SIMULTANEOUS IDENTIFICATION AND CONTROL

A Thesis submitted to the
University of Leicester
for the degree of
Doctor of Philosophy

by

David Peter Turtle, B.Sc.

June 1970

UMI Number: U632529

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

UMI U632529

Published by ProQuest LLC 2015. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

X75 3061809

I Thesis submitted to the

University of Cambridge

for the degree of

Doctor of Philosophy

THESIS

384 284

10 11 71

David Peter Turner, BSc

1971

ABSTRACT

The established identification techniques of Astrom and Bohlin and Box and Jenkins are designed to estimate the parameters of both process and disturbance models from input/output records. The methods require open-loop experimentation on the process with a perturbation on the input from an external source. The resulting output fluctuations are often unacceptable for normal production, especially if the output exhibits drift. It is more desirable to attempt to identify and control simultaneously, avoiding the need for a prior experimental identification mode.

This thesis presents some techniques of simultaneous identification and control for single-input, single-output, linear dynamic systems. The methods, based on the Box and Jenkins control scheme, are simple to apply and require a relatively small amount of computation. A particularly valuable technique exploits a separation effect for processes with large time-delays. This method of simultaneous identification and control is applied to an industrial chemical process and provides sufficient information to allow for the design of a control system which substantially reduces the output variations.

Other techniques, presented in this thesis, which are not restricted by the magnitude of the process time-delay are not as yet generally applicable to industrial process-control problems.

ACKNOWLEDGEMENTS

The author is indebted to his supervisor Dr. P.H. Phillipson for suggesting the topic of research and for his continual interest, encouragement and guidance throughout the last three years.

To Professor G.D.S. MacLellan, the members of staff and research students of the Engineering Department at Leicester University, and several others the author wishes to express his thanks for their many helpful comments and discussions. Thanks are also due to the Science Research Council for their financial support from October 1967 to September 1969 and to the Engineering Department of Leicester University for the appointment of Research Demonstrator from October 1969.

The author also wishes to express his gratitude to Mrs. Helen Sheppard for preparing the typescript.

CONTENTS

Abstract	i
Acknowledgements	ii
Notation	v
CHAPTER I : Introduction	
1.1. Introduction to the problem	1.
1.2. Linear models and control theory	2.
1.3. Methods of identification	6.
1.4. A new approach to identification	10.
CHAPTER II : The Box and Jenkins control system	
2.1. Derivation of the control law	11.
2.2. Steady-state analysis of the control system	14.
2.3. System sensitivity	17.
2.4. Estimation of disturbance models	19.
2.5. The design of linear predictors	25.
2.6. Summary	30.
CHAPTER III : Updating the Box and Jenkins control scheme	
Introduction	32.
3.1. The updating equation	32.
3.2. Updating the process model	34.
3.3. Updating the predictor	36
3.4. The separation effect	41.
3.5. The separation effect as a means of identification	48.
3.6. Summary and conclusions	50.
CHAPTER IV : Simulation studies	
Introduction	52.

4.1. Estimation of autoregressive models	52.
4.2. Prediction	55.
4.3. Updating analyses	57.
4.4. Summary	68.
CHAPTER V : An industrial investigation	
5.1. Introduction	70.
5.2. Determining the empirical model	72.
5.3. Application of the separation technique of identification	74.
5.4. Redesign of the control system	83.
5.5. Summary and conclusions	93.
CHAPTER VI : Further methods of simultaneous identification and control	
Introduction	95.
6.1. Generalisation of the updating equation	95.
6.2. The separation effect for mixed models	98.
6.3. A general method of simultaneous identification and control	100.
6.4. Sensitivity of the general method to estimation errors	107.
6.5. Some problems associated with simultaneous identification and control	111.
CHAPTER VII : Conclusions	
7.1. Simultaneous identification and control as a practical proposition	115.
7.2. Applications	115.
7.3. Topics for further research	116.
References	128.
Publications	132.

NOTATION

The following standard notation is used throughout this thesis. Other symbols will be defined where necessary.

T	:	sampling interval
t	:	discrete time in integral multiples of the sampling interval
$\mathcal{Z}$	:	$\mathcal{Z}$ -transform defined as a forward time-shift of one sample interval, e.g. $\mathcal{Z}x_t = x_{t+1}$
s	:	Laplace transform operator
$\{a_t\}$	:	residual sequence
$\{c_t\}$	:	process input sequence
$\{e_t\}$	:	error sequence
$\{x_t\}$	:	process output-related disturbance sequence
$\{E_t\}$	:	measured process output sequence
$\{X_t\}$	:	estimated disturbance sequence
$\{\zeta_t\}$	:	sequence of independent random variables with a zero-mean, otherwise referred to as white noise
$\{E_t^d\}$	:	desired output sequence, also referred to as the set-point
$A(\mathcal{Z}^{-1})$	:	denominator polynomial of the process pulse transfer function
$B(\mathcal{Z}^{-1})$	:	numerator polynomial of the process pulse transfer function
$H(\mathcal{Z}^{-1})$	:	polynomial representing the predictor
$H^*(\mathcal{Z}^{-1})$	:	polynomial representing the optimum predictor
$M(\mathcal{Z}^{-1})$	:	process-model impulse-response polynomial
$P(\mathcal{Z}^{-1})$	:	process impulse-response polynomial

$Q(z^{-1})$	:	autoregressive polynomial in the disturbance model
$R(z^{-1})$	:	autoregressive polynomial in the model of the estimated disturbance
$S(z^{-1})$	:	moving-average polynomial in the disturbance model
$W(z^{-1})$	:	polynomial representing the controller
k	:	number of sample intervals in the process time-delay
l	:	number of sample intervals in the process-model time- delay
K	:	steady-state gain of the process
N	:	number of data samples
ϵ	:	expectation operator
μ	;	ensemble average
σ	:	standard deviation
λ	:	variance of the white-noise sequence $\{\zeta_t\}$
τ	:	supplementary time variable
Δ	:	difference operator defined as $1 - z^{-1}$
$\underline{\theta}$	:	vector of parameters
$\underline{\Phi}$	:	matrix of observations
opt	:	optimum value

CHAPTER IINTRODUCTION1.1. Introduction to the problem

Many theoretical techniques have now been developed for statistical identification. However, practical applications have in the past been limited by restraints on the type of allowable experimentation. More recently emphasis has been placed on the development of identification techniques which incur a minimum of economic penalty in their application. To achieve this there appears to be a need for methods of identification which apply to processes operating in a closed-loop mode ¹. The aim of this work has been to develop such a method.

The most significant area for the application of statistical identification is in the chemical process industry; this field offers a wide range of well-defined regulatory control problems. Since the analytical design of non-linear stochastic control systems has not yet reached the point of practical applicability ², interest at present lies in the control and identification of linear processes. Control algorithms for a known, time-variable, linear process have been developed using the theory of optimal control in combination with Kalman filtering ³; the two problems of control and state estimation have been treated independently by applying the separation theorem of optimal control. This approach employs the state-variable representation of dynamic systems although similar, but less general, algorithms exist for the conventional transfer function representations.

In line with most of the recent literature, discrete models of dynamic systems will be used throughout this work. Of the many reasons given for using discrete models the extensive application of the digital computer in the chemical process industry is undoubtedly the most

significant. It will also be convenient to consider only single-input-single-output systems, although extensions to the multivariable case will be mentioned where appropriate. For this simple class of systems the transfer function representation is probably more convenient than state-variables, and will be employed throughout this thesis. The next few sections will outline the various established approaches to the identification problem as a basis for the consideration of simultaneous identification and control in the remainder.

1.2. Linear models and control theory

This section considers the representation of single-input-single-output, linear discrete dynamic systems and the corresponding control algorithms.

Firstly consider a stochastic signal $[x_t]$, represented as the output of a constant-coefficient linear filter whose input is a sequence of independent random variables $[\zeta_t]$; i.e.

$$x_t = F(z^{-1})\zeta_t \quad (1.2.1)$$

$$\text{where } F(z^{-1}) = \sum_{i=0}^{\infty} f_i z^{-i}$$

and $[\zeta_t, t = -\infty, \dots, 0, \dots, \infty]$ is defined by

$$\begin{aligned} \epsilon[\zeta_t] &= 0 \\ \epsilon[\zeta_{t+i} \zeta_{t+j}] &= 0 \quad \text{for } i \neq j \\ &= \lambda \quad \text{for } i = j \end{aligned} \quad (1.2.2)$$

As Kalman⁴ notes, a white noise input $[\zeta_t]$ is employed because of its obviously advantageous statistical properties, and a linear filter is assumed since, in practice, insufficient a priori information is available to allow precise definition of a non-linear model.

From eqns. (1.2.1), (1.2.2) it follows that

$$\epsilon[x_t x_{t-j}] = \lambda \sum_{i=0}^{\infty} f_i f_{i+j} \text{ for all } j \quad (1.2.3)$$

Now, eqn. (1.2.3) can still be satisfied if either f_0 or λ is chosen arbitrarily. There will therefore be no loss of generality in the following if f_0 is taken to be unity. An alternative representation for the random process $[x_t]$ is then

$$x_t = \frac{1 + S(z^{-1})}{1 - Q(z^{-1})} \zeta_t \quad (1.2.4)$$

$$\text{where } S(z^{-1}) = \sum_{i=1}^m S_i z^{-i} \text{ and } Q(z^{-1}) = \sum_{i=1}^n q_i z^{-i}$$

If $S(z^{-1}) = 0$ in eqn. (1.2.4) the random process $[x_t]$ is said to be an autoregressive process; also, if $Q(z^{-1}) = 0$, $[x_t]$ is a moving-average process. Otherwise the model of eqn. (1.2.4) will be referred to as mixed moving-average-autoregressive.

For eqn. (1.2.1) or (1.2.4) to represent a weak-stationary random process the linear filter must be stable; i.e. there must be a bounded output $[x_t]$ for a bounded input $[\zeta_t]$. The filter in eqn. (1.2.4) is stable if and only if all its poles lie inside the unit circle in the z -plane; i.e. the roots of $1 - Q(z^{-1}) = 0$ must lie inside the unit circle. Also, if the polynomial $1 + S(z^{-1})$ is constrained to have all its roots inside the unit circle then the model of eqn. (1.2.4) is invertible and identifiable⁵. Stationarity, invertibility and their implications are thoroughly discussed by Box & Jenkins⁵.

Some allowance can be made in the model of eqn. (1.2.4) for types of non-stationary random processes. Consider a signal $[w_t]$ consisting of a stationary random process $[x_t]$ superimposed on an 1th-order deterministic trend $[u_t]$; i.e.

$$w_t = u_t + x_t = \sum_{i=0}^1 V_i t^i + \frac{1 + S(z^{-1})}{1 - Q(z^{-1})} \zeta_t \quad (1.2.5)$$

where the V_i are constants. It can be shown that by differencing 1 times, eqn. (1.2.5) reduces to

$$\Delta^1 w_t = 1V_p + \frac{1 + S(z^{-1})}{1 - Q(z^{-1})} \zeta'_t$$

where $\Delta = 1 - z^{-1}$ and ζ'_t is another white noise process. The zero-mean random process $w'_t = w_t - \epsilon[w_t]$ can then be represented by the integrated model

$$w'_t = \frac{1 + S(z^{-1})}{\Delta^1(1 - Q(z^{-1}))} \zeta'_t \quad (1.2.6)$$

The model of eqn. (1.2.6) describes an accumulated random process⁶ which can also represent a more general evolutionary type of non-stationary behaviour.

The general discrete representation of a linear noisy dynamic system can now be formed using eqn. (1.2.4) as a model for the process random disturbance $[x_t]$; i.e.

$$E_t = \frac{z^{-k}B(z^{-1})}{1 + A(z^{-1})} c_t + \frac{1 + S(z^{-1})}{1 - Q(z^{-1})} \zeta_t \quad (1.2.7)$$

where $[E_t, t = 0, 1, \dots]$ is the measurable process output, $[c_t, t = 0, 1, \dots]$ is the process input and k is the process pure time-delay. The polynomials $A(z^{-1})$, $B(z^{-1})$ are given by

$$A(z^{-1}) = \sum_{i=1}^I a_i z^{-i}, \quad B(z^{-1}) = \sum_{i=0}^J b_i z^{-i}$$

Box & Jenkins⁵, amongst others, employ the model of eqn. (1.2.7) to form an optimum control algorithm for a linear discrete system; the derivation will be presented in Chapter II.

A variation of eqn. (1.2.7) is used by Astrom^{7,8} in his important contributions to stochastic identification and control; viz

$$[1 + A'(z^{-1})] E_t = z^{-k} B'(z^{-1}) c_t + \lambda [1 + S'(z^{-1})] \xi_t \quad (1.2.8)$$

where $1 + A'(z^{-1}) = [1 + A(z^{-1})][1 - Q(z^{-1})] = 1 + \sum_{i=1}^L a'_i z^{-i}$

$$B'(z^{-1}) = B(z^{-1}) [1 - Q(z^{-1})] = \sum_{i=0}^L b'_i z^{-i}$$

$$1 + S'(z^{-1}) = [1 + S(z^{-1})][1 + A(z^{-1})] = 1 + \sum_{i=1}^L S'_i z^{-i}$$

and where any, or all, of the coefficients $a'_1, \dots, a'_L, b'_1, \dots, b'_L, S'_1, \dots, S'_L$ can be zero.

The optimum control algorithm which Astrom develops for the system eqn. (1.2.8) is similar to that of Box & Jenkins. The general form of the equation is

$$c_t = - \frac{z^k H(z^{-1}) [1 + A(z^{-1})]}{B(z^{-1}) [1 - H(z^{-1})]} E_t \quad (1.2.9)$$

where $H(z^{-1}) = \sum_{i=0}^K h_i z^{-k-i}$

is the k -step ahead linear predictor for the disturbance $[x_t]$. The derivation of optimum linear predictors is given in Chapter II.

Other forms of the model (1.2.7) have been used for identification or control. For instance, the process may be represented by its impulse response $P(z^{-1})$ where

$$P(z^{-1}) = B(z^{-1}) [1 + A(z^{-1})]^{-1} = \sum_{i=0}^{\infty} p_i z^{-i} \quad (1.2.10)$$

In practice, the process impulse response can often be truncated with a negligible loss of accuracy; i.e. if the process is inherently stable then the series $p_0, p_1, \dots, p_i \dots$ converges to zero with increasing i .

Hence

$$P(z^{-1}) \approx \sum_{i=0}^I p_i z^{-i} \quad (1.2.11)$$

A process described by the truncated impulse response of eqn. (1.2.11) is referred to as a finite-memory process.

The choice of a particular form of the model eqn. (1.2.7) must be related to the individual problem. However, some general comments can be made; firstly, the model of eqn. (1.2.7) will, in general, contain fewer parameters than the other models - the parsimonious representation recommended by Box & Jenkins^{5,9}. This advantage may be outweighed by the less complex structure of control (and identification) algorithms for the Astrom (1.2.8) and finite-memory (1.2.11) models. Secondly, a non-minimum phase process presents problems in the implementation of the control law (1.2.9), since the inverse process model will then be unstable; the problem is less difficult to overcome when the process is represented as in eqns. (1.2.7) and (1.2.8). Finally, the control law (1.2.9) can easily be generalised to multi-input processes¹⁰ although extension to multiple outputs is less straightforward.

1.3. Methods of identification

The computation of the control algorithm (1.2.9) obviously requires knowledge of the parameters of the system eqn. (1.2.7). It is sometimes possible to derive estimates of these parameters from mechanical, chemical and thermodynamical analyses of the system. More often it is only possible to obtain estimates of the model order and rough approximations to the process parameters. In the latter situation it is usually necessary to employ a statistical technique of identification.

The many techniques existing for estimation of the parameters of a linear noisy dynamic system differ not only in their theoretical bases but also in the required form of experimentation. For purposes of discussion the methods will be broadly classified; firstly, the theoretical approaches (as suggested by Balakrishnan & Peterka¹) -

- (i) minimization of output error,
- (ii) minimization of equation error,

(iii) statistical error criterion.

Class (i) techniques compare system and model outputs to give an output error $[e_t]$, which is then minimised in a mean-square sense. The method has most often been used to identify the process impulse response (1.2.11),

$$\text{i.e.} \quad E_t = \sum_{i=0}^I p_i c_{t-k-i} + x_t \quad (1.3.1)$$

where the relationship between correlation functions is used, viz

$$\phi_{Eo}(1) = \sum_{i=0}^I p_i \phi_{cc}(1-k-i) \text{ for } k < 1 < (k+I) \quad (1.3.2)$$

assuming $\phi_{xc}(1) = 0$

Minimization of output error also forms the basis of many model adjustment techniques¹¹.

The second class forms an equation error, from (1.2.7)

$$e_t = [1 + A(z^{-1})] E_t - B(z^{-1}) c_{t-k} \quad (1.3.3)$$

Then from a set of data pairs $[E_t, c_t, t = 1, \dots, N]$, N equations (1.3.3) can be formed; in matrix form

$$\underline{e} = \underline{\Phi} \underline{\theta} \quad (1.3.4)$$

where $\underline{e}$ is an N -vector of e 's, $\underline{\Phi}$ is an $(N \times 2n)$ matrix of E 's and c 's, and $\underline{\theta}$ is a $2n$ -vector of parameters a_i, b_i . Weighted least squares is normally used to solve eqn. (1.3.4) for $\underline{\theta}$ ¹². The major problem with this technique is the elimination of bias caused by non-zero correlation between $\underline{e}$ and $\underline{\Phi}$ ¹. Stochastic approximation algorithms^{13,14} have been extensively employed to obtain estimates of the parameter vector $\underline{\theta}$; these methods, however, often impose restrictive conditions on the form of the model (1.2.7).

Class (iii) techniques minimise a function $(\underline{\theta} - \hat{\underline{\theta}})^T (\underline{\theta} - \hat{\underline{\theta}})$ where $\hat{\underline{\theta}}$ is an estimate of $\underline{\theta}$. The method of maximum likelihood¹⁵

provides a general approach to the solution of this problem. However, for the models of eqns. (1.2.7) and (1.2.8) the method does not yield a straightforward solution, since the estimates are non-linear in the parameters. Box & Jenkins⁹ and Astrom⁸ have proposed estimation techniques based on maximum likelihood but various computational problems still remain to be solved; e.g. for a multi-peak likelihood function. It has been noted¹⁶ that linear-least-squares can be used when the disturbance $[x_t]$ is white noise.

Identification techniques can also be classified according to the type of data required from experimentation, i.e.

- (a) input-output pairs from normal-operating records,
- (b) output measurements from the open-loop process which is excited by a random perturbation,
- (c) output measurements from the closed-loop process with an added input perturbation.

Many techniques claim the use of process normal-operating records. One of the earliest applications of statistical identification (by Florentin et al¹⁷) provided evidence that such records are not, in general, satisfactory for estimation purposes. The reason is that most normal-operating records represent an attempt at manual feedback control; in these circumstances the assumption in eqn. (1.3.2) that the input $[c_t]$ is uncorrelated with the disturbance $[x_t]$ is not, in general, valid. This assumption is an integral part of the output and equation error methods ((i) and (ii) above) and it can also be shown¹⁸ that statistical error methods produce biased results when there is feedback. Consider the system given by eqn. (1.2.7) with

$$c_t = - f(E_t) + w_t \quad (1.3.5)$$

where f is an unknown function representing manual feedback and $\{w_t\}$ is a random process uncorrelated with the disturbance $\{x_t\}$. Then from

eqns. (1.2.7) and (1.3.5),

$$E_t + \frac{z^{-k}B(z^{-1})}{1 + A(z^{-1})} f(E_t) = \frac{1 + S(z^{-1})}{1 - Q(z^{-1})} \zeta_t + \frac{z^{-k}B(z^{-1})}{1 + A(z^{-1})} w_t \quad (1.3.6)$$

A class (iii) identification procedure will only be able to estimate the closed-loop system (1.3.6), and since this contains f a solution for $B(z^{-1})$, $A(z^{-1})$ is not, in general, possible. A further probable disadvantage of normal-operating records is that input changes will be infrequent so that a long record is required for accurate estimation; such records may not satisfy Astrom's⁸ condition of a sufficiently exciting input.

To overcome the limitations of normal-operating records many authors have suggested the use of an externally applied random input perturbation; this is usually a psuedo-random binary sequence because of its desirable statistical properties¹¹, and the convenience of on-site implementation. The input perturbation can either be superimposed on the existing process input or used as the only input. Many applications have been cited in the literature, notably by Briggs et al¹⁹ using a correlation technique, Astrom⁷ using maximum-likelihood, Clarke²⁰ using generalized least-squares. These methods have one feature in common - the collection of a set of data pairs $\{E_t, c_t\}$ from a process operated in open-loop. Unfortunately this is only practical for a small number of industrial chemical processes. In many systems the disturbance contains some form of drift which will usually cause large output deviations if the process is operated in open-loop. This situation is unlikely to be condoned by plant managers who must give priority to the economics of production.

Also, consider the system

$$E_t = P(z^{-1})c_{t-k} + x_t \quad (1.3.7)$$

where $\{c_t\}$ is a white-noise sequence of variance v .

$$\text{Then } \text{var} \{E_t\} = \nu \sum_{i=0}^I p_i^2 + \text{var} [x_t] \quad (1.3.8)$$

That is, the white noise input perturbation increases the open-loop output variance by a factor proportional to the variance of the perturbation. Also, the period of data collection will be dependent on the ratio of disturbance variance to perturbation variance for a given accuracy of estimation²¹.

A random perturbation can also be added to an existing input while the process is under feedback control. If the feedback function f in eqn. (1.3.6) is known a priori it will then be possible to solve for all the system parameters of eqn. (1.2.7). Although the method again introduces an extra source of output error, operation in a closed-loop mode is more desirable from an economic point of view. Box & Jenkins²² propose that their estimation procedure can be applied to a closed-loop system although practical applications have experienced some difficulty in achieving reliable results²³.

1.4. A new approach to identification

The most successful of the reported applications of statistical identification have employed a random input perturbation in a prior experimental mode. However, optimum stochastic control systems, for which these methods are intended to provide design information, are most easily justified in cases where fluctuation of the output incurs a serious economic penalty². It is in precisely such situations that perturbation and open-loop operation of the process are least acceptable. It therefore appears desirable to attempt to identify and control simultaneously without the need for prior experimentation.

The approach to simultaneous identification and control presented in this thesis is based on the Box & Jenkins control system, which will be described in detail before proceeding to develop the identification procedures.

CHAPTER IITHE BOX & JENKINS CONTROL SCHEME2.1. Derivation of the control law

In this section a control scheme based on the work of Box & Jenkins^{9,26} will be derived from first principles. The scheme will not be precisely as proposed by Box & Jenkins since they were not concerned with changes in the desired value. The following theorem presents the optimum stochastic control law for a single-input, single-output, linear discrete dynamic system where the desired value is a known deterministic signal.

THEOREM 2.1.

Consider the linear discrete dynamic system represented by

$$E_t = z^{-k} P(z^{-1}) c_t + x_t \quad (2.1.1)$$

where $P(z^{-1}) = \sum_{i=0}^{\infty} p_i z^{-i}$ and $\sum_{i=0}^{\infty} |p_i| < \infty$,

i.e. the process $P(z^{-1})$ is inherently stable.

Let the system random disturbance $\{x_t\}$ be represented by an infinite-order moving-average model,

$$x_t = [1 + S(z^{-1})] \zeta_t \quad (2.1.2)$$

where $S(z^{-1}) = \sum_{i=1}^{\infty} s_i z^{-i}$

Define the system error $\{e_t\}$ as the difference between the measured and desired outputs,

$$e_t = E_t - E_t^d \quad (2.1.3)$$

Then (i) the optimum control law which minimizes the mean-square error $L = \epsilon\{e_t^2\}$ is given by,

$$c_t = \frac{1}{P(z^{-1})} \left[E_t^d - z^k H_T^*(z^{-1}) x_t \right] \quad (2.1.4)$$

where $H_T^*(z^{-1})$, the optimum k^{th} -step ahead linear predictor for $\{x_t\}$, is defined by

$$H_T^*(z^{-1}) = \frac{S_T(z^{-1})}{1 + S(z^{-1})} \quad ; \quad S_T(z^{-1}) = \sum_{i=k}^{\infty} S_i z^{-i} \quad (2.1.5)$$

and $[\cdot]_T$ is the truncation operator defined in Appendix I,

(ii) at the optimum, the output error is given by

$$\text{opt}(e_t) = [1 + S(z^{-1}) - S_T(z^{-1})] \zeta_t \quad (2.1.6)$$

and the cost function by

$$L_{\min} = \lambda \left[1 + \sum_{i=1}^{k-1} S_i^2 \right] \quad (2.1.7)$$

Proof

From eqns. (2.1.1) and (2.1.2)

$$E_t = \sum_{i=0}^{\infty} p_i c_{t-k-i} + \zeta_t + \sum_{i=1}^{\infty} S_i \zeta_{t-i} \quad (2.1.8)$$

It is noted, from eqn. (2.1.8) that due to the process pure time-delay z^{-k} the output at time t can only be influenced by an input at time $(t-k)$. We therefore minimise L with respect to c_{t-k} .

