho_k^2}{\sigma_z^4} \text{var}(V) \quad (\text{A4.23})$$

We will in turn evaluate the variance and covariance terms of (A4.23). First though we state the result, Anderson (1957) that if $\underline{z} \sim \text{MVN}(\underline{0}, \underline{\Sigma})$, with identical variances then

$$E(z_a z_{a+s} z_b z_{b+s+t}) = \sigma_z^4 (\rho_s \rho_{s+t} + \rho_{b-a} \rho_{b-a+t} + \rho_{b-a+s+t} \rho_{b-a-s}) \quad (\text{A4.24})$$

A4.3.1. Evaluation of $\text{var}(V)$

$$\begin{aligned} \text{var}(V) &= \text{cov} \left(\frac{1}{n} \sum_a \lambda^{N-a} z_a z_a, \frac{1}{n} \sum_b \lambda^{N-b} z_b z_b \right)^2 \\ &= \frac{1}{n^2} E \left(\sum_a \lambda^{N-a} z_a z_a \right)^2 - \left(E \left(\frac{1}{n} \sum_a \lambda^{N-a} z_a z_a \right) \right)^2 \end{aligned} \quad (\text{A4.25})$$

Assuming that $N-a = N-b = N$ then

$$\text{var}(V) = \frac{1}{n^2} \sum_{a,b} \lambda^{N-a} \lambda^{N-b} E(z_a z_a z_b z_b) - \sigma_z^4 \quad (\text{A4.26})$$

Using (A4.24) with $s=t=0$ we find

$$\text{var}(V) = \frac{\sigma_z^4}{n^2} \sum_{a,b} \lambda^{N-a} \lambda^{N-b} (\rho_0^2 + 2\rho_{b-a}^2) \quad (A4.27)$$

The summation may be expanded as follows

$$\begin{aligned} \sum_{a,b} \lambda^{N-a} \lambda^{N-b} \rho_{b-a}^2 = & \\ & \lambda^{N-1} (\rho_0^2 \lambda^{N-1} + \rho_1^2 \lambda^{N-2} + \dots + \rho_{N-1}^2 \lambda^0) + \\ & \lambda^{N-2} (\rho_{-1}^2 \lambda^{N-1} + \rho_0^2 \lambda^{N-2} + \dots + \rho_{N-2}^2 \lambda^0) + \\ & \lambda^0 (\rho_{-N+1}^2 \lambda^{N-1} + \dots + \rho_0^2 \lambda^0) \end{aligned} \quad (A4.28)$$

The N terms involving ρ_0^2 sum to

$$\frac{1 - \lambda^{2N}}{1 - \lambda^2} \cdot \rho_0^2 \quad (A4.29)$$

The $(N-1)$ terms involving ρ_1^2 sum to

$$\lambda \cdot \frac{1 - \lambda^{2(N-1)}}{1 - \lambda^2} \cdot \rho_1^2 \quad (A4.30)$$

and those involving ρ_{-1}^2 sum to

$$\lambda \cdot \frac{1 - \lambda^{2(N-1)}}{1 - \lambda^2} \cdot \rho_{-1}^2 \quad (A4.31)$$

Collecting all the terms of (A4.28) we find

$$\sum_{a,b} \lambda^{N-a} \lambda^{N-b} \rho_{b-a}^2 = \sum_{j=0}^{N-1} (\rho_j^2 + \rho_{-j}^2) \lambda^j \cdot \frac{1 - \lambda^{2(N-j)}}{1 - \lambda^2}$$

$$= \sum_{j=-N+1}^{N-1} \rho_j^2 \lambda^{|j|} \cdot \frac{1 - \lambda^{2(N-|j|)}}{1 - \lambda^2} \quad (\text{A4.32})$$

therefore

$$\text{var}(V) = 2 \frac{\sigma_z^4}{n^2} \cdot \sum_{j=-N+1}^{N-1} \rho_j^2 \lambda^{|j|} \cdot \frac{1 - \lambda^{2(N-|j|)}}{1 - \lambda^2} \quad (\text{A4.33})$$

since $\rho_0 = 1$.

A4.3.2 Evaluation of $\text{Cov}(V, C_k)$

$$\text{cov}(V, C_k) = \text{cov}\left(\frac{1}{n} \sum_a \lambda^{N-a} z_a z_a, \frac{1}{n} \sum_b z_b z_{b+k}\right)$$

$$= \frac{\sigma_z^4}{n^2} \sum_{a,b} \lambda^{N-a} \lambda^{N-b} (\rho_0 \rho_k + 2\rho_{b-a} \rho_{b-a+k}) - \sigma_z^4 \rho_0 \rho_k \quad (\text{A4.34})$$

The last result was obtained by using (A4.34) with $s = 0$, $t = k$.

Simplifying (A4.34) we find

$$\text{cov}(V, C_k) = \frac{2\sigma_z^4}{n^2} \cdot \sum_{j=-N+1}^{N-1} \rho_j \rho_{j+k} \lambda^{|j|} \cdot \frac{1 - \lambda^{2(N-|j|)}}{1 - \lambda^2} \quad (\text{A4.35})$$

A4.3.3 Evaluation of $\text{var}(C_k)$

$$\begin{aligned} \text{var}(C_k) &= E\left(\frac{1}{n} \sum_a \lambda^{N-a} z_a z_{a+k}, \frac{1}{n} \sum_b \lambda^{N-b} z_b z_{b+k}\right) \\ &= \frac{\sigma_z^4}{n^2} \sum_{a,b} \lambda^{N-a} \lambda^{N-b} (\rho_k^2 + \rho_{b-a}^2 + \rho_{b-a+k} \rho_{b-a-k}) \sigma_z^4 \rho_k^2 \end{aligned} \quad (\text{A4.36})$$

where we have used (A4.24) with $s = k$, $t = 0$.

Proceeding in a similar manner as before we obtain

$$\text{var}(C_k) = \frac{\sigma_z^4}{n^2} \sum_{j=-N+1}^{N-1} (\rho_j^2 + \rho_{j+k} \rho_{j-k}) \lambda^{|j|} \cdot \frac{1 - \lambda^{2(N-|j|)}}{1 - \lambda^2} \quad (\text{A4.37})$$

Substituting these results into (A4.23) we obtain finally

$$\begin{aligned} \text{var}(\hat{r}_k) &= \frac{1}{n^2} \sum_{j=-N+1}^{N-1} (\rho_j^2 + \rho_{j+k} \rho_{j-k} - 4\rho_k \rho_{j+k} \rho_j + 2\rho_k^2 \rho_j^2) \\ &\quad \lambda^{|j|} \cdot \frac{1 - \lambda^{2(N-|j|)}}{1 - \lambda^2} \end{aligned} \quad (\text{A4.38})$$

If we assume that the theoretical autocorrelations are zero beyond lag q , and that $q \ll N$ then (A4.38) reduces to

$$\text{var}(\hat{r}_k) = \frac{1}{n^2} \frac{1 - \lambda^{2N}}{1 - \lambda^2} \left(1 + 2 \sum_{j=1}^q \rho_j^2 \lambda^{2j}\right) \quad (\text{A4.39})$$