Now, substituting eqn. (2.1.8) into eqn. (2.1.3) and expanding,

$$\begin{aligned} L = \{e_t^2\} = \epsilon \left\{ \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_i p_j c_{t-k-i} c_{t-k-j} + \zeta_t^2 + \sum_{i=1}^{\infty} S_i^2 \zeta_{t-i}^2 + (E_t^d)^2 \right. \\ \left. + 2 \sum_{i=0}^{\infty} p_i c_{t-k-i} \zeta_t + 2 \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} p_i S_j c_{t-k-i} \zeta_{t-j} \right. \\ \left. - 2E_t^d \sum_{i=0}^{\infty} p_i c_{t-k-i} - 2E_t^d \zeta_t - 2E_t^d \sum_{i=1}^{\infty} S_i \zeta_{t-i} \right\} \quad (2.1.9) \end{aligned}$$

But $c_t = f_1(E_t, E_{t-1}, \dots, E_t^d, E_{t-1}^d, \dots) = f_2(c_{t-k}, c_{t-k-1}, \dots, E_t^d, \dots, x_t, x_{t-1}, \dots)$

or $c_t = f_3(c_{t-k}, c_{t-k-1}, \dots, E_t^d, \dots, \zeta_t, \zeta_{t-1}, \dots)$, implying that c_t and ζ_{t+i} are independent for $i > 0$; i.e. $\epsilon \{ c_{t-k} \zeta_{t-i} \} = 0$ for $i < k$.

This result is used to reduce eqn. (2.1.9) before differentiating L

with respect to c_{t-k} to find a supremum of L . It then follows that

$$\frac{\partial L}{\partial c_{t-k}} = \epsilon \left\{ 2p_0 \sum_{i=0}^{\infty} p_i c_{t-k-i} + 2p_0 \sum_{j=k}^{\infty} S_j \zeta_{t-j} - 2p_0 E_t^d \right\} = 0 \text{ for max or min.}$$

Then L is minimised when

$$\sum_{i=0}^{\infty} p_i c_{t-k-i} + \sum_{j=k}^{\infty} S_j \zeta_{t-j} = \epsilon \{E_t^d\} \quad (2.1.10)$$

If $\{E_t^d\}$ is a known deterministic signal we can assume that $\epsilon\{E_t^d\} = E_{t-k}^d$, so eqn. (2.1.10) becomes

$$z^{-k} P(z^{-1}) c_t + S_T(z^{-1}) \zeta_t = E_{t-k}^d \quad (2.1.11)$$

$$\text{where } S_T(z^{-1}) = \sum_{j=k}^{\infty} S_j z^{-j}$$

and $[\cdot]_T$ is the k -th order truncation operator defined in Appendix I.

The optimum control algorithm is, from eqn. (2.1.11), given by

$$c_t = \frac{1}{P(z^{-1})} [E_t^d - z^k S_T(z^{-1}) \zeta_t] \quad (2.1.12)$$

Alternatively, substituting for ζ_t from eqn. (2.1.2).

$$c_t = \frac{1}{P(z^{-1})} \left[E_t^d - z^k H_T^*(z^{-1}) x_t \right]$$

$$\text{where } H_T^*(z^{-1}) = \frac{S_T(z^{-1})}{1 + S(z^{-1})}, \text{ Q.E.D(i)}$$

The design of an optimum predictor $H_T^*(z^{-1})$ will be presented in section 2.5.

It now follows, from eqns. (2.1.12), (2.1.1) and (2.1.3), that the optimum error is given by

$$\text{opt } \{e_t\} = [1 + S(z^{-1}) - S_T(z^{-1})] \zeta_t$$

$$\text{or } \text{opt } \{e_t\} = \sum_{i=1}^{k-1} S_i \zeta_{t-i} + \zeta_t$$

$$\text{and } L_{\min} = \{ \text{opt}(e_t)^2 \} = \lambda \left[1 + \sum_{i=1}^{k-1} S_i^2 \right] \text{ Q.E.D (ii)}$$

Fig.2.1. THE BOX & JENKINS CONTROL SCHEME

It is apparent from the form of the optimum control law (2.1.4) that c_t cannot be calculated without knowledge of $P(z^{-1})$, $H_T^*(z^{-1})$, k and $\{x_t\}$. Let $M(z^{-1})$, $H(z^{-1})$ and l be estimates of $P(z^{-1})$, $H_T^*(z^{-1})$ and k respectively. Then we can form a signal

$$X_t = E_t - z^{-1} M(z^{-1}) c_t \quad (2.1.13)$$

and from eqn. (2.1.1)

$$X_t = x_t + [z^{-k} P(z^{-1}) - z^{-1} M(z^{-1})] c_t$$

where if $M(z^{-1}) = P(z^{-1})$ and $l = k$,

$$X_t = x_t$$

The signal $\{X_t\}$ is, therefore, an estimate of $\{x_t\}$ and will be referred to as the estimated disturbance.

The practical implementation of the control law (2.1.4) will then be

$$c_t = \frac{1}{M(z^{-1})} \left[E_t^d - z^{-1} H(z^{-1}) X_t \right] \quad (2.1.14)$$

This is illustrated in a block-diagram of the control system shown in Fig. 2.1.

2.2. Steady-state analysis of the control system

In practice it is not realistic to assume a perfect knowledge of the process pulse transfer function or the disturbance model. It is, therefore, of interest to analyse the performance of the control system where allowance is made for estimation errors.

The control system of Fig. 2.1 is completely described by eqns. (2.1.1), (2.1.13) and (2.1.14). By combining these equations it can be shown that the system output is given by

$$E_t = \frac{M(z^{-1}) [1 - H(z^{-1})] x_t + z^{-k} P(z^{-1}) E_t^d}{M(z^{-1}) [1 - H(z^{-1})] + z^j P(z^{-1}) H(z^{-1})} \quad (2.2.1)$$

where $j = 1 - k$ is the difference between the process model and process

time delays. This system has been designed to (i) force the output $\{E_{t+k}\}$ to follow the set-point $\{E_t^d\}$, and (ii) minimise the effects of the random disturbance $\{x_t\}$. Then by considering the level of the system output the following criterion of performance can be formulated:-

CRITERION 2.1

The steady-state value of the ensemble average of the output should equal the steady-state value of the set-point, i.e.

$$\lim_{t \rightarrow \infty} \{\mu_E(t)\} = \lim_{t \rightarrow \infty} \{E_t^d\} \quad (2.2.2)$$

where $\mu_E(t)$ is the ensemble average of $\{E_t\}$ at time t .

Then from eqn. (2.2.1),

$$\begin{aligned} \lim_{t \rightarrow \infty} \{\mu_E(t)\} &= \lim_{t \rightarrow \infty} \left\{ \frac{M(z^{-1}) [1-H(z^{-1})]}{M(z^{-1}) [1-H(z^{-1})] + z^j P(z^{-1}) H(z^{-1})} \mu_x(t) \right\} \\ &+ \lim_{t \rightarrow \infty} \left\{ \frac{z^{-k} P(z^{-1})}{M(z^{-1}) [1-H(z^{-1})] + z^j P(z^{-1}) H(z^{-1})} E_t^d \right\} \end{aligned}$$

where $\mu_x(t)$ is the ensemble average of $\{x_t\}$ at time t .

It follows that eqn. (2.2.2) is true if and only if the following two criteria are simultaneously satisfied:-

CRITERION 2.2

$$\lim_{t \rightarrow \infty} \left\{ \frac{M(z^{-1}) [1-H(z^{-1})]}{M(z^{-1}) [1-H(z^{-1})] + z^j P(z^{-1}) H(z^{-1})} \mu_x(t) \right\} = 0 \quad (2.2.3)$$

CRITERION 2.3

$$\lim_{t \rightarrow \infty} \left\{ \frac{z^{-k} P(z^{-1})}{M(z^{-1}) [1-H(z^{-1})] + z^j P(z^{-1}) H(z^{-1})} E_t^d \right\} = \lim_{t \rightarrow \infty} \left\{ E_t^d \right\} \quad (2.2.4)$$

Now in chemical process control problems the disturbance $\{x_t\}$ can usually be represented by a zero-mean stationary random process superimposed on a deterministic 'drift' or 'trend'.⁵ This type of non-stationary disturbance can be described by a variation in the ensemble

average of $\{x_t\}$, where for practical processes the variation is bounded.

Then from Theorem A.2 in Appendix II, criterion 2.2 is satisfied if and only if

$$\frac{M(z^{-1}) [1-H(z^{-1})]}{M(z^{-1})[1-H(z^{-1})] + z^j P(z^{-1})H(z^{-1})} \bigg|_{z=1} = 0 \quad (2.2.5)$$

which reduces to the condition,

$$H(z^{-1}) \bigg|_{z=1} = 1 \quad (2.2.6)$$

It is also usual, in the chemical process industry, for set-point changes to be limited to step functions. Then by applying the z -transform final-value theorem²⁴ to eqn. (2.2.4), criterion 2.3 is satisfied for a unit step change in E_t^d if

$$\frac{z^{-k} P(z^{-1})}{M(z^{-1})[1-H(z^{-1})] + z^j P(z^{-1})H(z^{-1})} \bigg|_{z=1} = 1 \quad (2.2.7)$$

which is satisfied if

$$\left. \begin{aligned} H(z^{-1}) \bigg|_{z=1} &= 1 \\ \text{or } M(z^{-1}) \bigg|_{z=1} &= P(z^{-1}) \bigg|_{z=1} \end{aligned} \right\} \quad (2.2.8)$$

It can also be shown, by a similar argument to that presented in Appendix II, that criterion 2.3 is more generally satisfied by eqn. (2.2.8) if $\{E_t^d\}$ is a bounded deterministic signal over the finite interval $0 < t < \infty$.

By combining the conditions of eqns. (2.2.6) and (2.2.8) it can be seen that the steady-state performance criterion 2.1 is satisfied if and only if,

$$H(z^{-1}) \bigg|_{z=1} = 1$$

i.e. the steady-state gain of the predictor must be unity.

2.3. System sensitivity

It is well known that, under special circumstances, optimal control solutions may be highly sensitive to parameter variations. It is therefore of interest to investigate the sensitivity of the system (2.2.1) to errors in the process and disturbance models ⁷. For this purpose let

$$P(z^{-1}) = \frac{B(z^{-1})}{1 + A(z^{-1})}, \quad M(z^{-1}) = \frac{M_1(z^{-1})}{1 + M_2(z^{-1})}, \quad H(z^{-1}) = \frac{G(z^{-1})}{1 + F(z^{-1})}$$

where the polynomials B, A, M_1, M_2, G and F are of finite order. Then eqn. (2.2.1) becomes

$$E_t = \frac{[M_1(1+A)(1+F-z^{-1}G)] x_t + [z^{-k}B(1+M_2)(1+F)] E_t^d}{M_1(1+A)(1+F-z^{-1}G) + z^{-k}BG(1+M_2)} \quad (2.3.1)$$

where, for convenience, the z^{-1} argument of all the polynomials has been omitted.

The dynamical system represented by eqn. (2.3.1) has a number of modes ²⁵ equal to the degree of the characteristic polynomial

$$M_1(1+A)(1+F-z^{-1}G) + z^{-k}BG(1+M_2) \quad (2.3.2)$$

where each mode corresponds to a zero of the function (2.3.2). Now the inputs $\{x_t\}$ and $\{E_t^d\}$ might excite all the modes associated with the characteristic polynomial. This is acceptable if the modes are stable; if, however, the function (2.3.2) contains an unstable mode the system represented by eqn. (2.3.1) will be subject to an exponentially increasing error.

For small errors in the process and disturbance model parameters the modes of (2.3.2) will be close to those of the polynomial

$$B(1+A)(1+F) \quad (2.3.3)$$

n.b. the function (2.3.3) is obtained by setting $M_1=B$, $M_2=A$ and $l=k$ in (2.3.2).

Fig. 2.2. SUBOPTIMUM CONTROL SYSTEM FOR A NON-MINIMUM PHASE PROCESS

The function (2.3.3) will have unstable modes if any of its zeros lie outside the unit circle in the z -plane. Now, it has already been assumed (following eqn. (2.1.1)) that the process is inherently stable, i.e. that the function $1+A(z^{-1})$ has all its zeros inside the unit circle (n.b. most practical chemical processes satisfy this assumption). If it is further assumed that the predictor $H(z^{-1})$ is a stable filter - a feature which can be incorporated in the design choice - then the sensitivity of the system (2.3.1) depends only on the zeros of $B(z^{-1})$.

The polynomial $B(z^{-1})$ will have one or more zeros outside the unit circle if the process dynamics are non-minimum phase. In this situation it is of practical importance to derive a sub-optimal control law which is insensitive to parameter variations. Following the work of Astrom⁷, let

$$B(z^{-1}) = B_1(z^{-1}) B_2(z^{-1}) \quad (2.3.4)$$

where B_1 and B_2 contain roots inside and outside the unit circle respectively. Now consider the control system of Fig. 2.2., where the controller $W'(z^{-1})$ excludes the polynomial $B_2(z^{-1})$ but compensates for the change in steady-state gain, i.e.

$$W'(z^{-1}) = \frac{\gamma [1 + A(z^{-1})]}{B_1(z^{-1})} \quad (2.3.5)$$

$$\text{and } \gamma = \left[B_2(z^{-1}) \Big|_{z=1} \right]^{-1}$$

The system equation (2.3.1) then reduces to

$$E_t = z^{-k} \gamma B_2(z^{-1}) E_t^d + [1 - \gamma B_2(z^{-1}) H'(z^{-1})] x_t \quad (2.3.6)$$

This sub-optimal system will be insensitive to parameter variations if the predictor $H(z^{-1})$ is a stable filter. It should also be noted that the steady-state and average performance criteria are still satisfied if

$$H'(z^{-1}) \big|_{z=1} = 1$$

The coefficients of the polynomials $H'(z^{-1})$ and $D(z^{-1})$ can be calculated by equating coefficients of z^{-1} in the formula,

$$1 + S(z^{-1}) = D(z^{-1}) [1 - Q(z^{-1})] + [1 + S(z^{-1})] \gamma B_2(z^{-1}) H'(z^{-1}) \quad (2.3.7)$$

where $D(z^{-1}) = 1 - \gamma B_2(z^{-1}) H'(z^{-1})$

Eqn. (2.3.7) is equivalent to the formula used by Astrom⁷ in his sub-optimal system design.

2.4. Estimation of disturbance models

This section considers some methods of estimating the parameters of a model for a random disturbance $\{x_t\}$. The resulting techniques will be used in the identification procedures which are presented in the next few chapters.

The general linear model for a stationary random process is given by eqn. (1.2.4) as

$$x_t = \frac{1 + S(z^{-1})}{1 - Q(z^{-1})} \zeta_t \quad (2.4.1)$$

where $S(z^{-1}) = \sum_{i=1}^m S_i z^{-i}$ and $Q(z^{-1}) = \sum_{i=1}^n q_i z^{-i}$.

Box and Jenkins⁵ show that the problem of estimating the parameters S_i, q_i from a record $\{x_t, t = 1, \dots, N\}$ can be solved by the method of maximum likelihood. For the model of eqn. (2.4.1) the log-likelihood function is given by

$$l(\underline{\theta}) = -N \log \sigma_a - \frac{1}{2\sigma_a^2} \sum_{t=1}^n a_t^2 \quad (2.4.2)$$

where $\{a_t, t = 1, \dots, N\}$ is the sequence of residuals, with variance

$$\sigma_a^2, \text{ defined by} \quad a_t = \frac{1 - \hat{Q}(z^{-1})}{1 + \hat{S}(z^{-1})} x_t \quad (2.4.3)$$

and $\hat{Q}$, $\hat{S}$ are estimates of Q , S respectively. Then, for normally distributed residuals, the maximum of the log-likelihood function (2.4.2) coincides with the minimum sum of squares of the residuals.

Given estimates $\hat{Q}(z^{-1})$ and $\hat{S}(z^{-1})$ the residual sequence $\{a_t, t = 1, 2, \dots, N\}$ can be calculated iteratively from eqn. (2.4.3). The minimum sum of squares is then found by a hill-climbing technique (e.g. Astrom⁸ uses a Newton-Raphson algorithm). Unfortunately no simple analytic solution is available for this general problem due to its non-linear nature; i.e. the function $\partial a_t / \partial \theta_i$ is not independent of $\underline{\theta}$. Box & Jenkins⁵ propose that reliable estimates of the S_i , q_i are best obtained by an extensive graphical study of the likelihood function followed by a more precise search in the neighbourhood of the maximum. This procedure may avoid the many problems arising from multi-peak likelihood functions.³²

The estimation problem is greatly simplified if the disturbance model of eqn. (2.4.1) is purely autoregressive, i.e. if $S(z^{-1}) = 0$. The maximum likelihood method then reduces to a simple linear-least-squares solution.

In practice it may often be possible to transform the mixed model of eqn. (2.4.1) into a pure autoregressive model; consider

$$[1 + S(z^{-1})]^{-1} [1 - Q(z^{-1})] = 1 - R(z^{-1}) = 1 - \sum_{i=1}^{\infty} r_i z^{-i} \quad (2.4.4)$$

Then if the function $1 + S(z^{-1})$ is invertible (i.e. if all the zeros of $1 + S(z^{-1})$ lie inside the unit circle) the series $r_1, r_2, \dots, r_i, \dots$ converges to zero with increasing i . It is then possible to truncate $R(z^{-1})$ such that

$$R'(z^{-1}) = \sum_{i=1}^n r'_i z^{-i} \approx R(z^{-1})$$

where the error depends on the magnitude of n .

There is some practical justification for using only autoregressive models to describe chemical process disturbances. Firstly,

the low-pass filtering effect of most chemical processes is likely to introduce dominant autoregressive terms in the disturbance model. Secondly, it has been found in practice that an autoregressive model of 2 or 3 terms is, in general, sufficient to describe the disturbance. Some practical and simulation results to this effect will be presented in section 4.1. Finally, the disadvantage that an autoregressive model will, in general, contain more parameters than the corresponding mixed model is clearly outweighed by the ease and efficiency of autoregressive parameter estimation. It will also be found convenient, in Chapter III, to use the autoregressive representation of a random signal in the development of simultaneous identification and control procedures.

Consider a stationary random process $\{x_t, t = 1, 2, \dots, N\}$ and an autoregressive model,

$$x_t = Q(z^{-1})x_t + \zeta_t \quad (2.4.5)$$

$$\text{where } Q(z^{-1}) = \sum_{i=1}^n q_i z^{-i}$$

Then for an estimated model $\hat{Q}(z^{-1})$ a sequence of residuals $\{a_t, t = n+1, n+2, \dots, N\}$ can be formed from

$$a_t = [1 - \hat{Q}(z^{-1})]x_t, t = n+1, \dots, N \quad (2.4.6)$$

The $(N-n)$ equations (2.4.6) can be written in matrix form as

$$\underline{a} = \underline{x}_{n+1} - \underline{\Phi} \hat{\underline{g}} \quad (2.4.7)$$

where $\underline{x}_i$ and $\underline{a}$ are $(N-n)$ -vectors given by

$$\underline{x}_i^T = (x_i \ x_{i+1} \ \cdot \ \cdot \ \cdot \ x_{N+i-n-1}) \text{ for } i = 1, \dots, n+1$$

$$\underline{a}^T = (a_{n+1} \ a_{n+2} \ \cdot \ \cdot \ \cdot \ a_N)$$

$\hat{\underline{g}}$ is an n -vector of parameter estimates,

$$\hat{\underline{q}}^T = (\hat{q}_1 \quad \hat{q}_2 \quad \cdot \quad \cdot \quad \cdot \quad \hat{q}_n)$$

and $\underline{\Phi}$ is the $(N-n) \times n$ observation matrix whose columns are the vectors $\underline{x}_n, \underline{x}_{n-1}, \dots, \underline{x}_1$.

The sum of squares of the residuals is therefore given by

$$S = \underline{a}^T \underline{a} = \left[\underline{x}_{n+1}^T - \hat{\underline{q}}^T \underline{\Phi}^T \right] \left[\underline{x}_{n+1} - \underline{\Phi} \hat{\underline{q}} \right] \quad (2.4.8)$$

Then the vector $\hat{\underline{q}}$ which minimises the sum of squares S is given by the well-known linear-least-squares result ²⁷,

$$\hat{\underline{q}} = (\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T \underline{x}_{n+1} \quad (2.4.9)$$

and at the minimum

$$S_{\min} = \underline{x}_{n+1}^T \underline{x}_{n+1} - \hat{\underline{q}}^T \underline{\Phi}^T \underline{\Phi} \hat{\underline{q}} \quad (2.4.10)$$

It can also be shown that,

$$\epsilon\{\hat{\underline{q}}\} = \epsilon\{\underline{q}\} + \epsilon\{(\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T \underline{\zeta}\} \quad (2.4.11)$$

where $\underline{\zeta}$ is an $(N-n)$ - vector defined by

$$\underline{\zeta}^T = (\zeta_{n+1} \quad \zeta_{n+2} \quad \cdot \quad \cdot \quad \cdot \quad \zeta_N)$$

From eqn. (2.4.11) it can be seen that $\hat{\underline{q}}$ is an unbiased estimate of $\underline{q}$ if $\epsilon\{(\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T \underline{\zeta}\} = 0$. This condition may be satisfied if $\epsilon\{\underline{x}_i\} = 0$; $\hat{\underline{q}}$ is then asymptotically unbiased ²⁸.

To ensure that the n vectors $\underline{x}_i$ have zero means we can form a set of averages

$$V_i = \frac{1}{N-n} \sum_{j=i}^{N+i-n-1} x_j \quad \text{for } i = 1, \dots, n+1 \quad (2.4.12)$$

and a new set of zero-mean vectors $\underline{x}'_i$ defined by

$$\underline{x}'_i = \underline{x}_i - V_i \underline{e} \quad \text{for } i = 1, \dots, n+1 \quad (2.4.13)$$

where $\underline{e}$ is an $(N-n)$ unit vector. A new observation matrix $\underline{\Phi}'$ can also

be formed from the vectors $\underline{x}'_i$, $i = 1, \dots, n$. Then an asymptotically unbiased estimate $\hat{\underline{g}}$ of $\underline{g}$ can be obtained by replacing $\underline{x}_{n+1}$ and $\underline{\Phi}$ by $\underline{x}'_{n+1}$ and $\underline{\Phi}'$ in eqn. (2.4.9).

Now consider an element of the matrix $\underline{\Psi} = (\underline{\Phi}'^T \underline{\Phi})$,

$$\psi_{i,j} = \sum_{k=n+1}^N (x_{k-i} - v_{n+1-i})(x_{k-j} - v_{n+1-j}) \quad (2.4.14)$$

also
$$\psi_{i+h,j+h} = \sum_{k=n+1}^N (x_{k-i-h} - v_{n+1-i-h})(x_{k-j-h} - v_{n+1-j-h}) \quad (2.4.15)$$

for all permissible h .

Then from eqns. (2.4.14) and (2.4.15)

$$\epsilon\{\psi_{i,j}\} = \epsilon\{\psi_{i+h,j+h}\} = (N-n)\gamma_{i-j} \text{ for all permissible } h, i, j.$$

where γ_{i-j} is the $(i-j)^{\text{th}}$ covariance coefficient of $\{x_t\}$.

Similarly it can be shown that

$$\epsilon[(\underline{\Phi}'^T \underline{x}'_{n+1})_i] = (N-n)\gamma_i \text{ for } i = 1, \dots, n.$$

That is, all the elements of $\underline{\Psi}$ and $\underline{\Phi}'^T \underline{x}'_{n+1}$ can be approximated by covariance coefficients of $\{x_t\}$. Hence eqn. (2.4.9) can be approximated

by

$$\hat{\underline{g}} = \underline{P}_n^{-1} \underline{\rho}_n \quad (2.4.18)$$

where the common factor $(N-n)\gamma_0$ cancels from $\epsilon\{\underline{\Psi}\}$ and $\epsilon\{\underline{\Phi}'^T \underline{x}'_{n+1}\}$ to give

$$\underline{\rho}_n^T = (\rho_1 \quad \rho_2 \quad \cdot \quad \cdot \quad \cdot \quad \rho_n)$$

$$\underline{P}_n = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \rho_n \\ \rho_1 & 1 & \rho_1 & \cdot & \cdot & \cdot \\ \rho_2 & \rho_1 & 1 & \cdot & \cdot & \cdot \\ \cdot & & & & & \\ \cdot & & & & & \\ \rho_n & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

and where ρ_i is the i^{th} autocorrelation coefficient of $\{x_t\}$.

Durbin²⁹ has shown that eqn. (2.4.18) can be solved by the following iterative relations,

$$\hat{q}_{n,n} = \frac{\rho_n - \sum_{i=1}^{n-1} \hat{q}_{n-1,i} \rho_{n-i}}{1 - \sum_{j=1}^{n-1} \hat{q}_{n-1,j} \rho_j} \quad (2.4.19)$$

$$\hat{q}_{n,j} = \hat{q}_{n-1,j} - \hat{q}_{n,n} \hat{q}_{n-1,n-j} \text{ for } j = 1, \dots, n-1 \quad (2.4.20)$$

where $\hat{q}_{n,i}$ is an estimate of the i^{th} parameter of an n^{th} -order autoregressive model.

The equations (2.4.19) and (2.4.20) involve far less computation than the linear-least-squares solution of eqn. (2.4.9). The resulting rough estimates can therefore be of use in the choice of the model order.

Box & Jenkins⁵ introduce a statistic u_i defined by

$$u_i = \frac{\hat{q}_i}{\sigma_i} \quad (2.4.21)$$

where σ_i is the standard deviation of $\hat{q}_i$ and is estimated by

$$\sigma_i \approx \frac{\Omega}{\sqrt{N-i}}$$

Then if the estimates $\hat{q}_i$ of q_i are normally distributed, u_i is a unit normal random variable. Consider the null hypothesis that a particular parameter q_i is zero. Then if, for this value of i , $|u_i| > 1.96$ the null hypothesis can be rejected with probability 0.95. Conversely if $|u_i| < 1.96$ the null hypothesis is accepted. It is then possible to choose the optimum model order by successively calculating $\hat{q}_{11}, \hat{q}_{22}, \dots$ from eqns. (2.4.19) and (2.4.20) until a parameter $\hat{q}_{n+1,n+1}$ is found for which the null hypothesis is satisfied. The model order is then chosen to be n and the exact linear-least squares algorithm (2.4.9) is used to calculate the final estimates $\hat{q}_1, \hat{q}_2, \dots, \hat{q}_n$. Details of the

Fig.2.3. FLOW CHART FOR ESTIMATION OF AUTOREGRESSIVE MODELS

procedure are shown in the flow diagram of Fig. 2.3.

It now remains to test the adequacy of the estimated model and to calculate the estimation errors. Box & Jenkins propose several tests for 'goodness of fit', mostly based on an analysis of the residual sequence $\{a_t\}$. Since this sequence is an estimate of the white noise $\{\zeta_t\}$ it is sufficient to show that $\{a_t\}$ satisfies various criteria of statistical independence (e.g. zero autocorrelation).

Finally the covariance matrix of the estimates $\hat{\underline{g}}_1$ is given by

$$\begin{aligned} \epsilon\{(\hat{\underline{g}} - \underline{g})(\hat{\underline{g}} - \underline{g})^T\} &= \epsilon\{(\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T \underline{\zeta} \underline{\zeta}^T \underline{\Phi} (\underline{\Phi}^T \underline{\Phi})^{-1}\} \\ &= \lambda (\underline{\Phi}^T \underline{\Phi})^{-1} \end{aligned} \quad (2.4.22)$$

since $\epsilon\{\underline{\zeta} \underline{\zeta}^T\} = \lambda \underline{I}$

where $\underline{I}$ is the unit matrix.

2.5. The design of linear predictors

In the derivation and analysis of the Box & Jenkins control scheme it has been shown that the predictor $H(\underline{z}^{-1})$ plays a vital role in the performance of the overall system. This section will, in consequence, be concerned with the design of practical predictors based on a model for the random disturbance.

Firstly, consider a stationary random disturbance $\{x_t\}$ represented by an infinite-order moving-average model,

$$x_t = [1 + R(\underline{z}^{-1})] \zeta_t \quad (2.5.1)$$

$$\text{where } R(\underline{z}^{-1}) = \sum_{i=1}^{\infty} r_i \underline{z}^{-i}$$

It is desired to predict $x_{\tau+k}$ given $\{x_t, t = 1, 2, \dots, \tau\}$ such that the mean-square error is minimised; the error is defined by

$$e_{\tau}(k) = x_{\tau+k} - \hat{x}_{\tau}(k) \quad (2.5.2)$$

where $x_{\tau+k}$ and $\hat{x}_{\tau}(k)$ are respectively the true value and the predicted value of $\{x_t\}$ at time $(\tau+k)$. It has been shown by Kalman⁴ that the

random variable $\hat{x}_T(k)$ which minimizes the mean-squared error $\epsilon\{e_T^2(k)\}$ is the conditional expectation of x_{T+k} given $\{x_t, t = 1, 2, \dots, T\}$, i.e.

$$\hat{x}_T(k) = \epsilon\{x_{T+k} \mid x_1, x_2, \dots, x_T\} \quad (2.5.3)$$

Then from eqn. (2.5.1),

$$\hat{x}_T(k) = \epsilon\{\zeta_{T+k} + \sum_{i=1}^{\infty} r_i \zeta_{T+k-i} \mid x_1, \dots, x_T\} \quad (2.5.4)$$

But from the definition of $\{\zeta_t\}$,

$$\begin{aligned} \epsilon\{\zeta_{T+i} \mid x_1, \dots, x_T\} &= 0 \text{ if } i > 0 \\ &= \zeta_{T+i} \text{ if } i \leq 0 \end{aligned} \quad (2.5.5)$$

Eqn. (2.5.4) therefore reduces to

$$\hat{x}_T(k) = \sum_{i=k}^{\infty} r_i \zeta_{T+k-i} = R_T(z^{-1}) \zeta_{T+k} \quad (2.5.6)$$

where $[.]_T$ is the k^{th} -order truncation operator.

Substituting for ζ_{T+k} from eqn. (2.5.1).

$$\hat{x}_T(k) = \frac{R_T(z^{-1})}{1 + R(z^{-1})} x_{T+k} \quad (2.5.7)$$

$$\text{giving } H_T^*(z^{-1}) = \frac{R_T(z^{-1})}{1 + R(z^{-1})} \quad (2.5.8)$$

Then from eqns. (2.5.2) and (2.5.7) the error of prediction is given by

$$e_T(k) = [1 - H_T^*(z^{-1})] x_{T+k} = [1 + R(z^{-1}) - R_T(z^{-1})] \zeta_{T+k} \quad (2.5.9)$$

The preceding three equations apply equally to the mixed model of eqn. (2.4.1) since we can write

$$1 + R(z^{-1}) = [1 + S(z^{-1})] [1 - Q(z^{-1})]^{-1}$$

The optimum predictor for a mixed model can then be written, from eqn.

(2.5.8), as

$$H_T^*(z^{-1}) = \{[1 + S(z^{-1})] [1 - Q(z^{-1})]^{-1}\}_T \frac{[1 - Q(z^{-1})]}{[1 + S(z^{-1})]} \quad (2.5.10)$$

Now the predictor, as a linear discrete filter, is stable if all its poles lie inside the z -plane unit circle. The mixed model predictor of eqn. (2.5.10) will therefore be stable if the function $1 + S(z^{-1})$ is invertible.

It has been shown in section 2.4 that it is desirable to use autoregressive models to describe chemical process disturbances. Optimum predictors designed from autoregressive models also have some useful practical properties. Firstly it should be noted that a predictor for a pure autoregressive process, for which $S(z^{-1}) = 0$, is itself a moving-average filter and therefore always stable. Secondly, from Theorem A.3.1., an optimum predictor for an n^{th} -order autoregressive model has only n non-zero coefficients. Then if eqn. (2.5.10) is used to calculate $H_T^*(z^{-1})$ it is only necessary to compute the first $(n+k)$ coefficients of the inverse $(1-Q(z^{-1}))^{-1}$. Thirdly a k^{th} -step ahead predictor can be calculated from a simple iterative procedure given in Theorem A.3.2 - the procedure is started with $H_1^*(z^{-1}) = Q(z^{-1})$. These results are derived in detail in Appendix III.

Now consider the relationship,

$$\hat{x}_T(k) = H_T^*(z^{-1})x_{T+k} \quad (2.5.11)$$

The predictor $H_T^*(z^{-1})$ should satisfy the condition that $\hat{x}_T(k)$ is an unbiased estimate of x_{T+k} , or equivalently

$$\lim_{T \rightarrow \infty} \{\mu_e(\tau)\} = 0$$

where $\mu_e(\tau)$ is the ensemble average of the prediction error at time τ . By applying Theorem A.2 of Appendix II to eqn. (2.5.9) it follows that for a random process $\{x_t\}$ which may have a time-varying mean,

$$\lim_{T \rightarrow \infty} \{\mu_e(\tau)\} = [1 - H_T^*(z^{-1})]_{z=1} \mu_x(\tau)$$

The prediction $\hat{x}_T(k)$ is therefore an unbiased estimate of x_{T+k} if

$$H_T^*(z^{-1}) \Big|_{z=1} = 1 \quad (2.5.12)$$

The condition of eqn. (2.5.12) is identical to that imposed on the Box & Jenkins control system by Criterion 2.1. (see section 2.2). It is therefore of considerable practical importance to study the type of disturbance model which yields a predictor of unity steady-state gain.

Consider the integrated autoregressive process (see section 1.2) defined by

$$x_t = \frac{1}{(1-z^{-1})(1-Q(z^{-1}))} \zeta_t \quad (2.5.13)$$

and let

$$[1 + R(z^{-1})] = [1 - Q(z^{-1})]^{-1} \quad (2.5.14)$$

Then from eqn. (2.5.10),

$$\begin{aligned} H_T^*(z^{-1}) &= \{[1 + R(z^{-1})][1 - z^{-1}]^{-1}\}_T [(1 - z^{-1})(1 - Q(z^{-1}))] \\ &= \left[\sum_{i=0}^k r_i z^{-k} + \sum_{i=0}^{k+1} r_i z^{-k-1} + \dots \right] [(1 - z^{-1})(1 - Q(z^{-1}))] \\ &= \left[\sum_{i=0}^k r_i z^{-k} + r_{k+1} z^{-k-1} + \dots \right] [1 - Q(z^{-1})] \end{aligned}$$

$$\text{Hence } H_T^*(z^{-1})|_{z=1} = \left[1 + \sum_{i=1}^{\infty} r_i \right] \left[1 - \sum_{j=1}^n q_j \right]$$

and from eqn. (2.5.14),

$$H_T^*(z^{-1})|_{z=1} = 1$$

That is, an optimum predictor derived from an integrated autoregressive model always has unity steady-state gain. If however the disturbance $\{x_t\}$ has been assumed stationary then the autoregressive model will not, in general, include a difference term $(1 - z^{-1})$ and the optimum predictor will not, in general, satisfy eqn. (2.5.12). Under these circumstances it is necessary to design a sub-optimum predictor which will satisfy the steady-state performance criteria.

Consider a stationary random process $\{x_t\}$, and an estimated model

$$x_t = \frac{1}{1 - Q(z^{-1})} \zeta_t \quad (2.5.15)$$

Now define an adapted model to be used in the design of a sub-optimum predictor; in particular let

$$1 - Q'(z^{-1}) = [1 - Q(z^{-1})][1 - g(z^{-1})] \quad (2.5.16)$$

$$\text{where } g(z^{-1})|_{z=1} = 1$$

It is noted that since the function $1 - g(z^{-1})$ contains $(1 - z^{-1})$ as a factor, a predictor designed from the model of eqn. (2.5.16) will have unity steady-state gain.

A convenient form for $g(z^{-1})$ has been found to be an averaging function, viz

$$g(z^{-1}) = \frac{1}{N} \sum_{i=1}^N z^{-i} \quad (2.5.17)$$

From Theorem A.3.2 of Appendix III the k^{th} -step ahead predictor for the model of eqn. (2.5.16) is given by

$$H_k(z^{-1}) = \sum_{i=1}^{k-1} q'_i H_{k-i}(z^{-1}) z^{-i} + \sum_{i=k}^n q'_i z^{-i} \quad \text{for } k < n \quad (2.5.18)$$

and from eqn. (2.5.9),

$$e_r(k) = [1 - H_k(z^{-1})] x_{r+k} = \frac{[1 - H_k(z^{-1})]}{[1 - Q(z^{-1})]} \zeta_{r+k} \quad (2.5.19)$$

Then substituting eqn. (2.5.18) into eqn. (2.5.19) leads to

$$e_r(k) = [1 - g(z^{-1})] \zeta_{r+k} + \sum_{i=1}^{k-1} q'_i e_r(k-i) \quad (2.5.20)$$

For example, for one-step ahead prediction, $k = 1$ and

$$e_r(1) = [1 - g(z^{-1})] \zeta_{r+1} \quad (2.5.21)$$

$$\text{hence } \text{var} \{e_r(1)\} = \epsilon \{e_r^2(1)\} = \left\{ 1 + \sum_{i=1}^N \left(\frac{1}{N} \right)^2 \right\} \lambda$$

and since the optimal error variance is λ ,

$$\text{var } \{e_T(1)\} = \frac{N+1}{N} \text{var } \{\text{opt}(e_T(1))\}$$

Also from eqn. (2.5.20), for $k = 2$,

$$e_T(2) = [1 - g(z^{-1})] \zeta_{T+2} + q_1' e_T(1)$$

and substituting for $e_T(1)$ from eqn. (2.5.21) gives

$$e_T(2) = [1 - g(z^{-1})] [1 + q_1 z^{-1} + \frac{1}{N} z^{-1}] \zeta_{T+2}$$

If N is large, such that $q_1 \gg \frac{1}{N}$, then

$$e_T(2) \underline{\Omega} [1 - g(z^{-1})] [1 + q_1 z^{-1}] \zeta_{T+2} = [1 - g(z^{-1})] \text{opt } \{e_T(2)\}$$

$$\text{and } \text{var } \{e_T(2)\} \underline{\Omega} \frac{N+1}{N} \text{var } \{\text{opt } (e_T(2))\}$$

In general, for large N ,

$$\begin{aligned} e_T(k) &\underline{\Omega} [1 - g(z^{-1})] \text{opt } (e_T(k)) \\ \text{and } \text{var } \{e_T(k)\} &\underline{\Omega} \frac{N+1}{N} \text{var } \{\text{opt } (e_T(k))\} \end{aligned} \quad (2.5.22)$$

Thus, the introduction of the function $1 - g(z^{-1})$ increases the prediction error variance by a factor of approximately $(N+1)/N$, where N is large. In practice a value of N of the order of 20 appears suitable (see, for example, the simulation results presented in section 4.2).

2.6. Summary

This chapter has presented, in detail, a derivation and analysis of the Box & Jenkins control scheme, together with some results from linear estimation and prediction theory. It has been shown that the optimum control error is equal to the error of prediction, i.e.

$$\text{opt } \{e_t\} = [1 - H^*(z^{-1})] x_t$$

It has also been shown that there will be zero steady-state control error, under the conditions most commonly associated with industrial chemical

processes, if and only if

$$H(z^{-1})|_{z=1} = 1$$

This point has often been neglected in the literature and should be emphasised. Even if the system disturbance is a constant-mean stationary random signal there will still be a steady-state error after a set-point change if the steady-state gains of the process model and the process are unequal. Since a control engineer must always allow for the effects of estimation errors (and to some extent uncharacteristic behaviour) it is logical to require that the predictor should always have a steady-state gain of unity. It should be noted that the equivalent condition for the continuous-time case is that there must be integral control.

The use of autoregressive models for random disturbances has also been emphasised in the previous two sections. The most significant advantages of autoregressive models over mixed or moving-average models can be summarised as follows:-

- (i) a unique analytical solution of the estimation problem exists only when the model under consideration is purely autoregressive,
- (ii) an autoregressive model yields a stable predictor which is easily adapted to satisfy the steady-state criterion of unity steady-state gain,
- (iii) practical results (as presented in Chapters IV and V) show that for chemical process disturbances an autoregressive model gives at least as 'good a fit' as the more general mixed model.

CHAPTER IIIUPDATING THE BOX & JENKINS CONTROL SCHEMEIntroduction

It has been shown in Chapter I that there is a need for new methods of identifying process and disturbance dynamics. In particular the production requirements of many industrial chemical plants restrict the form of allowable experimentation to operation in a closed-loop. Identification then leads to an updating of the existing control system parameters - an optimisation procedure. This chapter will consider some methods of optimising a control system from analyses of closed-loop data.

The fundamental concept of this chapter can be stated as follows:- if the process is operated under the Box & Jenkins automatic feedback control system then the theoretical model for $\{X_t\}$ can be equated to a model estimated from a record of $\{X_t, t = 1, 2, \dots N\}$. The resulting 'updating equation' is developed in section 3.1.

Three methods of optimising the control system are suggested by the updating equation and these are derived in sections 3.2 to 3.4. Of particular practical significance is the separation effect which will be shown in section 3.5., to have a more general application to conventional control systems.

3.1. The updating equation

Consider the estimated disturbance signal $\{X_t\}$ in the Box & Jenkins control system. From eqns. (2.1.1), (2.1.13) and (2.1.14) it can be shown that

$$X_t = \frac{M(z^{-1})x_t + [z^{-k}P(z^{-1}) - z^{-1}M(z^{-1})]E_t^d}{M(z^{-1})[1-H(z^{-1})] + z^jP(z^{-1})H(z^{-1})} \quad (3.1.1)$$

where the notation is that of section 2.1. If the set-point $\{E_t^d\}$ is

is constant it is shown in Appendix IV that $\{X_t\}$ can be represented as a function of $\{x_t\}$ plus a constant level due to $\{E_t^d\}$. There is no loss of generality in the following if this level is considered to be zero.

Let the inverse process model be given by

$$[M(z^{-1})]^{-1} = W(z^{-1}) = \sum_{i=0}^{\infty} w_i z^{-i} \quad (3.1.2)$$

and let the disturbance $\{x_t\}$ be represented by an m^{th} -order autoregressive model,

$$x_t = \frac{1}{1 - Q(z^{-1})} \zeta_t, \quad Q(z^{-1}) = \sum_{i=1}^m q_i z^{-i} \quad (3.1.3)$$

Finally define a function $V(z^{-1})$ by

$$V(z^{-1}) = z^j P(z^{-1}) W(z^{-1}) - 1 = \sum_{i=j}^{\infty} v_i z^{-i} \quad (3.1.4)$$

Then eqn. (3.1.1) can be written in the form

$$X_t = \frac{1}{[1 - Q(z^{-1})][1 + H(z^{-1})V(z^{-1})]} \zeta_t \quad (3.1.5)$$

Now consider the denominator polynomial of eqn. (3.1.5) and let

$$1 - R(z^{-1}) = [1 - Q(z^{-1})][1 + H(z^{-1})V(z^{-1})] = 1 - \sum_{i=1}^{\infty} r_i z^{-i} \quad (3.1.6)$$

If both $P(z^{-1})$ and $W(z^{-1})$ are inherently stable filters, i.e. if

$$\sum_{i=0}^{\infty} |p_i| < \infty \quad \text{and} \quad \sum_{i=0}^{\infty} |w_i| < \infty$$

it follows that $\sum_{I=-j}^{\infty} |v_I| < \infty$

Also

$$H(z^{-1})V(z^{-1}) = \sum_{i=-j}^{\infty} \sum_{I=-j}^i v_I h_{i-I} z^{-1-i}$$

so that if $H(z^{-1})$ is an inherently stable filter,

$$\sum_{i=-j}^{\infty} \left| \sum_{I=-j}^{\infty} v_I h_{i-I} \right| < \infty$$

Then the polynomial $R(z^{-1})$ in eqn. (3.1.6) also satisfies the condition

$$\sum_{i=1}^{\infty} |r_i| < \infty$$

which implies that the series $r_1, r_2, \dots, r_i, \dots$ converges to zero as $i \rightarrow \infty$. In practice the polynomial $R(z^{-1})$ can be truncated after the n^{th} coefficient r_n such that for large n the relationship

$$R(z^{-1}) \approx \sum_{i=1}^n r_i z^{-i}$$

holds with an arbitrarily small error.

Now if the process is operated under the Box & Jenkins automatic feedback control system over the interval $0 \leq t \leq N$ a set of data $\{X_t, t = 1, 2, \dots, N\}$ can be recorded. This data can be used to estimate an autoregressive model for $\{X_t\}$ of the form,

$$X_t = \frac{1}{1 - \hat{R}(z^{-1})} \zeta_t, \quad \hat{R}(z^{-1}) = \sum_{i=1}^n \hat{r}_i z^{-i} \quad (3.1.7)$$

where, by comparison with eqn. (3.1.5),

$$1 - \hat{R}(z^{-1}) \approx 1 - R(z^{-1}) = [1 - Q(z^{-1})] [1 + H(z^{-1})V(z^{-1})] \quad (3.1.8)$$

This equation will be referred to as 'the updating equation'.

3.2. Updating the process model

By rearranging eqn. (3.1.8) we have

$$V(z^{-1}) = \frac{Q(z^{-1}) - R(z^{-1})}{H(z^{-1})[1 - Q(z^{-1})]} \quad (3.2.1)$$

and from the definition of $V(z^{-1})$ in eqn. (3.1.4) it follows that

$$z^j P(z^{-1}) = \frac{Q(z^{-1}) - R(z^{-1}) + H(z^{-1})[1 - Q(z^{-1})]}{H(z^{-1})[1 - Q(z^{-1})]} \cdot M(z^{-1}) \quad (3.2.2)$$

Now in eqn. (3.2.2) the polynomials $H(z^{-1})$, $M(z^{-1})$ are known since they determine the control system. Then if estimates $\hat{Q}(z^{-1})$, $\hat{R}(z^{-1})$ of $Q(z^{-1})$, $R(z^{-1})$ are available an estimate $\hat{P}(z^{-1})$ of $P(z^{-1})$ can be calculated from

$$z^j \hat{P}(z^{-1}) = \frac{\hat{Q}(z^{-1}) - \hat{R}(z^{-1}) + H(z^{-1}) [1 - \hat{Q}(z^{-1})]}{H(z^{-1}) [1 - \hat{Q}(z^{-1})]} \cdot M(z^{-1}) \quad (3.2.3)$$

where $j = 1 - k$ and l , the process model time-delay, is known.

The estimate $\hat{Q}(z^{-1})$ could be obtained from a record of the uncontrolled process output; i.e. if the process input $\{c_t\}$ is held constant the process output $\{E_t\}$ equals the disturbance $\{x_t\}$, such that $\hat{Q}(z^{-1})$ could be estimated from an analysis of $\{E_t\}$. This technique has previously been suggested by Florentin et al¹⁷. However, the method assumes that the disturbance model $Q(z^{-1})$ is the same for the set of uncontrolled-output data as for the set of closed-loop data. This assumption will, in general, only be valid if the disturbance is stationary and the number of data samples N is large. If $Q(z^{-1})$ does vary the resulting $\hat{P}(z^{-1})$ may be a highly inaccurate estimate of $P(z^{-1})$. It is also apparent that the method does not completely solve the simultaneous identification and control problem due to the necessity for a prior open-loop experimental run.

An alternative method does not require a priori knowledge of the disturbance model. If the control system is run twice with controllers $W_1(z^{-1})$ and $W_2(z^{-1})$ respectively, then two updating equations can be formed,

$$\left. \begin{aligned} 1 - \hat{R}_1(z^{-1}) &\underline{\Omega} [1 - Q(z^{-1})] [1 + H(z^{-1})V_1(z^{-1})] \\ 1 - \hat{R}_2(z^{-1}) &\underline{\Omega} [1 - Q(z^{-1})] [1 + H(z^{-1})V_2(z^{-1})] \end{aligned} \right\} \quad (3.2.4)$$

where the subscripts 1,2 refer to the first and second runs respectively. Once again it has been assumed that the disturbance model $Q(z^{-1})$ is the same for both sets of data. Now from eqn. (3.2.4)

$$\frac{1 - \hat{R}_1(z^{-1})}{1 - \hat{R}_2(z^{-1})} = \frac{1 + H(z^{-1})[z^j P(z^{-1}) W_1(z^{-1}) - 1]}{1 + H(z^{-1})[z^j P(z^{-1}) W_2(z^{-1}) - 1]}$$

and by rearranging

$$z^j \hat{P}(z^{-1}) = \frac{[\hat{R}_1(z^{-1}) - \hat{R}_2(z^{-1})][1 - H(z^{-1})]}{H(z^{-1})\{W_2(z^{-1})[1 - \hat{R}_1(z^{-1})] - W_1(z^{-1})[1 - \hat{R}_2(z^{-1})]\}} \quad (3.2.5)$$

where $\hat{P}(z^{-1})$ is an estimate of $P(z^{-1})$. Also from eqns. (3.2.4), (3.2.5)

$$1 - \hat{Q}(z^{-1}) = \frac{W_2(z^{-1})[1 - \hat{R}_1(z^{-1})] - W_1(z^{-1})[1 - \hat{R}_2(z^{-1})]}{[W_2(z^{-1}) - W_1(z^{-1})][1 - H(z^{-1})]} \quad (3.2.6)$$

where $\hat{Q}(z^{-1})$ is an estimate of $Q(z^{-1})$

There are two major disadvantages of this alternative method of updating the Box & Jenkins control scheme. Firstly, unless the controllers $W_1(z^{-1})$ and $W_2(z^{-1})$ are substantially different the polynomials $\hat{R}_1(z^{-1})$ and $\hat{R}_2(z^{-1})$ will be approximately equal and the coefficients of the numerator and denominator polynomials of eqns. (3.2.5), (3.2.6) will be close to zero. Then even for small estimation errors in $\hat{R}_1(z^{-1})$, $\hat{R}_2(z^{-1})$ the resulting estimates $\hat{P}(z^{-1})$, $\hat{Q}(z^{-1})$ may be subject to unacceptably large errors. Secondly, there is no way of checking the validity of the assumption that $Q(z^{-1})$ is constant so that the accuracy of the estimates cannot be determined.

It therefore appears that the techniques of this section are unreliable except for a very limited class of systems. Some results from simulations of such systems will be presented in section 4.3.

3.3. Updating the predictor

It was shown in section 2.1 that the signal $\{X_t\}$ is an estimate of the system disturbance $\{x_t\}$. Thus an estimate of the optimum predictor $\hat{H}^*(z^{-1})$ can be obtained from a model for $\{X_t\}$.

Now, consider the following predictor updating procedure : let the process be operated by the Box & Jenkins control system. After

every N sample intervals an autoregressive model for $\{X_t\}$ is estimated from a set of data $\{X_t, t = (i-1)N + 1, \dots, iN\}$, where $i = 1, 2, \dots$. Let the model at the i^{th} iteration be given by

$$X_t = \frac{1}{1 - \hat{R}^i(z^{-1})} \zeta_t, \quad \hat{R}^i(z^{-1}) = \sum_{j=1}^n \hat{r}_j^i z^{-j} \quad (3.3.1)$$

Then at every $(iN)^{\text{th}}$ sample for $i = 1, 2, \dots$ the predictor is updated such that

$$H^i(z^{-1}) = \{[1 - \hat{R}^i(z^{-1})]^{-1}\}_T [1 - \hat{R}^i(z^{-1})] \quad (3.3.2)$$

where $[\cdot]_T$ is the 1^{th} -order truncation operator.

It will now be shown that, under certain conditions, this predictor updating procedure converges to an optimum system. Consider the first updating (i.e. $i = 1$), then from eqn. (3.1.8)

$$1 - \hat{R}^1(z^{-1}) \stackrel{\Omega}{=} [1 - Q(z^{-1})] [1 + H(z^{-1})V(z^{-1})] \quad (3.3.3)$$

and if
$$1 + G(z^{-1}) = [1 - Q(z^{-1})]^{-1}, \quad G(z^{-1}) = \sum_{j=1}^{\infty} g_j z^{-j} \quad (3.3.4)$$

it follows from eqn. (3.3.2) that

$$H^1(z^{-1}) \stackrel{\Omega}{=} \frac{\{[1 + G(z^{-1})] [1 + H(z^{-1})V(z^{-1})]^{-1}\}_T}{[1 + G(z^{-1})] [1 + H(z^{-1})V(z^{-1})]^{-1}} \quad (3.3.5)$$

Consider the function,

$$H(z^{-1})V(z^{-1}) = H(z^{-1})[z^{1-k}P(z^{-1})W(z^{-1}) - 1]$$

If $k > 1$ the polynomial $V(z^{-1})$ contains no terms in positive powers of z . Also $H(z^{-1})$ contains only terms in z^{-1} for $i > 1$.

Then from Appendix I,

$$H(z^{-1}) = H_T(z^{-1})$$

$$\text{and} \quad H(z^{-1})V(z^{-1}) = H_T(z^{-1})V(z^{-1}) = [H_T(z^{-1})V(z^{-1})]_T$$

where $[\cdot]_T$ is the 1^{th} -order truncation operator.

Also from property (v) of Appendix I,

$$[(1+H_T(z^{-1})V(z^{-1}))^{-1} = 1 + \{ [1+H_T(z^{-1})V(z^{-1})]^{-1} \}_T$$

and it follows that

$$\begin{aligned} \{ [1+G(z^{-1})][1+H_T(z^{-1})V(z^{-1})]^{-1} \}_T &= \{ [1+G(z^{-1})] + [1+G(z^{-1})] \{ [1+H_T(z^{-1})V(z^{-1})]^{-1} \}_T \}_T \\ &= G_T(z^{-1}) + [1+G(z^{-1})] \{ [1+H_T(z^{-1})V(z^{-1})]^{-1} \}_T \end{aligned}$$

We can now write eqn. (3.3.5) in the form

$$H^1(z^{-1}) = \frac{G_T(z^{-1})[1+H_T(z^{-1})V(z^{-1})]}{[1+G(z^{-1})]} + \frac{\{ [1+H_T(z^{-1})V(z^{-1})]^{-1} \}_T}{[1+H_T(z^{-1})V(z^{-1})]^{-1}} \quad (3.3.6)$$

but from property (iv) of Appendix I,

$$\{ [1+H_T(z^{-1})V(z^{-1})]^{-1} \}_T = - H_T(z^{-1})V(z^{-1}) [1-H_T(z^{-1})V(z^{-1})]^{-1}$$

so that eqn. (3.3.7) reduces to

$$H^1(z^{-1}) = \frac{G_T(z^{-1})[1+H_T(z^{-1})V(z^{-1})]}{[1+G(z^{-1})]} - H_T(z^{-1})V(z^{-1}) \quad (3.3.7)$$

Assumption (a) : for any interval $(i-1)N+1 \leq t \leq iN$ the disturbance model is $[1-Q(z^{-1})]^{-1} = 1 + G(z^{-1})$. That is, $G^{i+1}(z^{-1}) = G^i(z^{-1}) = G(z^{-1})$ for all i .

Then eqn. (3.3.8) can be generalised to

$$H^i(z^{-1}) = \frac{G_T(z^{-1})[1+H_T^{i-1}(z^{-1})V(z^{-1})]}{[1+G(z^{-1})]} - H_T^{i-1}(z^{-1})V(z^{-1}) \quad (3.3.8)$$

Assumption (b) : the predictor-updating procedure converges to an equilibrium such that $H^i(z^{-1}) \rightarrow H^0(z^{-1})$ as $i \rightarrow \infty$.

Then at the equilibrium,

$$H_T^0(z^{-1}) = \frac{G_T(z^{-1})[1+H_T^0(z^{-1})V(z^{-1})]}{[1+G(z^{-1})]} - H_T^0(z^{-1})V(z^{-1}) \quad (3.3.9)$$

hence

$$H_T^O(z^{-1}) = \frac{G_T(z^{-1})}{[1+G(z^{-1})] + V(z^{-1})[1+G(z^{-1}) - G_T(z^{-1})]} \quad (3.3.10)$$

Now from eqn. (2.2.1) the system output at the equilibrium is given by

$$E_t = \frac{[1-H_T^O(z^{-1})][1+G(z^{-1})]}{[1+H_T^O(z^{-1})V(z^{-1})]} \zeta_t$$

and substituting for $H_T^O(z^{-1})$ from eqn. (3.3.10),

$$E_t = \frac{[1+G(z^{-1})][1+G(z^{-1}) - G_T(z^{-1})][1+V(z^{-1})]}{[1+G(z^{-1})][1+V(z^{-1})]} \zeta_t$$

hence

$$E_t = [1+G(z^{-1}) - G_T(z^{-1})] \zeta_t = [1-H_T^*(z^{-1})] x_t \quad (3.3.11)$$

which describes the optimum output error. That is, the predictor updating procedure, if convergent, produces an optimum system.

Convergence of the procedure can be established as follows.

At the i^{th} iteration let

$$H^i(z^{-1}) = H^O(z^{-1}) + h^i(z^{-1})$$

$$\text{where initially } H(z^{-1}) = H^O(z^{-1}) + h(z^{-1})$$

Then from eqn. (3.3.7),

$$H^O(z^{-1}) + h^i(z^{-1}) = \frac{G_T(z^{-1})[1+H^O(z^{-1})V(z^{-1})+h(z^{-1})V(z^{-1})]}{[1+G(z^{-1})]} - V(z^{-1})[H^O(z^{-1}) + h(z^{-1})]$$

and using eqn. (3.3.9)

$$h^i(z^{-1}) = h(z^{-1})V(z^{-1}) \left[\frac{G_T(z^{-1})}{1+G(z^{-1})} - 1 \right]$$

It follows that in general,

$$h^i(z^{-1}) = h(z^{-1})[V(z^{-1})]^i \left[\frac{G_T(z^{-1})}{1+G(z^{-1})} - 1 \right]^i \quad (3.3.12)$$

Then the updating procedure is convergent if

$$\lim_{i \rightarrow \infty} h^i(z^{-1}) = \lim_{i \rightarrow \infty} h(z^{-1}) [V(z^{-1})]^i \left[\frac{G_T(z^{-1})}{1 + G(z^{-1})} - 1 \right]^i = 0 \quad (3.3.13)$$

Now if the polynomial $\hat{R}^i(z^{-1})$ of the autoregressive model for $\{X_t\}$ has n coefficients then from Theorem A.3.2 of Appendix III, the predictor $H^i(z^{-1})$ also has only n non-zero coefficients. Hence at each iteration only the first n terms of the polynomial (3.3.12) influence the new predictor. It follows that convergence is determined by the first n terms of the polynomial (3.3.13).

Let

$$[V(z^{-1})] \left[\frac{G_T(z^{-1})}{1 + G(z^{-1})} - 1 \right] = 1 - d_0 + D(z^{-1}) \quad (3.3.14)$$

$$\text{where } D(z^{-1}) = \sum_{j=1}^{\infty} d_j z^{-j}$$

Then convergence is determined by the first n terms of the polynomial

$$\begin{aligned} [1 - d_0 + D(z^{-1})]^i &= (1 - d_0)^i + i(1 - d_0)^{i-1} D(z^{-1}) + \dots \\ &+ \frac{i(i-1) \dots (i-n+2)}{i!(n-1)} (1 - d_0)^{i-n+1} [D(z^{-1})]^{n-1} + \dots \end{aligned}$$

and for $i \gg n$ convergence is particularly dependent on $(1 - d_0)$. Hence a necessary condition for convergence of the procedure is that

$$|1 - d_0| < 1 \quad (3.3.15)$$

But from eqn. (3.3.14) and the definition of $V(z^{-1})$ in eqn. (3.1.9),

$$\begin{aligned} \text{for } j < 0, \quad d_0 &= 0 \\ \text{for } j = 0, \quad d_0 &= p_0 w_0 = p_0 / m_0 \end{aligned}$$

Thus the convergence condition (3.3.15) can only be satisfied if $j = 0$ and

$$|1 - p_o/m_o| < 1$$

$$\text{i.e. if } 0 < \frac{p_o}{m_o} < 2 \quad (3.3.16)$$

Eqn. (3.3.16) is a necessary but not sufficient condition for convergence of the procedure. Neither can the nature of the convergence be established; e.g. is it monotonically convergent?

These doubts concerning convergence added to the rather restrictive assumptions that l equals k and $Q(z^{-1})$ remains constant, do not recommend the procedure as a practically viable method of optimisation. The simulation results to be presented in section 4.3 do, however, show that predictor updating can be successfully applied to the proposed limited class of systems - in all these cases convergence is established after less than five iterations.

3.4. The separation effect

A special case of the updating equation leads to a simple solution to the problem of simultaneous identification and control. From eqn. (3.1.8)

$$1 - \hat{R}(z^{-1}) \underline{\Omega} [1 - Q(z^{-1})] + H(z^{-1}) [1 - Q(z^{-1})] [z^j P(z^{-1}) W(z^{-1})^{-1}] \quad (3.4.1)$$

$$\text{where } \hat{R}(z^{-1}) = \sum_{i=1}^n \hat{r}_i z^{-i} ; Q(z^{-1}) = \sum_{i=1}^m q_i z^{-i} ; H(z^{-1}) = \sum_{i=0}^{\infty} h_i z^{-1-i}$$

Consider the second term on the right-hand side of eqn. (3.4.1),

$$\begin{aligned} H(z^{-1}) [1 - Q(z^{-1})] [z^{1-k} P(z^{-1}) W(z^{-1})^{-1}] &= h_o p_o w_o z^{-k} + \dots & \text{for } k < 1 \\ &= h_o (p_o w_o - 1) z^{-1} + \dots & \text{for } k = 1 \\ &= -h_o z^{-1} + \dots & \text{for } k > 1 \end{aligned} \quad (3.4.2)$$

then the lowest power of z in this expression is $z^{-\beta}$ where

$$\beta = \min(1, k) \quad (3.4.3)$$

Thus if $m < \beta$ the two polynomials on the right-hand side of eqn. (3.4.1) contain no common powers of z and we can write

$$1 - \sum_{i=1}^m \hat{r}_i z^{-i} \underline{\Omega} 1 - \sum_{i=1}^m q_i z^{-i} \quad (3.4.4)$$

$$\text{and } -\hat{R}_T(z^{-1}) \underline{\Omega} H(z^{-1})[1-Q(z^{-1})][z^j P(z^{-1})W(z^{-1})-1] \quad (3.4.5)$$

where $[\cdot]_T$ is the β th-order truncation operator. The special case where $m < \beta$, which allows the updating equation (3.4.1) to be split into the two equations (3.4.4), (3.4.5), will be referred to as the separation effect.

The m disturbance model parameters can now be estimated from eqn. (3.4.4), viz

$$\hat{q}_i = \hat{r}_i \quad \text{for } i = 1, 2, \dots, m \quad (3.4.6)$$

and substituting $\hat{Q}(z^{-1})$ for $Q(z^{-1})$ in eqn. (3.4.5),

$$-\hat{R}_T(z^{-1}) \underline{\Omega} H(z^{-1}) \left[1 - \sum_{i=1}^m \hat{r}_i z^{-i} \right] [z^j P(z^{-1})W(z^{-1})-1]$$

and rearranging,

$$z^j \hat{P}(z^{-1}) = \frac{H(z^{-1}) \left[1 - \sum_{i=1}^m \hat{r}_i z^{-i} \right] - \hat{R}_T(z^{-1})}{H(z^{-1}) \left[1 - \sum_{i=1}^m \hat{r}_i z^{-i} \right] W(z^{-1})} \quad (3.4.7)$$

Since $H(z^{-1})$ and $W(z^{-1})$ are known and the $\hat{r}_i$, $i = 1, \dots, n$ can be calculated from a set of data $\{X_t, t = 1, 2, \dots, N\}$, it is possible to compute the estimate $\hat{P}(z^{-1})$ by manipulation of the z -transforms in eqn. (3.4.7). Alternatively the estimate $\hat{P}(z^{-1})$ can be computed by equating coefficients in the expression,

$$z^{1-k} \sum_{i=0}^{n-\beta} \hat{p}_i z^{-i} \sum_{i=0}^{\infty} h_i z^{-1-i} \left[1 - \sum_{i=1}^m \hat{r}_i z^{-i} \right] \sum_{i=0}^{\infty} w_i z^{-i} = \sum_{i=0}^{\infty} h_i z^{-1-i} \left[1 - \sum_{i=1}^m \hat{r}_i z^{-i} \right] - \sum_{i=\beta}^n \hat{r}_i z^{-i} \quad (3.4.8)$$

Fig.3.1. FLOW CHART FOR ESTIMATION OF
AUTORESSIVE MODELS USED FOR UPDATING

where it should be noted that since only n coefficients $\hat{r}_i$ are available it is only possible to estimate the first $(n-\beta+1)$ coefficients of $P(z^{-1})$. The control system is updated by calculating an optimum predictor from the autoregressive disturbance model $\hat{Q}(z^{-1})$ and a controller from the inverse of the process $\hat{P}(z^{-1})$.

It has been noted that the separation technique is only applicable to a system where the number of parameters in the disturbance model is less than the number of samples in the process (or process model) time-delay. The extent of this limitation can be judged from the fact that most chemical process disturbances can be represented by autoregressive models of not more than 3 parameters. Many examples of processes with sufficiently large time-delays exist in the chemical industry, notably cement and paper-making, plastics, polymerisation and distillation processes. Also in some cases it may be possible to satisfy the requirement that $m < \beta$ by increasing the system sampling frequency; this has the effect of increasing l and k while m , in general, remains constant.

Detailed application of the separation technique is straightforward and proceeds as follows:

- (i) from a set of data $\{X_t, t = 1, 2, \dots, N\}$ the autoregressive parameter estimates $\hat{r}_i$ of the model eqn. (3.1.12), are computed by the linear-least-squares algorithm of eqn. (2.4.9). The procedure for choosing the model order n is similar to that presented in section 2.4. except that the iteration continues until several successive parameters $\hat{r}_{i,i}$ are found to satisfy the null hypothesis. The flow diagram of Fig. 3.1 shows the detailed procedure for choosing n . Finally the covariance matrix of the estimates is calculated from eqn. (2.4.22)
- (ii) A statistic α_i is defined by

$$\alpha_i = \frac{\hat{r}_i}{\sqrt{\text{var}(\hat{r}_i)}} \quad (3.4.9)$$

where $\text{var}(\hat{r}_i)$ is found from the covariance matrix of the estimates. Then α_i is a unit normal random variable under the null hypothesis that the parameter r_i differs insignificantly from zero. Hence any insignificant parameters can be eliminated from the polynomial $\hat{R}(z^{-1})$ by the following:-

$$\begin{aligned} \text{for } i = 1, 2, \dots, n, \text{ if } |\alpha_i| < 1.96, \hat{r}_i &\approx 0 \\ \text{if } |\alpha_i| > 1.96, \hat{r}_i &\neq 0 \end{aligned}$$

(iii) The reduced polynomial $\hat{R}(z^{-1})$ can now be examined for evidence of the separation effect. Consider the polynomial

$$\hat{R}(z^{-1}) = \hat{r}_1 z^{-1} + \hat{r}_2 z^{-2} + \dots + \hat{r}_n z^{-n}$$

and let $|\alpha_i| < 1.96$ for $i = m+1, \dots, \beta-1$ where $\beta > m$. Then $\hat{R}(z^{-1})$ reduces to

$$\hat{R}(z^{-1}) = \hat{r}_1 z^{-1} + \dots + \hat{r}_m z^{-m} + \hat{r}_\beta z^{-\beta} + \dots + \hat{r}_n z^{-n}$$

since $\hat{r}_i = 0$ for $i = m+1, \dots, \beta-1$

In this case separation clearly occurs after the m^{th} sample and it follows that the disturbance model $Q(z^{-1})$ contains m coefficients and that $\beta = \min(1, k)$. Also

$$\hat{R}_T(z^{-1}) = \sum_{i=\beta}^n \hat{r}_i z^{-i}$$

(iv) The process pure time-delay is readily determined from an examination of the particular structure of eqn. (3.4.8). Since the process-model time-delay l is known, k is determined as follows,

a) if $\beta < l$ then, from eqn. (3.4.8), k must equal β .

b) if $\beta = l$ then $k > l$. The precise value of k is best found by solving for $z^j P(z^{-1})$ in eqns. (3.4.7) or (3.4.8).

It should be noted that the reduction procedure of (iii) above may estimate β to be greater than $\min(1, k)$. In this situation $k > l$

and, as before, the precise value of k is found by solving for $\hat{z}^j \hat{p}(z^{-1})$.

This experimental procedure can easily be adapted to meet the needs of a particular example. The simulation results of section 4.3 and the industrial investigation of section 5.4 show various similar methods that have been successfully employed.

In practice it is always important to be able to calculate the errors of estimation. For the disturbance model parameters this is straightforward : from eqn. (3.4.6)

$$\epsilon\{\hat{q}_i\} = \epsilon\{\hat{r}_i\} = r_i - q_i \quad \text{for } i = 1, 2, \dots, m$$

$$\text{and var}\{\hat{q}_i\} = \epsilon\{\hat{q}_i - q_i\}^2 = \epsilon\{\hat{r}_i - r_i\}^2 = \text{var}\{\hat{r}_i\}$$

so that $\hat{q}_i$ is an unbiased estimate of q_i with a variance easily found from the major diagonal of the covariance matrix $(\hat{\underline{r}} - \underline{r})(\hat{\underline{r}} - \underline{r})^T$. Now consider the estimation errors of the process parameters p_i ; let

$$\hat{R}(z^{-1}) = R(z^{-1}) + \Phi(z^{-1}) \quad (3.4.10)$$

$$\text{where} \quad \Phi(z^{-1}) = \sum_{i=1}^{\infty} \phi_i z^{-i}$$

and $\Phi(z^{-1})$ is the difference between the estimated and true polynomials $R(z^{-1})$.

Then from eqn. (3.4.6),

$$\begin{aligned} \hat{Q}(z^{-1}) &= \hat{R}(z^{-1}) - \hat{R}_T(z^{-1}) \\ &= R(z^{-1}) - R_T(z^{-1}) + \Phi(z^{-1}) - \Phi_T(z^{-1}) \end{aligned}$$

and substituting for $\hat{Q}$ and $\hat{R}_T$ in eqn. (3.4.7), and omitting the z^{-1} argument,

$$\hat{z}^j \hat{p} = \frac{H[1 - R + R_T - \Phi + \Phi_T] - R_T - \Phi}{HW[1 - R + R_T - \Phi + \Phi_T]} \quad (3.4.11)$$

But the true polynomial $z^j P(z^{-1})$ is given by

$$z^j P = \frac{H[1-Q] - R_T}{HW[1-Q]} = \frac{H[1-R+R_T] - R_T}{HW[1-R+R_T]}$$

so that eqn. (3.4.11) can be written in the form

$$z^j \hat{P} - z^j P = \frac{\hat{R}_T}{HW} \left[\frac{1}{1-Q} - \frac{1}{1-\hat{Q}} \right] - \frac{\hat{\Phi}_T}{HW(1-Q)} \quad (3.4.12)$$

It follows that

$$\hat{p}_0 - p_0 = - \frac{\phi_\beta}{h_0 w_0} \quad (3.4.13)$$

$$\hat{p}_1 - p_1 = - \frac{\hat{r}_\beta \phi_1}{h_0 w_0} - \frac{\phi_{\beta+1}}{h_0 w_0} - \frac{\phi_\beta}{h_0 w_0} \left[\frac{h_1}{h_0} + \frac{w_1}{w_0} + q_1 \right] \quad (3.4.14)$$

Now, since $\epsilon\{\phi_i\} = 0$ for $i = 1, 2, \dots, n$, from eqn. (3.4.13)

$$\epsilon\{\hat{p}_0 - p_0\} = 0$$

That is, $\hat{p}_0$ is an unbiased estimate of p_0 . Unfortunately a similar result does not apply for $\hat{p}_i$, $i = 1, \dots, n - \beta$. For instance, from eqn. (3.4.14)

$$\epsilon\{\hat{p}_1 - p_1\} = \epsilon\left\{-\frac{(r_\beta + \phi_\beta)\phi_1}{h_0 w_0}\right\} = \epsilon\left\{-\frac{\phi_\beta \phi_1}{h_0 w_0}\right\} \neq 0 \quad (3.4.15)$$

The bias can, however, be calculated since h_0 and w_0 are known and $\epsilon\{\phi_\beta \phi_1\}$ is the term $\epsilon\{(\hat{r}_\beta - r_\beta)(\hat{r}_1 - r_1)\}$ in the covariance matrix of the estimates.

The variance of the estimates can be calculated as follows,

$$\begin{aligned} \text{var}\{\hat{p}_0\} &= \epsilon\{\hat{p}_0 - p_0\}^2 = \left(\frac{1}{h_0 w_0}\right)^2 \epsilon\{\phi_\beta^2\} \\ &= \frac{\text{var}\{\hat{r}_\beta\}}{(h_0 w_0)^2} \end{aligned} \quad (3.4.16)$$

where $\text{var}\{\hat{r}_\beta\}$ is the element on the β^{th} column and row of the covariance

matrix of the $\hat{r}$'s. Also, from eqn. (3.4.14)

$$\text{var}\{\hat{p}_1\} = \left(\frac{1}{h_o w_o} \right)^2 \epsilon \{ \phi_1^2 r_\beta^2 + \phi_{\beta+1}^2 + \gamma^2 \phi_\beta^2 + \phi_1 \phi_{\beta+1} r_\beta + \gamma \phi_\beta^2 + \gamma \phi_\beta \phi_{\beta+1} \} \quad (3.4.17)$$

$$\text{where } \gamma = \frac{h_1}{h_o} + \frac{w_1}{w_o} + q_1$$

An approximate $\text{var}\{\hat{p}_1\}$ can then be calculated by substituting $\hat{r}_\beta$, $\hat{q}_1$ for r_β , q_1 . Similar but more complex expressions hold for $\text{var}\{\hat{p}_i\}$, $i = 2, \dots, n - \beta$.

If the estimation errors are normally distributed (but not necessarily with a zero mean) we can write

$$\left. \begin{aligned} | \hat{p}_o - p_o | &\leq 2\sqrt{(\text{var}\{\hat{p}_o\})} \\ \text{or } | \hat{p}_o - p_o | &\leq \frac{2\sigma_\beta}{h_o w_o} \end{aligned} \right\} \text{ with probability 0.95}$$

where σ_β is the standard deviation of $\hat{r}_\beta$. Similarly the error in $\hat{p}_1$ lies in the range

$$- \frac{\epsilon\{\phi_\beta \phi_1\}}{h_o w_o} - \sqrt{(\text{var}\{\hat{p}_1\})} > \hat{p}_1 - p_1 > - \frac{\epsilon\{\phi_\beta \phi_1\}}{h_o w_o} + \sqrt{(\text{var}\{\hat{p}_1\})}$$

with probability 0.95. It should be noted that $\epsilon\{\phi_\beta \phi_1\}$ and $\text{var}\{\hat{p}_1\}$ are of the same order of magnitude - that is, very much less than 1. Then since the 95% confidence limit is determined by $\sqrt{(\text{var}\{\hat{p}_1\})}$ it follows that the bias is typically an order of magnitude smaller. Since similar, but more complex, expressions hold for the $\hat{p}_i$, $i = 2, \dots, n - \beta$ it appears that the bias can be reasonably assumed insignificant compared with the estimation error, for all the estimates $\hat{p}_i$.

It should also be noted that the estimation errors are highly dependent on the magnitude of $h_o w_o$. It is therefore desirable to choose h_o and $w_o = (m_o)^{-1}$ as large as possible in the control system to be used for identification.

Fig.3.2. CONVENTIONAL CONTROL SYSTEM

3.5. The separation effect as a means of identification

In the previous section it has been shown that the separation effect leads to a simple solution to the simultaneous identification and control problem for processes with large time-delays. The technique has so far been limited to the Box & Jenkins control system. It will now be shown that similar results hold for a conventional control system.

Consider the system of Fig. 3.2, where $G(z^{-1})$ describes the controller pulse transfer function. If $\{E_t^d, t = 1, 2, \dots, N\} = 0$ the output is given by

$$E_t = \frac{1}{1 + z^{-k}P(z^{-1})G(z^{-1})} x_t = \frac{1}{[1 - Q(z^{-1})][1 + z^{-k}P(z^{-1})G(z^{-1})]} \zeta_t \quad (3.5.1)$$

Then an autoregressive model can be estimated from a set of data

$\{E_t, t = 1, 2, \dots, N\}$, i.e.

$$E_t = \frac{1}{1 - \hat{R}(z^{-1})} \zeta_t, \quad \hat{R}(z^{-1}) = \sum_{i=1}^n \hat{r}_i z^{-i} \quad (3.5.2)$$

and by equating (3.5.1), (3.5.2)

$$1 - \hat{R}(z^{-1}) \underline{\Omega} [1 - Q(z^{-1})][1 + z^{-k}P(z^{-1})G(z^{-1})] \quad (3.5.3)$$

$$\text{where } Q(z^{-1}) = \sum_{i=1}^m q_i z^{-i}$$

Since $G(z^{-1})$ must contain only negative powers of z , it follows that if $m < k$, eqn. (3.5.3) can be separated into two equations,

$$\hat{q}_i = \hat{r}_i \text{ for } i = 1, 2, \dots, m \quad (3.5.4)$$

$$\text{and } z^{-k}\hat{P}(z^{-1}) = \frac{\hat{Q}(z^{-1}) - \hat{R}(z^{-1})}{G(z^{-1})[1 - \hat{Q}(z^{-1})]} = \frac{-\hat{R}_T(z^{-1})}{G(z^{-1})[1 - \hat{Q}(z^{-1})]} \quad (3.5.5)$$

where $[\cdot]_T$ is the k^{th} -order truncation operator.

It is now possible to design an optimum Box & Jenkins control system - i.e. the predictor $\hat{H}(z^{-1})$ is calculated from the autoregressive disturbance model (3.5.4) and the controller $W(z^{-1})$ from the inverse of $\hat{P}(z^{-1})$.

If the separation effect is to be used as a method of identification an obvious requirement is that it should be possible to design some form of automatic control system prior to the identification. In most cases the process will have been under manual control for some time before the implementation of an automatic control system. During this time the process operators will probably have developed empirical relationships for the most important control parameters - i.e. the steady-state gain and time-delay of the process. An empirical control system can then be designed from the experience of manual control. Even if no quantitative information is available it is usually possible to make approximations to the process steady-state gain and time-delay from experimental step responses.

The separation technique of identification is fundamentally different from the established methods of Box & Jenkins, Astrom etc. in that it is specifically designed for use with an automatic control system rather than with open-loop operating data. The most significant advantage that feedback control can have over open-loop operation is the elimination of drifts or trends in the output. Closed-loop operation, therefore, is less likely to cause a serious upset to normal production and in consequence should be more viable from the economic point of view.

An important practical feature of the separation technique is that no assumption has to be made concerning the stationarity of the disturbance. In fact, the method appears to be equally valid for a wide range of non-stationary disturbances. Even if the disturbance has time-varying characteristics it should still be possible to estimate the process impulse response, although the resulting disturbance model will be of little value in the design of an optimum predictor.

3.6. Summary and Conclusions

This chapter has presented some theoretical techniques of updating the Box & Jenkins control scheme. The methods, which are all based on analysis of the signal $\{X_t\}$, can be summarised as follows:-

- a) if the model $Q(z^{-1})$ of the disturbance $\{x_t\}$ is known a priori, and if a sufficiently long sample of data is available such that $Q(z^{-1})$ remains constant over successive records, then an estimate of the process impulse response can be calculated and the controller updated;
- b) if $Q(z^{-1})$ remains constant over successive records and if the control system can be operated with two substantially different controllers $W_1(z^{-1})$ and $W_2(z^{-1})$, then estimates of the process impulse response and $Q(z^{-1})$ can be calculated and the controller and predictor updated;

- c) if $Q(z^{-1})$ remains constant over successive records and if

$$(i) \quad l = k$$

$$(ii) \quad 0 < p_0 w_0 < 2$$

then the control system can be optimised by updating the predictor every N samples;

- d) if the number of parameters m in the disturbance model is less than both l and k then estimates of the process impulse response and the disturbance model can be calculated and both the controller and predictor updated.

Methods a), b), c) are thus only applicable to a limited class of systems, the assumptions implying that long samples of $\{X_t\}$ are available and that the disturbance is known a priori to be stationary. The validity of these assumptions is difficult to establish in a practical example - a feature not present in the separation technique

- d). If a process has a large time-delay the separation technique offers a straightforward method of simultaneous identification and con-

trol. It has a significant advantage over the more established techniques in that the process input perturbation is supplied by the inherent process disturbances rather than from an external source. It also requires a relatively small amount of computation and produces a unique analytic solution - not always the case for maximum likelihood methods.

It should be noted that the techniques of this chapter can only estimate the impulse response of the process. Although this is not considered to be a major practical disadvantage an attempt will be made in Chapter VI to generalise the methods to pulse-transfer function representations. Also it has so far been tacitly assumed that the process dynamics are time-invariant. However the methods are applicable to processes with slowly-varying dynamics - e.g. if a significant change only occurs over a large interval N (of the order of 200 samples) then it can be detected by updating every N samples.

Finally, although it has been shown that the separation technique applies equally to a conventional feedback control system it appears better suited to the Box & Jenkins configuration. The reason is the particular form of the autoregressive model for $\{X_t\}$ - i.e. the model $R(z^{-1})$ for $\{X_t\}$ can, in general, be adequately approximated by a finite-order polynomial $\hat{R}(z^{-1})$. This may not be the case with a conventional control system since the controller $G(z^{-1})$ may contain an integral term $(1-z^{-1})^{-1}$.

CHAPTER IV

SIMULATION STUDIES

Introduction

This chapter presents a series of results obtained in order to substantiate the theories of Chapters II and III and as an aid to the industrial investigation to be presented in Chapter V. The first section is concerned with the estimation of the parameters of an autoregressive model as a prelude to the second section on prediction and section 3 which deals with updating the Box & Jenkins control scheme.

The results of this chapter were obtained from a set of digital computer programs written in FORTRAN. As the programs required no special knowledge they are not included in this thesis. Also it should be noted that only a representative sample of results will be presented here.

Throughout the simulation studies it was found convenient to employ artificially-generated random signals - i.e. signals of known characteristics formed by passing a sequence of independent random numbers through a digital filter. They provide a simple means of comparing practical results with the corresponding theoretical values.

4.1. Estimation of the parameters of an autoregressive model.

Throughout the first 3 chapters the practical importance of autoregressive models for chemical process disturbances has been emphasised. This section will first compare two methods of autoregressive parameter estimation and then present some results which appear to justify the use of autoregressive models in practice.

In section 2.4 the exact method of linear-least squares and the approximate partial-autocorrelation method due to Durbin were proposed for autoregressive parameter estimation. Table 4.1 compares

TABLE 4.1. AUTOREGRESSIVE PARAMETER ESTIMATION

SERIES I : 200 samples

Parameter	Partial-auto- correlation estimate	Linear-least- squares estimate	True
q_1	0.52	0.58 ± 0.14	0.60
q_2	-0.34	-0.43 ± 0.15	-0.40
q_3	0.24	0.24 ± 0.14	0.20
λ	10542	9212	9107

SERIES II : 200 samples

q_1	0.54	0.58 ± 0.14	0.60
q_2	-0.33	-0.34 ± 0.15	-0.36
q_3	0.19	0.23 ± 0.14	0.22
λ	10601	9278	9107

SERIES III : 200 samples

q_1	1.70	1.73 ± 0.08	1.75
q_2	-0.80	-0.82 ± 0.08	-0.82
λ	1.24	1.12	1.03

SERIES IV : 200 samples

q_1	1.67	1.71 ± 0.08	1.70
q_2	-0.75	-0.78 ± 0.08	-0.80
λ	9.45	8.91	8.87

the estimates obtained for 4 random sequences generated by

$$\text{Series I} \quad ; x_t = \frac{1}{1-0.6z^{-1}+0.4z^{-2}-0.2z^{-3}} \zeta_t \quad ; \text{var}\{\zeta_t\} = \lambda = 9107 \quad (4.1.1)$$

$$\text{Series II} \quad ; x_t = (1+0.6z^{-1}) \zeta_t \quad ; \text{var}\{\zeta_t\} = \lambda = 9107 \quad (4.1.2)$$

$$\text{Series III} \quad ; x_t = \frac{1}{1-1.75z^{-1}+0.82z^{-2}} \zeta_t \quad ; \text{var}\{\zeta_t\} = \lambda = 1.03 \quad (4.1.3)$$

$$\text{Series IV} \quad ; x_t = \frac{1}{1-1.7z^{-1}+0.82z^{-2}} \zeta_t \quad ; \text{var}\{\zeta_t\} = \lambda = 8.87 \quad (4.1.4)$$

The sequence $\{\zeta_t\}$ was tested for randomness by calculating the autocorrelation function; it was found that the autocorrelation coefficients were insignificantly small for large samples of $\{\zeta_t\}$, e.g. < 0.08 for 200 samples. However it was noted that the coefficients differed significantly from record to record, e.g. $\epsilon\{\zeta_t \zeta_{t-1}\} = -0.04$ for $t = 1, 2, \dots, 200$ whereas $\epsilon\{\zeta_t \zeta_{t-1}\} = +0.07$ for $t = 601, \dots, 800$. As expected these variations were more marked in smaller samples. In conclusion it was thought that the sequence $\{\zeta_t\}$ could only be considered to be white noise for samples greater than about 200.

The estimates in Table 4.1 were calculated assuming a model of the form

$$x_t = \frac{1}{1 - \hat{Q}(z^{-1})} a_t \quad ; \quad \hat{Q}(z^{-1}) = \sum_{i=1}^n \hat{q}_i z^{-i}$$

where $\hat{q}_i, i = 1, \dots, n$ are estimates of the true parameters $q_i, i = 1, \dots, n$ and where $\{a_t\}$ is the sequence of residuals whose variance $\hat{\lambda}$ is an estimate of λ . The model order n was selected by the procedure shown in the flow diagram of Fig. 2.3. The 95% confidence limits for the linear-least-squares estimates are also shown in Table 4.1.

It can be clearly seen, from Table 4.1, that the linear-least-

TABLE 4.2. MODELS FOR BOX & JENKINS DATA

Data Series	Estimated model (Box & Jenkins)	Residual Variance
Concentration	$x_t = \frac{1-0.5z^{-1}}{1-0.9z^{-1}} a_t$	0.102
Temperature	$x_t = \frac{1-0.1z^{-1}-0.1z^{-2}}{(1-z^{-1})^2} a_t$	0.199
Viscosity	$x_t = \frac{1}{1-0.86z^{-1}} a_t$	0.089

Data Series	Estimated autoregressive model	Residual Variance
Concentration	$x_t = \frac{1}{1-0.42z^{-1}-0.21z^{-2}} a_t$	0.114
Temperature	$x_t = \frac{1}{(1-0.80z^{-1})(1-z^{-1})} a_t$	0.183
Viscosity	$x_t = \frac{1}{1-0.86z^{-1}} a_t$	0.089

squares method produced more reliable results. This was apparent in all the tested examples although the partial-autocorrelation estimates were found to be almost identical to the linear-least-squares estimates for large samples - e.g. for $N = 500$. As a consequence of these results the linear-least-squares technique was subsequently used exclusively for autoregressive parameter estimation.

Of particular interest is a comparison of the models for Series I and Series II. The pure moving-average process of eqn(4.1.2) was chosen to be a first-order approximation to the autoregressive process of eqn(4.1.1). Table 4.1 shows that the estimated autoregressive models for the two series are both third-order with very similar parameter values.

The second set of results was obtained from data quoted by Box & Jenkins⁵. The data consisted of sets of measurements of concentration, temperature and viscosity from an industrial chemical process. Box & Jenkins attempted to fit 'parsimonious' models to the data - i.e. models with the least number of parameters. Table 4.2 shows the parameter estimates derived by Box & Jenkins as compared with the parameter estimates for a pure autoregressive model. It can be seen that in no case does the autoregressive model contain more parameters than the corresponding 'parsimonious' model of Box & Jenkins. Also there is little difference between the residual variances of the mixed and autoregressive models.

A further set of data, collected from an industrial process, was available for analysis - 150 uncontrolled basis weight measurements from a paper machine. In this case the estimated autoregressive model was given by

$$x_t = \frac{1}{1 - 0.89z^{-1} + 0.19z^{-2}} a_t$$

TABLE 4.3. COMPARISON OF PREDICTORS

Series III : var x_t = 39.8 : 200 samples

Autoregressive Model $1-Q(z^{-1})$	Prediction Step k	Predictor $z^k H(z^{-1})$	Calculated Error Variance	Optimal Error Variance
$1-1.73z^{-1}+0.82z^{-2}$	1	$1.73 + 0.82z^{-1}$	1.17	1.12
"	2	$2.17 - 1.42z^{-1}$	4.50	4.06
"	4	$2.27 - 1.86z^{-1}$	16.44	15.23
$(1-0.8z^{-1})(1-z^{-1})$	1	$1.82 - 0.8z^{-1}$	1.29	1.12
"	2	$2.44 - 1.44z^{-1}$	5.11	4.06
"	4	$3.38 - 2.28z^{-1}$	20.56	15.23

Sub-optimal prediction : 200 samples

Autoregressive Model $1-Q(z^{-1})$	Averaging Function $g(z^{-1})$	Prediction Step k	Calculated Error Variance	Sub-optimal Error Variance
$1-1.73z^{-1}+0.82z^{-2}$	$\sum_{i=1}^{10} 0.1z^{-i}$	1	1.30	1.23
"	"	2	4.91	4.47
"	"	4	19.18	16.75
"	$\sum_{i=1}^{20} 0.05z^{-i}$	1	1.27	1.18
"	"	2	4.80	4.26
"	"	4	18.32	15.99

Thus the estimated autoregressive models for the four available industrial process disturbances all contained a maximum of two parameters. This evidence supports the hypothesis of section 2.4 (and section 3.4) that, in general, an industrial process disturbance can be adequately represented by an autoregressive model of not more than three parameters.

4.2. Prediction

This section compares the three forms of predictor proposed in section 2.5. Firstly, consider the optimum k-step ahead predictor for an autoregressive process $\{x_t\}$ defined by

$$\hat{H}(z^{-1}) = \{[1-Q(z^{-1})]^{-1}\}_T [1-Q(z^{-1})] \quad (4.2.1)$$

where $[\cdot]_T$ is the k^{th} -order truncation operator. Then from eqn.(2.5.9) the optimum error variance is given by

$$\text{opt} [\text{var} \{e_T(k)\}] = \left[1 + \sum_{i=1}^{k-1} S_i^2 \right] \lambda \quad (4.2.2)$$

$$\text{where } 1 + S(z^{-1}) = \sum_{i=1}^{\infty} S_i z^{-i} = [1-Q(z^{-1})]^{-1}$$

In practice the predictor would be derived from an estimate $1 - \hat{Q}(z^{-1})$ of the disturbance model and the error variance calculated from

$$\text{var} \{e_T(k)\} = \frac{1}{N} \sum_{\tau=1}^N [x_{\tau+k} - \hat{x}_{\tau}^-(k)]^2 \quad (4.2.3)$$

Examples of predictors and the calculated error variances for 200 samples of Series III are shown in the first 3 rows of Table 4.3. It can be seen that, in each case, the calculated error variance is close to the optimum; the difference is caused by the estimation errors in the autoregressive model for $\{x_t\}$.

It was noted in section 2.2 that in practice a predictor used in the Box & Jenkins control scheme must have a steady-state gain of unity if the steady-state performance criterion is to be satisfied.

Unfortunately the optimum predictor defined by eqn(4.2.1) will, in general, have a steady-state gain different from unity. However if a model is estimated for the differenced disturbance, i.e.

$$\Delta x_t = \frac{1}{1 - \hat{q}(z^{-1})} \zeta_t$$

such that $x_t = \frac{1}{(1-z^{-1})(1-\hat{q}(z^{-1}))} \zeta_t$ (4.2.4)

then a predictor derived from eqn.(4.2.4) always has unity steady-state gain. The second 3 rows of Table 4.3 show examples of the predictors and calculated error variances based on the integrated model (4.2.4) for 200 samples of Series III. It can be seen that, in each case, the calculated error variance is significantly greater than for the optimum predictor; in particular the increase is 10% for $k = 1$ but rises to 25% for $k = 4$. In general, a predictor based on an integrated model is a satisfactory approximation to the optimum for $k = 1, 2$ but becomes increasingly less accurate for larger steps ahead.

An alternative method of imposing the steady-state gain condition is to base the predictor on an adapted model of the form

$$[1 - \hat{q}(z^{-1})] = [1 - \hat{q}(z^{-1})] [1 - g(z^{-1})] \quad (4.2.5)$$

where, as proposed in section 2.5, $g(z^{-1})$ is chosen to be an averaging function -

$$g(z^{-1}) = \frac{1}{\nu} \sum_{i=1}^{\nu} z^{-i}$$

Then from eqn(2.5.22) the sub-optimal error variance is approximately given by

$$\text{var}\{e_T(k)\} \sim \frac{\nu + 1}{\nu} \text{opt}[\text{var}\{e_T(k)\}] \text{ for large } \nu. \quad (4.2.6)$$

Examples of sub-optimum prediction for 200 samples of Series III are shown in Table 4.3. It can be seen that, as expected from eqn(4.2.6), the error variance is smaller for $\nu = 20$ than for $\nu = 10$.

Fig.4.1. DIFFERENCED VERSION OF THE BOX & JENKINS CONTROL SCHEME

Also, in both cases, the calculated error variance is less than the corresponding value for the integrated model - a feature particularly apparent at large values of the predictor step k .

4.3. Updating Analyses

This section presents some examples of the application of the updating techniques of Chapter III to simulated processes. In each case the process $P(z^{-1})$ is simulated by a digital filter and the disturbance $\{x_t\}$ by an artificially generated random signal.

Firstly it should be noted that it is not always necessary to impose the condition that the predictor must have a steady-state gain of unity. If it is known a priori that

- a) the disturbance $\{x_t\}$ is a zero-mean stationary random signal,
- b) the set-point $\{E_t^d\}$ is constant,

than it can be shown that the alternative control scheme shown in Fig. 4.1 has zero steady-state errors. The system equation (2.2.1) becomes

$$\Delta E_t = \frac{M(z^{-1})[1-H(z^{-1})]\Delta x_t + z^{-k}P(z^{-1})\Delta E_t^d}{M(z^{-1})[1-H(z^{-1})] + z^k P(z^{-1})H(z^{-1})} \quad (4.3.1)$$

and

$$\epsilon\{E_t\} = \epsilon\{E_0 + \sum_{i=1}^k \Delta E_i\} = E_0$$

Thus if E_0 is set equal to E^d then the steady-state output is equal to the set-point. Similarly if $X_0 = 0$ then $\epsilon\{X_t\} = 0$, and the signal $\{X_t\}$ can be reconstructed by summing the $\{\Delta X_i\}$, $i = 1, \dots, t$.

It has been found convenient to consider only those systems which satisfy assumptions a) and b) for the purposes of simulating the methods of process-model and predictor updating - it should be noted that in some cases it is possible to extend their application to non-zero mean disturbances. The control system of Fig. 4.1 has therefore been used to test these techniques.

Process-model updating

The first method of updating the process-model, derived in

TABLE 4.4. PROCESS-MODEL UPDATE I : N=400

Disturbance - Series I - $\text{var}\{x_t\} = 12542$

Estimated disturbance model - $[1-\hat{Q}(z^{-1})] = 1-0.58z^{-1}+0.43z^{-2}-0.24z^{-3}$

Process	Process-model	Predictor	$\text{var}\{E_t\}$	$\text{var}\{X_t\}$
$z^{-k}P(z^{-1})$	$z^{-l}M(z^{-1})$	$z^lH(z^{-1})$		
$\frac{z^{-1}}{1-0.5z^{-1}}$	$2z^{-1}$	$0.58-0.43z^{-1}+0.24z^{-2}$	11560	19323

Autoregressive model for $\{X_t\} : \hat{R}(z^{-1})$		Updated process- model $\hat{P}(z^{-1})$		Process $P(z^{-1})$	
$\hat{r}_1$	0.87	$\hat{p}_0$	1.01	p_0	1.00
$\hat{r}_2$	-0.96	$\hat{p}_1$	0.51	p_1	0.50
$\hat{r}_3$	0.72	$\hat{p}_2$	0.28	p_2	0.25
$\hat{r}_4$	-0.35	$\hat{p}_3$	0.15	p_3	0.125
$\hat{r}_5$	0.15	$\hat{p}_4$	0.07	p_4	0.063

	Open-loop	Initial control	Updated control	Optimum
$\text{var}\{E_t\}$	12542	11560	9854	9107
$\text{var}\{X_t\}$	-	19323	13690	12542
D	1.00	0.92	0.79	0.73

section 3.2., assumes that a model for the disturbance $\{x_t\}$ can be estimated from a record of the uncontrolled output - i.e. when the process is operated open-loop. In the many simulations of this technique the following computational procedure was adopted:-

- (i) an estimate $1 - \hat{Q}(z^{-1})$ of the disturbance model was calculated from the first N samples of the disturbance series,
- (ii) a predictor $z^{-1}H(z^{-1})$ was then calculated from eqn.(4.2.1) and the process-model $z^{-1}M(z^{-1})$ taken to be a rough approximation to the process $P(z^{-1})$,
- (iii) the control system of Fig. 4.1 was simulated on a digital computer and a record of N samples of $\{X_t\}$ obtained, $\{x_t\}$ being the second set of N samples of the disturbance series.
- (iv) an autoregressive model $1 - \hat{R}(z^{-1})$ for $\{X_t\}$ was computed and the process-model updated from eqn.(3.2.3), i.e.

$$z^{-1}\hat{P}(z^{-1}) = \frac{\hat{Q}(z^{-1}) - \hat{R}(z^{-1}) + H(z^{-1})[1 - \hat{Q}(z^{-1})]}{H(z^{-1}) [1 - \hat{Q}(z^{-1})]} \cdot M(z^{-1})$$

- (v) further sets of N samples of $\{X_t\}$ and $\{E_t\}$ were then recorded for the updated control system in order to compare the performance of the system to the theoretical optimum.

Throughout this section it will be found convenient to use a figure of merit D defined by

$$D = \frac{\text{var } \{E_t\}}{\text{var } \{x_t\}} \quad (4.3.2)$$

as a means of comparing the performance of the system, before and after updating, with the open-loop case.

An example of the procedure is shown in Table 4.4, where in this case the disturbance was Series I and N was 400. It can be seen that the updated process-model was a close approximation to the true process. Also the figure of merit D, although only slightly less than

TABLE 4.5. PROCESS-MODEL UPDATE II : N=200

Disturbance - Series I - $\text{var}\{X_t\} = 12360$

Estimated disturbance model - $[1-\hat{Q}(z^{-1})] = 1-0.52z^{-1}+0.36z^{-2}-0.23z^{-3}$

Process	Process-model	Predictor	$\text{var}\{E_t\}$	$\text{var}\{X_t\}$
$z^{-k}P(z^{-1})$	$z^{-l}M(z^{-1})$	$z^{-l}H(z^{-1})$		
$\frac{4z^{-1}}{1-0.5z^{-1}}$	$6z^{-1}$	$0.52-0.36z^{-1}+0.23z^{-2}$	10499	15397

Autoregressive model for $\{X_t\}$: $\hat{R}(z^{-1})$		Updated process- model $P(z^{-1})$		Process $P(z^{-1})$	
$\hat{r}_1$	0.69	$\hat{p}_0$	4.08	p_0	4.00
$\hat{r}_2$	-0.74	$\hat{p}_1$	2.05	p_1	2.00
$\hat{r}_3$	0.56	$\hat{p}_2$	1.00	p_2	1.00
$\hat{r}_4$	-0.27	$\hat{p}_3$	0.59	p_3	0.50
$\hat{r}_5$	0.12	$\hat{p}_4$	0.25	p_4	0.25

	Open-loop	Initial control	Updated control	Optimum
$\text{var}\{E_t\}$	12360	10499	9763	9107
$\text{var}\{X_t\}$	-	15397	12842	12360
D	1.00	0.85	0.79	0.74

TABLE 4.6. PROCESS-MODEL UPDATE III : N=200

Disturbance - Series IV - $\text{var}\{X_t\} = 227.9$

Estimated disturbance model - $[1-\hat{Q}(z^{-1})] = 1-1.71z^{-1}+0.78z^{-2}$

Process	Process-model	Predictor	$\text{var}\{E_t\}$	$\text{var}\{X_t\}$
$z^{-k}P(z^{-1})$	$z^{-l}M(z^{-1})$	$z^{-l}H(z^{-1})$		
$\frac{4z^{-1}}{1-0.5z^{-1}}$	$6z^{-1}$	$1.71-0.78z^{-1}$	25.9	171.9

Autoregressive model for $\{X_t\}$: $\hat{R}(z^{-1})$		Updated process- model $\hat{P}(z^{-1})$		Process $P(z^{-1})$	
$\hat{r}_1$	2.30	$\hat{p}_0$	3.95	p_0	4.00
$\hat{r}_2$	-2.63	$\hat{p}_1$	2.02	p_1	2.00
$\hat{r}_3$	1.90	$\hat{p}_2$	0.93	p_2	1.00
$\hat{r}_4$	-0.69	$\hat{p}_3$	0.43	p_3	0.50

	Open-loop	Initial control	Updated control	Optimum
$\text{var}\{E_t\}$	227.9	25.9	8.7	8.9
$\text{var}\{X_t\}$	-	171.9	220.3	227.9
D	1.00	0.11	0.04	0.04

1 for the initial control system, was close to the optimum for the updated control system. A similar example is shown in Table 4.5, where this time a different section of Series I was used and N was reduced to 200. The results of updating were much the same as the first example.

As a check on the validity of assumption a) above for these two examples, the estimated disturbance model $1 - \hat{Q}(z^{-1})$ was compared with the model calculated from the set of disturbance data used in the simulated control system. In both cases it was found that the two models differed only slightly. However when a similar test was carried out on successive records of 100 samples of Series I, the two models were significantly different. In consequence an updating analysis based on $N = 100$ produced inaccurate estimates of the process parameters and D was greater than one for the updated control system.

A further example of process-model updating is shown in Table 4.6. where the disturbance was Series IV and N was 200. In this case D shows more clearly the improvement from open-loop to the initial control system and again to the updated control system. The calculated value of $\text{var}\{E_t\}$ was in fact less than the theoretical optimum, for the updated control system - due to the two figures being calculated over different lengths of data. Once again the disturbance model had not varied significantly over successive records of 200 samples.

In conclusion this method was found to produce accurate estimates of the process parameters given a stationary disturbance and sufficiently large samples of data. In all cases where it could be shown that $\hat{Q}(z^{-1})$ varied markedly the method proved unsatisfactory. It should be noted that these checks cannot be applied in practice so that some degree of uncertainty is always present in results produced by this method.

Two-controller updating

The second proposal of section 3.2 requires the process to be

TABLE 4.7. TWO-CONTROLLER UPDATING : N=200

Disturbance - Series IV - $\text{var}\{x_t\} = 227.9$

Process		Process-model I	Process-model II	Predictor	
$z^{-k}P(z^{-1})$		$z^{-1}M_1(z^{-1})$	$z^{-1}M_2(z^{-2})$	$z^{-1}H(z^{-1})$	
$\frac{z^{-1}}{1-0.5z^{-1}}$		0.67	$\frac{z^{-1}}{1-0.7z^{-1}}$	$\frac{0.8}{1-0.2z^{-1}}$	
Autoregressive model for $\{X_t\}$			Updated process-		Process
$\hat{R}_1(z^{-1})$ $\hat{R}_2(z^{-1})$		model $\hat{P}(z^{-1})$		$P(z^{-1})$	
$\hat{r}_1$	1.28 1.71	$\hat{p}_0$	1.08	p_0	1.00
$\hat{r}_2$	-0.81 -0.66	$\hat{p}_1$	0.55	p_1	0.50
$\hat{r}_3$	0.41 -0.15	$\hat{p}_2$	0.43	p_2	0.25
Estimated Disturbance			Disturbance Model		
Model $\hat{Q}(z^{-1})$			$Q(z^{-1})$		
$\hat{q}_1$	1.76		q_1	1.70	
$\hat{q}_2$	-0.90		q_2	-0.80	
Initial control			Updated	Optimum	
I II		control			
$\text{var}\{E_t\}$	13.1 32.2	10.4		8.9	
$\text{var}\{X_t\}$	34.1 316.2	244.4		227.9	
D	0.06 0.14	0.05		0.04	

operated in closed-loop with two substantially different controllers. Even in simulation it was often found difficult to meet this requirement and simultaneously decrease the output variance from the open-loop value. An example of the technique is shown in Table 4.7. The first control system with a controller $W_1(z^{-1})$ used samples 1 to 200 of Series IV and the second system, controller $W_2(z^{-1})$, used samples 201 to 400. The parameters of the autoregressive models $\hat{R}_1(z^{-1})$, $\hat{R}_2(z^{-1})$ for $\{X_t, t = 1, 2, \dots, 200\}$, $\{X_t, t = 201, \dots, 400\}$ are shown in Table 4.7. The updated process-model parameters $\hat{p}_i$, $i = 0, 1, 2$ were calculated from eqn.(3.2.5), i.e.

$$\hat{P}(z^{-1}) = \frac{[\hat{R}_1(z^{-1}) - \hat{R}_2(z^{-1})][1-H(z^{-1})]}{H(z^{-1})\{W_2(z^{-1})[1-\hat{R}_1(z^{-1})] - W_1(z^{-1})[1-\hat{R}_2(z^{-1})]\}}$$

and the disturbance model parameters $\hat{q}_1$, $\hat{q}_2$ from eqn.(3.2.6), i.e.

$$1 - \hat{Q}(z^{-1}) = \frac{W_2(z^{-1})[1-\hat{R}_1(z^{-1})] - W_1(z^{-1})[1-\hat{R}_2(z^{-1})]}{[W_2(z^{-1}) - W_1(z^{-1})][1-H(z^{-1})]}$$

It can be seen from Table 4.7 that, in this case, the updated polynomials $\hat{P}(z^{-1})$ and $\hat{Q}(z^{-1})$ were reasonably good estimates of the true polynomials $P(z^{-1})$ and $Q(z^{-1})$. However in most cases the updated parameters were found to be very sensitive to small estimation errors in the autoregressive parameters $\hat{r}_i$, and in consequence were often highly inaccurate estimates of the corresponding true values. In conclusion, this alternative method of process-model updating has been found to be difficult to apply and unreliable in estimating the process and disturbance models. It therefore appears that the method is unlikely to be of use for the solution of practical identification or updating problems.

Predictor updating

Section 3.3 proposed a method of optimising the Box & Jenkins control system by periodically updating the predictor. This technique

TABLE 4.8. PREDICTOR UPDATING : N=200

Disturbance - Series I - $\text{var}\{x_t\} = 12542$

Process	Initial process-	Initial predictor
$z^{-k}P(z^{-1})$	model $z^{-1}M(z^{-1})$	$z^{-1}H(z^{-1})$
$\frac{z^{-1}}{1-0.5z^{-1}}$	z^{-1}	$0.58z^{-1}-0.42z^{-2}+0.24z^{-3}$

Autoregressive models for $\{x_t\}$

	$\hat{R}^1(z^{-1})$	$\hat{R}^2(z^{-2})$	$\hat{R}^3(z^{-3})$	$\hat{R}^4(z^{-4})$	$\hat{R}^5(z^{-5})$	$\hat{R}^6(z^{-6})$
$\hat{r}_1$	0.52	0.68	0.42	0.62	0.52	0.54
$\hat{r}_2$	-0.66	-0.84	-0.49	-0.71	-0.64	-0.66
$\hat{r}_3$	0.43	0.71	0.42	0.48	0.51	0.55
$\hat{r}_4$	-0.23	-0.51	-0.17	-0.31	-0.22	-0.21
$\hat{r}_5$	-	0.40	-	0.16	0.22	0.22
$\hat{r}_6$	-	-0.21	-	-	-	-
$\hat{r}_7$						

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Optimum
$\text{var}\{E_t\}$	11256	10338	11458	9863	9172	9348	9107
$\text{var}\{x_t\}$	16910	14501	16389	14403	11950	12754	12542
D	0.90	0.82	0.91	0.79	0.73	0.75	0.73

was shown to be limited to cases with small differences between the process and process-model. In particular a necessary condition for convergence of the procedure is $0 < p_0 w_0 < 2$. For all the simulation studies in which $p_0 w_0$ was chosen outside this range the predictor updating procedure was divergent. It was also found that the technique was only reliable for large samples of data - e.g. for sample lengths much less than 200 the procedure rarely converged. On the other hand convergence was established in a large number of cases which satisfied the requirements of a constant disturbance model and small process/process-model errors.

A typical (successful) example of the results of updating the predictor every N samples is shown in Table 4.8. In this case Series I was used as the disturbance $\{x_t\}$ and the initial control system parameters were as shown in Table 4.8. Each time a record of 200 samples of $\{X_t\}$ became available an autoregressive model $\hat{R}(z^{-1})$ was estimated; the updated one-step ahead predictor, calculated from eqn.(4.2.1), then took the form

$$H^i(z^{-1}) = \hat{R}^i(z^{-1}), \quad i = 1, 2, \dots$$

Computed values of $\text{var}\{E_t\}$ and $\text{var}\{X_t\}$ for each set of 200 samples between updates are shown at the foot of Table 4.8. It can be seen, from the values of D , that the system converged towards the optimum and in particular that convergence was established by the 5th iteration.

This result was confirmed for a range of disturbances and process/process-model errors. In all cases, which satisfied the conditions $0 < p_0 w_0 < 2$ and $1 = k$, convergence could be established by the 5th iteration. However since $N = 200$ appeared to be the minimum sample length, a total of about 1000 samples was needed to optimise the system. Also, as is noticeable in the example of Table 4.8, the convergence of the procedure was not found to be monotonic. In fact in many cases, where convergence was eventually established, the first

or second updated control system performed less well than open-loop - i.e. D_1 , D_2 became greater than 1. Due to this undesirable feature of the convergence, to the limitations imposed on the process-model and disturbance, and to the relatively long time required for optimisation, this method does not appear to be suitable as a generally applicable means of updating a control scheme. However it may find a useful application in a final automatic 'tuning' of a control system after an alternative method of identification has produced estimates of the process and disturbance models.

Separation effect

So far it has been shown that an estimate of the process impulse-response can be calculated and the Box & Jenkins control system updated if the process disturbance is known a priori to be stationary. No such restriction is placed on the separation technique of section 3.4. In this case the limiting factor is the number of samples k in the process time-delay as compared with the number of parameters m in an autoregressive model for the disturbance. The results of section 4.1 suggest that, since m is unlikely to be greater than 3 for chemical process disturbances, the separation technique should generally apply to processes with time-delays of 4 or more sample intervals.

Many simulations of the Box & Jenkins control scheme have been carried out to test the separation technique of simultaneous identification and control: in all cases an attempt was made to simulate the conditions to be expected in practice. Since in most practical problems very little quantitative information concerning the disturbance will be available it is not, in general, possible to assume a priori that the disturbance is stationary. Under these circumstances an initial control system must guard against steady-state errors caused by a varying (or constant non-zero) mean disturbance; as discussed in section 2.2. this can only be achieved, in the Box & Jenkins control

TABLE 4.9. SEPARATION EFFECT I : N=400

Disturbance - Series III - $\text{var}\{x_t\} = 35.8$

Process $z^k P(z^{-1})$	Process model $z^{-1} M(z^{-1})$	Predictor $z^{-1} H(z^{-1})$
$0.70z^{-4} + 1.40z^{-5} + 0.20z^{-6}$	$2.20z^{-5} + 1.30z^{-6}$	0.7
$(1-0.2z^{-1})(1-0.4z^{-1})(1-0.6z^{-1})$	$(1-0.3z^{-1})(1-0.5z^{-1})$	$(1-0.3z^{-1})$

Autoregressive model for $\{x_t\}$: $\hat{R}(z^{-1})$		Updated process- model $\hat{P}(z^{-1})$		Process $P(z^{-1})$	
$\hat{r}_1$	1.73 ± 0.11	$\hat{p}_0$	0.75 ± 0.66	p_0	0.70
$\hat{r}_2$	-0.82 ± 0.20	$\hat{p}_1$	2.34 ± 0.85	p_1	2.24
$\hat{r}_3$	0.01 ± 0.22	$\hat{p}_2$	2.54	p_2	2.58
$\hat{r}_4$	-0.24 ± 0.22	$\hat{p}_3$	1.92	p_3	2.14
$\hat{r}_5$	0.63 ± 0.22	$\hat{p}_4$	1.42	p_4	1.54
$\hat{r}_6$	-0.52 ± 0.22				
$\hat{r}_7$	0.06 ± 0.20				
$\hat{r}_8$	-0.06 ± 0.12				

Estimated Disturbance Model $\hat{Q}(z^{-1})$		Disturbance Model $Q(z^{-1})$	
$\hat{q}_1$	1.73 ± 0.11	q_1	1.75
$\hat{q}_2$	-0.82 ± 0.20	q_2	-0.82

	Open-loop	Initial control	Updated control	Optimum
$\text{var}\{E_t\}$	35.80	30.14	16.56	15.23
D	1.00	0.84	0.46	0.43

scheme, by choosing a predictor with a steady-state gain of unity. This feature was therefore incorporated in all the simulations of the separation technique. In particular, simulation experience has shown that a first-order low-pass filter is a convenient form for the initial predictor.

As an example of the calculations involved in the application of the separation technique consider the case shown in Table 4.9. The system was simulated, as before, with an artificially-generated random signal as the process disturbance $\{x_t\}$. The simulated process $z^{-k}P(z^{-1})$, given in Table 4.9, was taken to be a 3rd-order pulse transfer function with the process-model a 2nd-order approximation; in particular $l = 5$ and the process-model gain was 10 as compared with $k = 4$ and a gain of 12 for the true process. The predictor $z^{-1}H(z^{-1})$ was chosen to be a first-order low-pass filter with a steady-state gain of unity and the controller $W(z^{-1})$ was taken to be the exact inverse of the process-model.

The simulated process was operated by the initial control system for 400 sample intervals with the first 400 numbers of Series III as the disturbance $\{x_t\}$; this yielded a record of $\{X_t, t = 1, 2, \dots, 400\}$. Following the computational procedure of section 3.4., an autoregressive model for $\{X_t\}$ was then calculated by the linear-least-squares algorithm of eqn.(2.4.9); the parameters of the estimated 8th-order model $\hat{R}(z^{-1})$ and the associated 95% confidence intervals are shown in Table 4.9. To establish separation the reduction procedure of section 3.4 (ii) is employed - i.e. any parameter estimate which lies inside its 95% confidence interval is assumed to be insignificantly different from zero. For example, since

$$\hat{r}_3 = 0.01 \pm 0.22$$

$$\text{then } r_3 \stackrel{\Omega}{\approx} 0$$

The reduced polynomial $\hat{R}(z^{-1})$ is then given by

$$\hat{R}(z^{-1}) = 1.73z^{-1} + 0.82z^{-2} - 0.24z^{-4} + 0.63z^{-5} - 0.52z^{-6}$$

Fig.4.2. IMPULSE-RESPONSES FOR SEPARATION EFFECT I

Now, since $l = 5$ separation, if it occurs, must be before the 5th sample. The elimination of $\hat{r}_3$ by the reduction therefore suggests that separation occurs at the second sample and that

$$m = 2 = \text{number of parameters in disturbance model}$$

$$\beta = 4 = \min(1, k)$$

It follows that since $l = 5$, k must equal 4 sample intervals.

The updating equations(3.4.6) and (3.4.7) then become

$$\begin{aligned} \hat{q}_i &= \hat{r}_i, \quad i = 1, 2 \\ \hat{z}^{-4} \hat{p}(\hat{z}^{-1}) &= \frac{H(\hat{z}^{-1})[1 - \hat{r}_1 \hat{z}^{-1} - \hat{r}_2 \hat{z}^{-2}] - \sum_{i=4}^8 \hat{r}_i \hat{z}^{-i}}{\hat{z}^5 H(\hat{z}^{-1})[1 - \hat{r}_1 \hat{z}^{-1} - \hat{r}_2 \hat{z}^{-2}] W(\hat{z}^{-1})} \end{aligned}$$

and substituting for $H(\hat{z}^{-1})$, $W(\hat{z}^{-1})$, from Table 4.9

$$\hat{z}^{-4} \hat{p}(\hat{z}^{-1}) = \frac{\{\hat{z}^{-5} [1 - \hat{r}_1 \hat{z}^{-1} - \hat{r}_2 \hat{z}^{-2}] - 1.43(1 - 0.3\hat{z}^{-1}) \sum_{i=4}^8 \hat{r}_i \hat{z}^{-i}\} \{2.2 + 1.3\hat{z}^{-1}\}}{[1 - \hat{r}_1 \hat{z}^{-1} - \hat{r}_2 \hat{z}^{-2}](1 - 0.3\hat{z}^{-1})(1 - 0.5\hat{z}^{-1})}$$

The updated process-model parameters $\hat{p}_0, \hat{p}_1, \dots, \hat{p}_4$ calculated from this equation are shown in Table 4.9, and compared with the corresponding true values. Fig. 4.2 shows the updated process-model impulse response together with the true process and initial process-model impulse responses.

As proposed in section 3.4 the estimation errors on the parameters $\hat{p}_0, \hat{p}_1$ and the bias on $\hat{p}_1$ can be calculated as follows:-

$$\text{from eqn. (3.4.15), } \text{var}\{\hat{p}_0\} = \frac{\epsilon\{\phi_\beta^2\}}{(h_0 w_0)^2}$$

where the ϕ_i are the errors $\hat{r}_i - r_i$, and since $\beta = 4$, $h_0 w_0 = 0.318$,

$$\text{var}\{\hat{p}_0\} = 10 \epsilon\{\phi_4^2\} = 0.11$$

such that the 95% confidence interval for $\hat{p}_0$ is given by

$$\hat{p}_0 = 0.75 \pm 0.66$$

Also from eqn.(3.4.15) the bias on $\hat{p}_1$ is given by

$$\epsilon\{\hat{p}_1 - p_1\} = - \epsilon\left\{\frac{\phi_\beta \phi_1}{h_o w_o}\right\} = 0.006$$

and the variance of $\hat{p}_1$ is calculated from eqn.(3.4.17), i.e.

$$\text{var}\{\hat{p}_1\} \underline{\Omega} \left(\frac{1}{h_o w_o}\right)^2 \epsilon\{\phi_1^2 r_\beta^2 + \phi_\beta^2 + \gamma^2 \phi_\beta^2 + \hat{r}_\beta \phi_1 \phi_{\beta+1} + \gamma \phi_\beta^2 + \gamma \phi_\beta \phi_{\beta+1}\}$$

In this example, $\gamma = \frac{h_1}{h_o} + \frac{w_1}{w_o} + \hat{q}_1 = 0.30 - 1.39 + 1.73 = 0.64$

$$\text{and } \text{var}\{\hat{p}_1\} = 0.18$$

$$\text{such that } \hat{p}_1 = 2.34 \pm 0.85$$

It is noted that the bias on $\hat{p}_1$ is approximately 100 times less than the standard deviation of $\hat{p}_1$ and is therefore neglected. Expressions for the bias and variance of the remaining parameters $\hat{p}_2, \dots, \hat{p}_4$ are very complex and numerical values cannot easily be computed. However it can be shown that the bias and variance of a parameter $\hat{p}_i, i = 2, \dots, 4$ are of the same order of magnitude as the more readily computed values for $\hat{p}_1$. For instance, in this example,

$$\begin{aligned} \epsilon\{\hat{p}_2 - p_2\} &= 0.06 \\ \text{and } \text{var}\{\hat{p}_2\} &\underline{\Omega} 0.90 \end{aligned}$$

The estimated disturbance model, shown in Table 4.9, was used to calculate a sub-optimum predictor (see section 4.2) for inclusion in an updated control system. With the updated process-model also included the new control system was close to the theoretical optimum - the values of $\text{var}\{E_t\}$ and D are shown at the foot of Table 4.9.

The above example is a straightforward application of the separation technique which follows exactly the computational procedure suggested in section 3.4. There are, however, many examples which

TABLE 4.10. SEPARATION EFFECT II : N=200

Disturbance : $x_t = (1-0.9z^{-1})^{-1}\zeta_t$: $\text{var}\{x_t\} = 4.01$

Process		Process-model		Predictor	
$z^{-k}P(z^{-1})$		$z^{-l}M(z^{-1})$		$z^{-1}H(z^{-1})$	
z^{-4}		z^{-4}		0.8	
$(1-0.5z^{-1})(1-0.3z^{-1})$		$(1-0.7z^{-1})$		$(1-0.2z^{-1})$	
Autoregressive model for $\{x_t\}$: $\hat{R}(z^{-1})$		Updated process- model $\hat{P}(z^{-1})$		Process $P(z^{-1})$	
$\hat{r}_1$	0.91 ± 0.14	$\hat{p}_0$	1.00 ± 0.25	p_0	1.00
$\hat{r}_2$	-0.02 ± 0.20	$\hat{p}_1$	0.86 ± 0.28	p_1	0.80
$\hat{r}_3$	0.04 ± 0.20	$\hat{p}_2$	0.48	p_2	0.49
$\hat{r}_4$	0.04 ± 0.20	$\hat{p}_3$	0.37	p_3	0.28
$\hat{r}_5$	-0.13 ± 0.19	$\hat{p}_4$	0.25	p_4	0.15
$\hat{r}_6$	0.19 ± 0.19	$\hat{p}_5$	0.14	p_5	0.07
$\hat{r}_7$	-0.08 ± 0.20				
$\hat{r}_8$	0.04 ± 0.20				
$\hat{r}_9$	-0.02 ± 0.20				
$\hat{r}_{10}$	-0.08 ± 0.14				
Estimated Disturbance		Disturbance			
Model $\hat{Q}(z^{-1})$		Model $Q(z^{-1})$			
$\hat{q}_1$ 0.91 ± 0.14		q_1 0.90			
	Open-loop	Initial control	Updated control	Optimum	
$\text{var}\{E_t\}$	4.01	2.16	1.94	1.87	
D	1.00	0.54	0.48	0.47	

TABLE 4.11. SEPARATION EFFECT III : N=200

Disturbance : $x_t = (1-1.5z^{-1}+0.7z^{-2})^{-1}\zeta_t$: $\text{var}\{x_t\} = 6.79$

Process		Process-model		Predictor	
$z^{-k}P(z^{-1})$		$z^{-l}M(z^{-1})$		$z^{lh}(z^{-1})$	
z^{-4}		z^{-4}		0.8	
$(1-0.5z^{-1})(1-0.3z^{-2})$		$(1-0.7z^{-1})$		$(1-0.2z^{-1})$	
Autoregressive model for $\{x_t\}$: $\hat{R}(z^{-1})$			Updated process- model $\hat{P}(z^{-1})$		Process $P(z^{-1})$
$\hat{r}_1$	1.52 ± 0.14	$\hat{p}_0$	1.00 ± 0.35	p_0	1.00
$\hat{r}_2$	-0.79 ± 0.26	$\hat{p}_1$	0.89 ± 0.45	p_1	0.80
$\hat{r}_3$	0.10 ± 0.28	$\hat{p}_2$	0.51	p_2	0.49
$\hat{r}_4$	0.00 ± 0.28	$\hat{p}_3$	0.36	p_3	0.28
$\hat{r}_5$	-0.15 ± 0.28	$\hat{p}_4$	0.17	p_4	0.15
$\hat{r}_6$	0.29 ± 0.28	$\hat{p}_5$	0.07	p_5	0.07
$\hat{r}_7$	-0.21 ± 0.28				
$\hat{r}_8$	0.11 ± 0.28				
$\hat{r}_9$	-0.06 ± 0.26				
$\hat{r}_{10}$	-0.01 ± 0.14				
Estimated Disturbance			Disturbance		
Model $\hat{Q}(z^{-1})$			Model $Q(z^{-1})$		
$\hat{q}_1$	1.52 ± 0.14		q_1	1.50	
$\hat{q}_2$	-0.79 ± 0.26		q_2	-0.70	
	Open-loop	Initial control	Updated control	Optimum	
$\text{var } \{E_t\}$	6.79	8.71	4.51	4.24	
D	1.00	1.28	0.73	0.68	

require modification to the procedure if the identification results are to be reliable. One type of problem is shown in the simulation study of Table 4.10.; the disturbance, process and control system parameters are shown at the head of the table. This time 200 samples of $\{X_t\}$ were used to calculate the autoregressive model $\hat{R}(z^{-1})$. It is apparent from the estimated parameter values $\hat{r}_i$ and the associated 95% confidence intervals that, in this case, separation cannot be established by the proposed reduction procedure. We could establish separation, however, either by inspection or by comparing several models $\hat{R}(z^{-1})$ computed from various samples of the available data $\{X_t, t = 1, 2, \dots, N\}$. The latter method exploits particular properties of the parameter estimates $\hat{r}_i, i = \beta, \beta + 1, \dots$; an example is given, in the application of the separation technique to an industrial process, in Chapter V.

In the example of Table 4.10 separation can be established by inspection of the form of the polynomial $\hat{R}(z^{-1})$. Although the parameters $\hat{r}_5, \hat{r}_6$ lie inside the corresponding 95% confidence intervals they are greater in magnitude than their standard deviations i.e. $\hat{r}_5, \hat{r}_6$ lie outside the corresponding 66% confidence intervals. It can therefore be tentatively assumed that the true parameters r_5, r_6 are non-zero and that r_2, r_3, r_4 are insignificantly different from zero. Then separation occurs at the 1st sample giving $m = 1$ and $\beta = 5$. On this basis the updated process-model parameters $\hat{p}_i, i = 0, 1, \dots, 5$ and the disturbance model $\hat{Q}(z^{-1})$ are as shown in Table 4.10.

The use of such an arbitrary method of establishing separation is justified in practice by the fact that any choice of m and β other than the correct one leads, in general, to an unrealistic updated process-model.

A second problem in the application of the separation technique is illustrated by the simulation study of Table 4.11. In this case the initial control system's figure of merit is greater than unity

TABLE 4.12. SEPARATION EFFECT IV : N=200

Disturbance : $x_t = (1-0.9z^{-1})^{-1}\zeta_t$: $\text{var}\{x_t\} = 4.01$

Process		Process-model		Predictor	
$z^{-k}P(z^{-1})$		$z^{-l}M(z^{-1})$		$z^{-1}H(z^{-1})$	
$\frac{2z^{-4}+0.75z^{-5}}{(1-0.5z^{-1})}$		$3z^{-5}$		$\frac{0.8}{1-0.2z^{-1}}$	
Autoregressive model for $\{x_t\}$: $\hat{R}(z^{-1})$		Updated process- model $\hat{P}(z^{-1})$		Process $P(z^{-1})$	
$\hat{r}_1$	0.86 ± 0.15	$\hat{p}_0$	2.13 ± 0.71	p_0	2.00
$\hat{r}_2$	-0.06 ± 0.19	$\hat{p}_1$	1.89 ± 1.21	p_1	1.75
$\hat{r}_3$	0.08 ± 0.19	$\hat{p}_2$	0.81	p_2	0.87
$\hat{r}_4$	-0.57 ± 0.19	$\hat{p}_3$	0.48	p_3	0.44
$\hat{r}_5$	0.67 ± 0.20	$\hat{p}_4$	0.27	p_4	0.22
$\hat{r}_6$	-0.33 ± 0.20				
$\hat{r}_7$	-0.02 ± 0.19				
$\hat{r}_8$	0.04 ± 0.19				
$\hat{r}_9$	-0.02 ± 0.19				
$\hat{r}_{10}$	0.08 ± 0.15				
Estimated Disturbance Model $\hat{Q}(z^{-1})$		Disturbance Model $Q(z^{-1})$			
$\hat{q}_1$ 0.86 ± 0.15		q_1 0.90			
	Open-loop	Initial control	Updated control	Optimum	
$\text{var}\{E_t\}$	4.01	3.85	2.12	1.87	
D	1.00	0.96	0.53	0.47	

- i.e. the controlled output variance is greater than the open-loop variance. This undesirable situation may arise if the initial control system has been designed from inaccurate empirical data concerning the process and disturbance dynamics. However it is thought that in most practical cases an initial simulation study will lead to the design of an acceptable control system. The problem is accentuated for processes with large time-delays due to a decrease in the allowable margin of error.

In some cases a relatively short run under automatic control with an increased output variance may be acceptable since the accuracy of the parameter estimates derived from the separation technique is not affected. This can clearly be seen from Table 4.10 where as in the previous example the updated control system is close to the theoretical optimum.

The final problem, which will undoubtedly be encountered in practice, is concerned with the magnitude of the estimation errors for the updated process-model. As discussed in section 3.4 these errors are particularly dependent on $|h_0 w_0|$. When applying the separation technique this must be an important factor in the design of an initial control system.

The simulation study shown in Table 4.12 is an example of an undesirable choice of initial control system. In this case $(h_0 w_0)^{-1}$ is 3.75 compared with process parameters of 2.00 or less. In consequence the estimation errors for $\hat{p}_0$, $\hat{p}_1$ are relatively large and the actual estimates are relatively inaccurate. However the values of D at the foot of Table 4.12 show that the updated control system is close to the optimum.

In conclusion, simulation studies have shown that the separation technique can produce sufficiently accurate estimates of the process and disturbance dynamics from closed-loop operating data, given some degree of engineering judgement. In particular, it has been shown that

it is not always desirable to employ the standard computational procedure of section 3.4., due to the individual character of specific problems. However in all tested cases, where the condition $m < \beta$ is satisfied, it has been found possible to obtain the required estimates by simple modifications to the computational procedure. The most commonly encountered modifications have been presented above and will be further demonstrated in the application of the separation technique to the industrial process of Chapter V.

This section has only considered artificial random signals with stationary characteristics - due to the difficulties in simulating realistic non-stationary signals. However some work has been done on the separation technique using non-stationary disturbances and results have proved satisfactory. In particular, the separation technique has been applied to a system whose disturbance is one of the industrial data series quoted by Box & Jenkins⁵ and analysed in Table 4.2. It should also be noted that the two artificial signals Series III and Series IV although stationary in theory (and in practice for large samples) are similar to realistic non-stationary signals over short intervals - in fact they are similar to the evolutionary type of slowly-varying signals which Box & Jenkins⁵ would represent by an integrated model. In consequence these two data Series have been used extensively in the testing of the separation technique - e.g. Example I in Table 4.9.

No results concerning the application of the separation effect to a conventional control system will be presented here since they are, in general, similar to those presented above. It should be noted that the Box & Jenkins control scheme can be restructured to the form of a conventional scheme if the set-point $\{E_t^d\}$ is constant.

4.4. Summary and conclusions

The simulation results presented in this chapter have shown some applications of methods of simultaneous identification and control.

They can be summarised as follows:-

- (i) the method of process-model updating can be successfully applied to a system where the disturbance model is known a priori (or can be found from a record of the open-loop output) and is constant over successive records. The updated process parameters are, in general, good estimates of the process parameters only if long records - i.e. greater than about 150 samples - are available;
- (ii) the method of two-controller updating is difficult to apply successfully due to the requirement of two substantially different control systems. It is sensitive to estimation errors in the autoregressive model for $\{X_t\}$ and is therefore unlikely to be of use in practice;
- (iii) system optimisation by the method of predictor-updating is possible only when the differences between the process and process-model are small and the disturbance model remains almost constant. It appears to be best suited as a 'final tuning' procedure;
- (iv) the separation technique is readily applicable to processes with large time-delays. It is a robust technique which is particularly suited for practical applications. For instance, a good estimate of the process can be obtained even if the disturbance is non-stationary, or the amount of data is small and by a specific design of the initial control system can be made insensitive to estimation errors.

All the methods are relatively simple to apply and require a minimum of computation. It remains to show that they can apply to an industrial process and in particular that methods of simultaneous identification and control are better suited to chemical processes than the established open-loop techniques.

Fig.5.1 A SCHEMATIC DIAGRAM OF THE PROCESS

CHAPTER VAN INDUSTRIAL INVESTIGATION5.1. Introduction

The techniques of Chapter III were originally developed to identify and control a particular industrial chemical process. A schematic diagram of this process is shown in Fig. 5.1. The raw materials A and B react at a certain temperature to form the product P. A measurement Q of the quality of the product is made in an off-line laboratory.

The outstanding characteristics of the process are a large residence time, which is dependent on the feed rate of the raw materials, and large process disturbances which cause unwanted quality variations. These disturbances originate from several different sources which can be categorised as follows:-

- (a) the process input B is subject to an additive disturbance which cannot be measured accurately or frequently; the resulting fluctuations in product quality are of the order of $\pm 50\%$ of the desired quality value.
- (b) The composition of raw material A varies due to inconsistent plant operation and the requirements of production schedules. This can cause quality variations of up to $\pm 10\%$ of the desired value.
- (c) Production scheduling and inventory control also require frequent changes in process throughput; these changes alter the reaction dynamics and often cause large transients in quality.
- (d) Reaction temperature has a direct effect on quality and is inter-related with other process variables. It is possible that the effect of a process disturbance on quality may be increased by the related temperature change.

Fig.5.2 UNCONTROLLED PROCESS QUALITY RECORD

Fig.5.3 MANUALLY CONTROLLED PROCESS QUALITY RECORD

- (e) Mechanical breakdowns and general plant degeneration often cause extreme variations in process behaviour.

An example of the effect of the disturbances on product quality is shown in Fig. 5.2. This process record was taken over a period of time when all the inputs were maintained constant; it therefore represents an example of uncontrolled quality variations. The fluctuations shown in Fig. 5.2 were substantially caused by the input-type disturbances of (a) and (b) above - these will be represented by a lumped input disturbance d_1 . Higher-frequency and more severe short-term fluctuations usually result from types (c), (d) and (e) - these are representable by a disturbance d_2 acting at some intermediate stage in the reaction.

In the past, quality has been controlled by manual manipulation of the process variables according to the judgement and experience of process operators. Input B was the major control variable although frequent adjustments were made to the composition and feed rate of input A and to temperature. A typical record of the product quality and the associated control actions on B is shown in Fig. 5.3. This record illustrates the effect of some high-frequency, short-term disturbances - e.g. samples 12 to 16. A comparison of the manual control record with the uncontrolled record of Fig. 5.2 does not suggest that control has significantly reduced the fluctuations in quality. However, it has not been found possible to operate the process for a prolonged period without some form of control action. This is due to the input disturbance d_1 causing quality to drift steadily away from its desired value. Manual control has been most effective when used to eliminate this drift.

The above information formed the basis of a feasibility study aimed at determining the merits of using a digital computer to implement an automatic control system for product quality. The conclusions were favourable and can be stated, in brief, as follows,

- (i) the most important aspect of the control problem was thought to be the reduction of quality variations caused by the process random disturbance;
- (ii) control could be based on the laboratory measurements of quality given a sufficiently small sampling interval. In fact, a satisfactory sampling interval T could be achieved without an excessive increase in the costs associated with laboratory measurements;
- (iii) the only immediately practical control variable was the input B - however, it was thought that the major disturbances (type d_1) could be controlled from this input;
- (iv) the Box & Jenkins feedback control system using B as the control variable was proposed : it was hoped that a 50% reduction in the amplitude of quality variations would be possible - i.e. a 75% reduction in variance;
- (v) any possible extension to a multi-input system was deferred until some experience had been gained in the use of a single-variable scheme.

The chosen sampling interval T will also be used as the basic time unit throughout this chapter.

A more precise study of the control problem started with the determination of an empirical process model.

5.2. Determining the empirical model

The following empirical relationships had been developed by the process operators from their experience of running the plant:-

- (a) the transportation lag between input B and Q could be calculated from

$$k_t = \alpha \beta_t y_t \quad (5.2.1)$$

where k_t , y_t and β_t are respectively the transportation lag, the process throughput and another measurable process variable, all taken at time t . The constant α - nominally 0.50 - could also be derived from a simplified physical model of the process. The re-

Fig.5.4 THE EMPIRICAL PROCESS MODEL

(a) Input Step +0.8 at $t=2T$

(b) Input Step -0.6 at $t=2T$

(c) Input Step +0.6 at $t=2T$

(d) Input Step -0.5 at $t=2T$

Fig.5.5 EXPERIMENTAL STEP RESPONSES

sulting transportation lag varied between 3 and 12 sample intervals.

(b) The steady-state gain between B and Q could be calculated from

$$\left. \frac{\delta Q}{\delta B} \right|_{Q=Q'} = 0.5Q' \quad (5.2.2)$$

where δQ is the change in the level of quality caused by a change δB at the input.

(c) The dynamic response of quality to changes in B was thought to be fast relative to the transportation lag; a reasonable approximation to the transfer function Q/B was taken to be two time constants of $\frac{1}{2}$ and 2 sample intervals.

An empirical process model based on these relationships is shown in Fig. 5.4. The non-linear gain of eqn.(5.2.2) is represented by $K(Q)$, and the measurement lag e^{-ft^s} represents the time taken for a laboratory analysis of the product - it varies between $\frac{1}{2}$ and $\frac{3}{4}$ of a sample interval.

Attempts to justify the empirical model of Fig. 5.4 from physico-chemical analyses of the process have not proved successful up to the present time. Furthermore, the only practical conventional technique of confirming the model was the step-response method. This involves the application of a predesigned set of step changes to the process input and measurement of the corresponding output responses. Unfortunately, large random fluctuations in quality distort the responses to such an extent that little quantitative information can be obtained. A few examples of experimental step responses are shown in Fig. 5.5. The input steps were chosen to be as large as possible under the prevailing process conditions. Responses (a) and (d) in Fig. 5.5 show distinct changes in steady-state level but with a marked difference in gain. On the other hand it is difficult to distinguish any significant response to the input steps in (b) and (c) - this illustrates the predominance of the random disturbances. From an extensive study of 40 such step

Fig.5.6 THE LINEAR DISCRETE PROCESS MODEL

responses it was only possible to conclude that the transportation lag, gain and time constants of the empirical model were of the right order.

At this stage it was proposed to use one of the more recent statistical identification techniques to obtain better estimates of the process and disturbance dynamics. However, the demands of production scheduling did not allow for the type of experimentation required by these techniques. In particular,

- (i) the process could not be operated open-loop for longer than about 50 sample intervals,
- (ii) normal-operation consisted of manual feedback control with frequent changes to process throughput, and
- (iii) an extra source of perturbation of sufficient magnitude to give accurate results could not be tolerated.

It follows from the discussion of section 1.3 that none of the established statistical identification techniques are applicable to this process.

The above limitations do not, however, affect the application of the separation technique of identification described in Chapter III. In fact this particular process exhibits all the characteristics which are required for the separation effect - i.e. a large time-delay and a disturbance predominantly autoregressive due to its passage through the low-pass filter process.

5.3. Application of the separation technique of identification

As described in section 3.5 the separation technique is best used in conjunction with the Box & Jenkins control system. The process model must, therefore, be represented as a single-input, single-output, linear discrete dynamic system. The corresponding empirical model is shown in Fig. 5.6. In this representation the additive noise $\{z_t\}$ is the effect on quality of all the disturbances mentioned in section 5.1. The process model time-delay l_t is the sum of the transportation lag k_t

Fig.5.7 THE PRACTICAL IMPLEMENTATION OF THE BOX & JENKINS CONTROL SCHEME

and the measurement lag f_t - in practice l_t must be integral so it is set equal to the integer part of $(k_t + f_t)$.

It should be noted that the process model pulse transfer function is the relationship between ΔQ_t and ΔB_t - the intermediate relationship between Q and B shown in Fig. 5.6 is not strictly valid due to the lack of a unique steady-state Q for a given constant B . The data-hold element is included to transform the series of impulses B_t into a staircase function. The process model pulse transfer function is then given by

$$z^{-1} M(z^{-1}) = z \left\{ \frac{1 - e^{-s}}{s} \cdot \frac{K e^{-l_t s}}{(1 + \frac{1}{2}s)(1 + 2s)} \right\} \quad (5.3.1)$$

where the sampling interval is T . Taking the z -transform,

$$z^{-1} M(z^{-1}) = \frac{(0.238 + 0.102 z^{-1}) K z^{-1}}{(1 - 0.606 z^{-1})(1 - 0.137 z^{-1})} \quad (5.3.2)$$

where K is the linearised steady-state gain derived from eqn.(5.1.2) and the time-delay l equals $l_t + 1$. Linearisation is achieved by approximating eqn.(5.1.2) by

$$K = 0.5 Q^d \quad (5.3.3)$$

where Q^d is the desired value of quality.

The controller $W(z^{-1})$ of the Box & Jenkins control scheme was therefore obtained by inverting the process model $M(z^{-1})$ given by eqn. (5.3.2). Since $W(z^{-1})$ was chosen to be the exact inverse of $M(z^{-1})$ a simplified version of the control scheme could be used - this is shown in Fig. 5.7. It has the significant practical advantage of avoiding the difficulty of coping with a change in the process-model time-delay.

The design of a predictor $H(z^{-1})$ was not, however, so straightforward. The only quantitative information available was the short record of uncontrolled quality variations shown in Fig. 5.2. Unfortunately an autoregressive model calculated from this set of data was

subject to excessively large estimation errors and in consequence could not justifiably be used to design a predictor. However, since the disturbance $\{x_t\}$ was thought to consist mainly of a slowly-varying drift the following predictor was considered to be a reasonable first approximation,

$$z^{-1}H(z^{-1}) = 1 \quad (5.3.4)$$

This predictor is optimal for a constant-valued disturbance but, since its steady-state gain is unity, it will also eliminate slow variations in product quality - see Appendix II. It should be noted that a Box & Jenkins control system with a predictor defined by eqn.(5.3.4) is identical to the conventional Smith controller, which has often been used to control processes with time-delays.

At this stage the performance of the control system was extensively studied by digital simulations. The process dynamics were represented by a digital filter and could be varied about the process-model values. A disturbance $\{x_t\}$ was artificially generated by passing a sequence of random numbers through a digital filter such that its characteristics were similar to those of the uncontrolled quality record. The simulated system performed satisfactorily - i.e. the output deviations were less than the disturbance - for the expected range of the process parameters.

Before attempting to run the process under fully-automatic control for a prolonged period, several short test runs were performed. These ensured correct program functioning and were helpful in introducing the process operators to the new procedures. The tests were conducted with first manual then automatic implementation of the calculated input values, and lasted for about 20 sample intervals. In all cases the system performed satisfactorily, so that it appeared unnecessary to make any major amendments to the control programs.

It was then decided to attempt a full-scale experimental run

TABLE 5.1. SYSTEM PARAMETERS : RUN I

Desired quality value	Q_o^d	4.2 units of quality
Prevailing process input	B_o	3.8 units of B
Process transportation lag	k_o	4.8 sample intervals
Quality measurement lag	f_o	0.75 sample intervals
Process-model time delay	l	6 sample intervals
Process-model gain	K	2.1 units of Q/B
Predictor pulse-transfer function	$z^{-l}H(z^{-1})$	1
Controller pulse-transfer function	$W(z^{-1})$	$\frac{(1-0.606z^{-1})(1-0.137z^{-1})}{0.50 + 0.214z^{-1}}$

FIG.5.8 PROCESS RECORD OF EXPERIMENTAL RUN I

under the conditions required by the separation technique of identification - i.e. that the set-point and process time-delay should remain constant. The plant managers were satisfied that these conditions could be met without a serious upset to production but retained the right to end the run if product quality exceeded certain specified limits.

Just before switching from manual to automatic control a set of initial signal values had to be inserted into the computer control program:-

$$Q_0 = Q_0^d = \text{desired quality value}$$

$$\Delta Q_0 = \Delta X_0 = \Delta B_0 = 0$$

$$B_0 = \text{prevailing process input}$$

For this first experimental run the control system parameters were as shown in Table 5.1.

A record of the quality, estimated disturbance and input values for Run I are shown in Fig. 5.8. It can be seen that quality was maintained within reasonably close limits of the desired value until about the 70th sample. The run was terminated at the 80th sample, however, due to an increasingly large fluctuation in quality which can be seen at the end of the record of Fig. 5.8. It should be noted that the magnitudes of the estimated disturbance $\{X_t\}$ and the input $\{B_t\}$ steadily increased as the run progressed, and became oscillatory towards the end. This suggested that the system may have been unstable. Alternatively, the excessively large fluctuations may have been caused by a large process disturbance which, it was subsequently established, took effect from the 70th sample. It could be concluded, however, that some redesign of the system was required before another computer-controlled run could be undertaken.

Although the record of $\{X_t\}$ obtained from Run I was shorter than hoped for, it was decided to attempt an identification analysis on

TABLE 5.2. AUTOREGRESSIVE PARAMETERS : RUN I

Number of Samples N			
	50	55	60
$\hat{r}_1$	0.39 ± 0.32	0.32 ± 0.30	0.30 ± 0.28
$\hat{r}_2$	0.23 ± 0.34	0.23 ± 0.32	0.27 ± 0.30
$\hat{r}_3$	$- 0.20 \pm 0.36$	$- 0.14 \pm 0.34$	$- 0.17 \pm 0.30$
$\hat{r}_4$	$- 0.52 \pm 0.34$	$- 0.53 \pm 0.32$	$- 0.52 \pm 0.30$
$\hat{r}_5$	0.04 ± 0.36	$- 0.08 \pm 0.34$	$- 0.12 \pm 0.32$
$\hat{r}_6$	0.71 ± 0.30	0.84 ± 0.26	0.94 ± 0.24
$\hat{r}_7$	$- 0.29 \pm 0.38$	$- 0.31 \pm 0.36$	$- 0.34 \pm 0.36$
$\hat{r}_8$	$- 0.44 \pm 0.34$	$- 0.39 \pm 0.34$	$- 0.38 \pm 0.34$
$\hat{r}_9$	0.22 ± 0.36	0.11 ± 0.36	0.09 ± 0.36
$\hat{r}_{10}$	0.18 ± 0.36	0.26 ± 0.36	0.29 ± 0.34
$\hat{r}_{11}$	$- 0.18 \pm 0.34$	$- 0.15 \pm 0.34$	$- 0.15 \pm 0.32$

Number of Samples N			
	65	70	75
$\hat{r}_1$	0.28 ± 0.28	0.29 ± 0.26	0.22 ± 0.24
$\hat{r}_2$	0.27 ± 0.28	0.26 ± 0.26	0.30 ± 0.24
$\hat{r}_3$	$- 0.16 \pm 0.28$	$- 0.15 \pm 0.28$	$- 0.13 \pm 0.26$
$\hat{r}_4$	$- 0.54 \pm 0.28$	$- 0.55 \pm 0.26$	$- 0.60 \pm 0.26$
$\hat{r}_5$	$- 0.10 \pm 0.30$	$- 0.08 \pm 0.28$	$- 0.07 \pm 0.30$
$\hat{r}_6$	0.90 ± 0.22	0.90 ± 0.20	0.86 ± 0.20
$\hat{r}_7$	$- 0.26 \pm 0.32$	$- 0.28 \pm 0.30$	$- 0.15 \pm 0.30$
$\hat{r}_8$	$- 0.47 \pm 0.30$	$- 0.44 \pm 0.30$	$- 0.51 \pm 0.26$
$\hat{r}_9$	0.15 ± 0.32	0.12 ± 0.30	0.06 ± 0.30
$\hat{r}_{10}$	0.25 ± 0.32	0.24 ± 0.30	0.31 ± 0.30
$\hat{r}_{11}$	$- 0.13 \pm 0.30$	$- 0.11 \pm 0.28$	$- 0.11 \pm 0.28$

the available data. The analytical procedure of section 3.4 was employed although it will be seen that several modifications were necessary in order to establish separation.

Firstly the linear-least-squares algorithm (eqn.(2.4.9)) was used to estimate the parameters of an autoregressive model for $\{\Delta X_t, t = 1, 2, \dots, 75\}$. Some difficulty was encountered in the choice of the model order n , due to the relatively large estimation errors. However, for models with order $n > 10$ the first 10 parameters $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_{10}$ remained approximately constant. It was therefore decided to choose $n = 11$ with the condition that further results should be checked by comparison with those for $n > 11$ (e.g. $n = 15$).

The parameters of the 11th-order model

$$\Delta X_t = \frac{1}{1 - \hat{R}(z^{-1})} \zeta_t, \quad \hat{R}(z^{-1}) = \sum_{i=1}^{11} \hat{r}_i z^{-i} \quad (5.3.5)$$

are shown in the last column of Table 5.2 - also shown are the 95% confidence limits of the parameters computed from the covariance matrix of the estimates. At this stage the reduction procedure, described in section 3.4, could have been used to eliminate those estimates which were found to lie inside the corresponding 95% confidence interval. It could not, however, be justified since the reduced polynomial contained only $\hat{r}_2, \hat{r}_4, \hat{r}_6, \hat{r}_8, \hat{r}_{10}$ and it was considered that several important parameters had thus been eliminated (e.g. $\hat{r}_1$). Also the reduction procedure could not satisfactorily establish separation.

Instead the analysis proceeded as follows:- rather than relying on an analysis of just the first 75 samples of ΔX_t it was decided to study the properties of several models each fitted to a different set of data. As before the linear-least-squares algorithm was used to compute estimates of the autoregressive parameters of an 11th-order model. The estimates $\hat{r}_i$, with corresponding 95% confidence limits, derived from a set of data $\{\Delta X_t, t = 1, 2, \dots, N\}$ are shown in Table 5.2.

Now the updating equation for Experimental Run I, where

$H(z^{-1}) = z^{-6}$, is given by

$$1 - \hat{R}(z^{-1}) \underline{\Omega} [1 - Q(z^{-1})] + z^{-6} [1 - Q(z^{-1})][z^{6-k} P(z^{-1}) W(z^{-1}) - 1] \quad (5.3.6)$$

$$\text{where } Q(z^{-1}) = \sum_{i=1}^m q_i z^{-i} \quad \text{and} \quad \hat{R}(z^{-1}) = \sum_{i=1}^n \hat{r}_i z^{-i}$$

Hence if separation occurs,

$$\hat{Q}(z^{-1}) = \sum_{i=1}^m \hat{r}_i z^{-i} \quad (5.3.7)$$

$$\text{and} \quad -\hat{R}_T(z^{-1}) \underline{\Omega} [1 - \hat{Q}(z^{-1})][z^{-k} P(z^{-1}) W(z^{-1}) - z^{-6}] \quad (5.3.8)$$

where $[\cdot]_T$ is the β^{th} -order truncation operator and $\beta = \min(1, k)$.

Consider the case where $k < 6$; then $\beta = k$ and the first term of $\hat{R}_T(z^{-1})$ in eqn.(5.3.8) is given by

$$-\hat{r}_k \underline{\Omega} p_0 w_0$$

This coefficient is independent of the disturbance model parameters $\hat{q}_1$ and should therefore remain constant in the several models for $\{\Delta X_t\}$, even if the $\hat{q}_i$ vary. From a study of Table 5.2 it can be seen that the parameter $\hat{r}_4$ exhibits this property. It could therefore be tentatively assumed that $k = 4$.

Another significant property of the autoregressive model follows from special cases of eqn.(5.3.8); i.e.

(i) if $k = 1 - 1$ then

$$-\hat{R}_T(z^{-1}) \underline{\Omega} p_0 w_0 z^{-k} + (p_0 w_1 + p_1 w_0 - p_0 w_0 \hat{q}_1 - 1) z^{-1} + \dots$$

(ii) if $k = 1 - 2$ then

$$\begin{aligned} &-\hat{R}_T(z^{-1}) \underline{\Omega} p_0 w_0 z^{-k} + (p_0 w_1 + p_1 w_0 - p_0 w_0 \hat{q}_1) z^{-k-1} \\ &+ [p_0 w_2 + p_1 w_1 + p_2 w_0 - (p_0 w_1 + p_1 w_0) \hat{q}_1 - p_0 w_0 \hat{q}_2 - 1] z^{-1} + \dots \end{aligned}$$

In both cases the parameter $\hat{r}_1$ will be close to unity if the controller $W(z^{-1})$ is roughly equal to the inverse of the process $P(z^{-1})$. Hence, since $l = 6$, the parameter $\hat{r}_6$ should be close to unity for all the models if the assumption that $k < l$ is correct. As Table 5.2 shows $\hat{r}_6$ is close to unity in all the models.

TABLE 5.3. UPDATED PROCESS-MODEL PARAMETERS : RUN I

	Number of Samples N					
	50	55	60	65	70	75
$\hat{p}_0$	0.25	0.25	0.25	0.26	0.26	0.29
$\hat{p}_1$	0.37	0.42	0.43	0.43	0.43	0.44
$\hat{p}_2$	0.52	0.52	0.50	0.50	0.50	0.52
$\hat{p}_3$	0.46	0.47	0.44	0.43	0.43	0.42
$\hat{p}_4$	0.44	0.44	0.39	0.41	0.40	0.41
$\hat{p}_5$	0.35	0.37	0.34	0.34	0.34	0.36
$\hat{p}_6$	0.18	0.17	0.14	0.15	0.17	0.17

Fig. 5.9 STEP RESPONSE OF UPDATED MODEL

These two properties were equally apparent in the 15th-order models for $\{\Delta X_t, t = 1, 2, \dots, N\}$.

While questioning the plausibility of the result that the process pure time-delay had been overestimated by 2 sample intervals it was found that the empirical relationship for the transportation lag (eqn.5.2.1) had been based on measurements of the time between an input change and a 'significant change' in quality. It was concluded that the meaning of a 'significant change' inferred by the process operators had led to a definite overestimate of the pure time-delay.

Under the assumption, then, that separation occurred at the 3rd sample in eqn.(5.3.6), and that $m = 3, k = 4$, the analysis could proceed to the final stage. From the updating equation (5.3.7) and (5.3.8),

$$\hat{q}_i = \hat{r}_i, i = 1, 2, 3$$

$$\text{and } -\hat{R}_T(z^{-1}) \underline{\Omega} \left[1 - \sum_{i=1}^3 \hat{r}_i z^{-i} \right] \left[z^{-4} P(z^{-1}) W(z^{-1}) - z^{-6} \right]$$

Then the updated process model was given by

$$\hat{P}(z^{-1}) = \frac{z^{-2} \left[1 - \sum_{i=1}^3 \hat{r}_i z^{-i} \right] - \sum_{i=4}^n \hat{r}_i z^{4-i}}{W(z^{-1}) \left[1 - \sum_{i=1}^3 \hat{r}_i z^{-i} \right]} \quad (5.3.9)$$

The parameters calculated from eqn.(5.3.9) for the 6 polynomials $\hat{R}(z^{-1})$ are shown in Table 5.3. As noted in section 3.4 only $n - \beta = 7$ parameters $\hat{p}_i$ are available from eqn.(5.3.9). The average of each set of estimates $\hat{p}_i$ was used to construct the updated process-model step response shown in Fig. 5.9 - the dotted lines represent the limits of the region containing the responses of all the six models. Similarly the updated process-model parameters calculated from a 15th-order autoregressive model $\hat{R}(z^{-1})$ produced step responses within the same region.

As an example of the theoretical estimation errors for the updated parameters consider the results for $N = 65$. Then from the

covariance matrix $(\hat{\underline{r}} - \underline{r})(\hat{\underline{r}} - \underline{r})^T$ it is found that,

$$\epsilon\{\phi_1^2\} = 0.019, \quad \epsilon\{\phi_2^2\} = 0.020, \quad \epsilon\{\phi_3^2\} = 0.020$$

and it follows that the 95% confidence intervals for the disturbance model parameters are given by

$$\begin{aligned} |\hat{q}_1 - q_1| &\leq 0.28 \text{ such that } q_1 = 0.28 \pm 0.28 \\ |\hat{q}_2 - q_2| &\leq 0.28 \text{ such that } q_2 = 0.27 \pm 0.28 \\ |\hat{q}_3 - q_3| &\leq 0.28 \text{ such that } q_3 = -0.16 \pm 0.28 \end{aligned}$$

Also from eqn.(3.4.16), since $\epsilon\{\phi_4^2\} = 0.18$

$$\text{var}\{\hat{p}_0\} = \epsilon\left\{\frac{\phi_4^2}{w_0^2}\right\} = 0.004$$

and the 95% confidence interval is

$$|\hat{p}_0 - p_0| \leq 0.17 \text{ such that } p_0 = 0.26 \pm 0.17$$

The bias on $\hat{p}_1$ is calculated from eqn.(3.4.15) as

$$\epsilon\{\hat{p}_1 - p_1\} = -m_0 \epsilon\{\phi_1 \phi_4\} = -0.0012$$

and the variance as

$$\begin{aligned} \text{var}\{\hat{p}_1\} = \epsilon\{\hat{p}_1 - p_1\}^2 \underline{\Omega} \epsilon\{m_0^2 [\phi_1^2 r_k^2 + \phi_k^2 \hat{q}_1^2 + \phi_{k+1}^2 + \hat{r}_k \phi_1 (\phi_{k+1} + \hat{q}_1 \phi_k)] \\ + m_1^2 \phi_{k+1}^2 + \hat{r}_k m_0 m_1 \phi_1 \phi_k + m_0 (m_0 \hat{q}_1 + m_1) \phi_k \phi_{k+1}\} \end{aligned}$$

where the m_i are the coefficients of $\underline{M}(\underline{z}^{-1}) = [\underline{W}(\underline{z}^{-1})]^{-1}$.

Hence $\text{var}\{\hat{p}_1\} \underline{\Omega} 0.01$

and the 95% confidence interval is given by

$$|p_1 - \hat{p}_1| \leq 0.20 \text{ such that } p_1 = 0.43 \pm 0.20$$

It is noted that the bias on $\hat{p}_1$ is almost 100 times less than the standard deviation of $\hat{p}_1$ and has therefore been neglected. Also all the estimates $\hat{p}_0, \hat{p}_1$ in Table 5.3 lie well within the above confidence intervals.

This identification analysis produced more consistent results than might have been expected both from the small amount and the oscillatory characteristics of the data. Although the disturbance

model parameters $\hat{q}_1$, $\hat{q}_2$, $\hat{q}_3$ varied appreciably from $N = 50$ to $N = 75$ the corresponding variation in the updated process-model parameters was small. This consistency must be evaluated with consideration for the relatively large estimation errors on the autoregressive and the updated process-model parameters.

The updated and empirical process-model step responses are both shown in Fig. 5.9 - by comparison it can be seen that the empirical model was in error in the following ways:-

- a) the time-delay l was overestimated by 50%
- b) the steady-state gain K was underestimated by 40%
- c) the dynamics could more adequately have been described by a third or fourth-order transfer function.

As a final check on the reliability of these identification results an attempt was made to reconstruct Run I in a digital simulation. The simulated process was, in fact, the updated process-model and the control system the same as in Run I. The simulation was performed with several artificially-generated disturbances whose characteristics were similar to the signal $\{X_t\}$ but without the oscillation. In all cases the system behaved in a similar fashion to Run I - of particular significance was the appearance of an oscillation in $\{X_t\}$ and $\{B_t\}$ which did not affect $\{Q_t\}$ until the system was subjected to a large disturbance. It was concluded that the system was on the verge of instability for small quality variations but became definitely unstable for larger variations. This explanation was justified by the fact that the non-linear process gain would have had the effect of increasing the error between the process and process-model as quality increased. The system could therefore have been expected to go unstable after a large positive disturbance.

In conclusion, the application of the separation technique of identification produced enough quantitative information and operating

experience to allow for the extensive redesign of the control system.

5.4. Redesign of the control system

After consultations with the plant managers the following objectives of a redesigned control system were established,

- a) reduction of the magnitude of the control actions ΔB_t ,
- b) limitation of the control variable B_t to a desired range,
- c) reduction of the high-frequency disturbances entering the feedback loop,
- d) implementation of the updated process-model.

The first 3 objectives were all aimed at decreasing the possibility of the system becoming unstable and at maintaining the process in an approximately linear operating region. In particular, objective (b) was considered to be essential if the limits of the plant's mechanical performance were not to be exceeded.

The following modifications to the control system were therefore proposed:-

- (i) inclusion of a low-pass filter before the predictor,
- (ii) implementation of a reduced-order version of the updated process-model,
- (iii) implementation of a saturation characteristic to limit B_t ,
- (iv) a back-calculation procedure to reduce the non-linear effects of saturation.

It will be shown in the following that each of these modifications helps to achieve at least one of the stated redesign objectives.

Firstly, consider the effects of introducing a low-pass filter before the predictor. It was pointed out in section 5.1 that short-term high-frequency disturbances affect product quality from time to time - the frequencies of these disturbances were thought to be in the range $\frac{1}{2T}$ to $\frac{1}{6T}$. Due to the magnitude of the process time-delay, frequencies in that range cannot be controlled and if allowed to enter

the feedback loop could have an adverse effect on the control system's performance. It is therefore desirable to reduce the magnitude of such disturbances before they enter the feedback loop. A correctly designed low-pass filter acting on the signal $\{X_t\}$ can have precisely this effect; e.g. a discrete filter of the form,

$$F(z^{-1}) = \frac{a}{1 - b(z^{-1})}, \quad a = 1 - b$$

has a cut-off frequency at $\frac{1}{2T} \ln \frac{1}{b}$. If the cut-off frequency is chosen to be $\frac{1}{6T}$ then $b = e^{-1}$ and

$$F(z^{-1}) = \frac{0.63}{1 - 0.37z^{-1}} \quad (5.4.1)$$

It should be noted that, if $H(1) = 1$, the steady-state criterion of section 2.2 is satisfied only if the filter $F(z^{-1})$ also has a steady-state gain of unity. This filter may also help to reduce the magnitudes of the control actions ΔB_t although a more significant decrease can be achieved from modification (ii) above.

Consider the form of the controller in Run I; expanding eqn. (5.3.2) as a numerator polynomial,

$$W(z^{-1}) = 2.12 - 2.50z^{-1} + 1.30z^{-2} - 0.51z^{-3} + \dots$$

for a process-model gain of 2.0. It is noted that the first few coefficients of $W(z^{-1})$ are significantly greater than the controller gain of 0.5. Then it can be seen that the control actions

$$\Delta B_t = W(z^{-1}) [\Delta Q_t^d - z^{-1}H(z^{-1})\Delta X_t]$$

could be reduced in magnitude by decreasing the controller coefficients with respect to the steady-state gain.

Also consider the updated process-model impulse response,

$$z^{-1}\hat{P}(z^{-1}) = 0.26z^{-1}[1 + 1.6z^{-1} + 1.97z^{-2} + 1.70z^{-3} + \dots]$$

and its inverse

$$[\hat{P}(z^{-1})]^{-1} = 3.85 - 6.16z^{-1} + 0.23z^{-2} + \dots \quad (5.4.2)$$

where the controller gain is 0.37. As before the coefficients of

$[\hat{P}(z^{-1})]^{-1}$ are significantly greater than the steady-state gain. This effect is caused by the first sample of the process-model being small compared with subsequent samples - a feature typical of high-order transfer functions.

The updated process-model could be approximated by a reduced-order model with a correspondingly increased time-delay, e.g.

$$z^{-1}\hat{P}(z^{-1}) \approx z^{-1-1}M'(z^{-1}) = 0.68z^{-1-1} + 0.51z^{-1-2} + 0.44z^{-1-3} + \dots$$

$$\text{such that} \quad [M'(z^{-1})]^{-1} = 1.47 - 1.10z^{-1} - 0.13z^{-2} + \dots \quad (5.4.3)$$

In this case a reduction in order has been achieved by adding $\hat{p}_0$ to $\hat{p}_1$ to form the new first sample m'_0 . By comparing eqns.(5.4.2) and (5.4.3) it can be seen that a controller formed from $M'(z^{-1})$ will produce significantly smaller control actions than a controller formed from $\hat{P}(z^{-1})$.

It was noted at the end of the last section that both the process gain and time-delay had been mis-estimated by the empirical model. Consider first the pure time-delay l - this was calculated from the relationship (5.2.1) for the transportation lag k_t , and the measurement lag f_t . If the reduced-order model $M'(z^{-1})$ is to be used in the redesigned control system the updated time-delay l' should be one sample interval less than the Run I value; i.e.

$$\text{if} \quad l = \text{Integer}[k_t + 0.75] + 1 = 6$$

$$\text{then} \quad l' = \text{Integer}[k'_t + 0.75] + 1 = 5$$

From eqn.(5.2.1), $k_t = \alpha\beta_t y_t$ and if $k'_t = \alpha'\beta_t y_t$ then α' should be chosen

$$\text{in the range} \quad \frac{13}{17} \alpha < \alpha' < \frac{17}{21} \alpha \quad (5.4.4)$$

Since simulations suggested that the redesigned system was liable to go unstable if the process time-delay was underestimated (i.e. $l' < l$) it was decided to compromise by choosing $\alpha' = 0.45$.

Now consider the linearised process-model gain K calculated from eqn.(5.3.3). At $Q^d = 4$ units in Run I, K was taken as 2.0 whereas the updated process-model gain K' was found to be 2.8. Therefore the relationship (5.3.3) could be updated to

$$K' = 0.70 Q^d \quad (5.4.5)$$

The third system modification was concerned with the practical limits of the control variable B_t . There were two important factors involved:-

- (i) B_t could not exceed a specified value due to the limitation of the plant structure and obviously could never be negative,
- (ii) for values of B_t outside a specified range (which could vary) process behaviour was unpredictable.

It was therefore important to be able to restrict B_t to a given range at any time. This could be achieved by including a saturation characteristic in the redesigned control system, with the calculated B_t as an input and the implemented B_t as the output. Typically the saturation characteristic would be defined by,

for calculated $B_t > 7$ units, implemented $B_t = 7$ units
 " " $B_t \leq 1$ unit, " $B_t = 1$ unit
 " " $1 < B_t < 7$, implemented $B_t = \text{calculated } B_t$.

The inclusion of such a characteristic would however lead to a non-linear control system. At this point the investigation was continued by simulation; results showed that with the redesigned controller and filter included, only a large high-frequency disturbance (i.e. of type d_2) caused the input limits to be exceeded. It was also evident that the system ceased to function satisfactorily thereafter.

Since, as has already been discussed, large high-frequency disturbances cannot be controlled it was decided to further minimise their effects by limiting the output Q_t such that the calculated B_t could not exceed the given limits. This procedure will be referred to as back-calculation and can be described as follows:- let the measured quality value at time t be Q_t , and let the calculated value of the estimated disturbance be X_t . From Fig. 5.7 it can be shown that the calculated control action is given by

Fig. 5.10 THE REDESIGNED CONTROL SCHEME

$$\Delta B_t = W(z^{-1}) [\Delta Q_t^d - z^{-1}H(z^{-1})\Delta X_t] \quad (5.4.6)$$

and the absolute value of the control variable is

$$B_t = B_{t-1} + \Delta B_t$$

Then if this value of B_t lies outside the range $1 < B_t < 7$ a new control action $\Delta B'_t$ is defined by

$$B_t = B_{t-1} + \Delta B'_t = 1 \text{ or } 7 \text{ units}$$

It is noted that $\Delta B'_t$ is the largest control action that can be implemented, due to the saturation characteristic. Then the value Q'_t of quality which would have produced a control action $\Delta B'_t$ is calculated from

$$\Delta Q'_t = \frac{1}{z^{-1}H(z^{-1})W(z^{-1})} [W(z^{-1})\Delta Q_t^d - \{1-H(z^{-1})\}\Delta B'_t] \quad (5.4.7)$$

$$\text{and } Q'_t = Q_{t-1} + \Delta Q'_t$$

where eqn.(5.4.7) has been derived from eqn.(5.4.6). Back-calculation also produces a new value $\Delta X'_t$ defined by

$$\Delta X'_t = \frac{1}{z^{-1}H(z^{-1})W(z^{-1})} [W(z^{-1})\Delta Q_t^d - B'_t]$$

The result of this back-calculation procedure is that the control system can never demand a change in the input B_t which cannot be implemented. It has the advantage that the control system retains its linearity while rejecting extreme variations in quality.

A block-diagram of the redesigned control scheme is shown in Fig. 5.10. This system was used in a further set of digital simulations to check the performance of the redesigned elements. It was particularly noticeable that whenever a large disturbance caused the input B_t to saturate the back-calculation procedure was only required for a few samples; i.e. in most cases it was only necessary to replace one or two successive extreme values of Q_t in order to maintain B_t within the given limits.

It may also have been possible, at this stage, to include the

TABLE 5.4. SYSTEM PARAMETERS : RUN II

Desired quality value	Q_o^d	3.0 units of quality
Prevailing process input	B_o	3.6 units of B
Process transportation lag	k_o	6.4 sample intervals
Quality measurement lag	f_o	0.75 sample intervals
Process-model time-delay	l	8 sample intervals
Process-model gain	K	2.1 units of Q/B
Predictor pulse-transfer function	$z^{-1}H(z^{-1})$	1
Controller pulse-transfer function	$W(z^{-1})$	$1.54-1.16z^{-1}-0.11z^{-2}+0.10z^{-3}$

Fig.5.11 PROCESS RECORD OF EXPERIMENTAL RUN II

non-linear gain of eqn.(5.2.2) in the process-model and to design a predictor other than unity. It was thought, however, that further information was required before an attempt could be made to implement an optimum system.

The redesigned control system was next tested in a second Experimental Run - starting conditions and control system parameters are shown in Table 5.4. This time no restrictions were placed on the process operation so that quality set-point and process throughput could be changed whenever necessary.

A record of the start of Run II is shown in Fig. 5.11. It can be seen that the process throughput was changed twice where at each change the model time-delay was automatically updated - to 6 sample intervals at the 20th sample and to 5 sample intervals at the 70th sample. Also the quality set-point was changed at the 54th sample; it should be noted that this change was performed in two steps so that a very large control action was not required. The record shows a significant improvement over any of the previous records - i.e. Figs. 5.2, 5.3, and 5.8. It is particularly noticeable that the control actions were considerably smaller than the Run I values.

It should be noted, at this point, that the process records could not fairly be compared by means of a figure of merit. Any measure, such as quality variance, would be subject to considerable distortion from a large process upset.

The control system of Run II did not continue to function satisfactorily after the set-point changes at the 54th and 55th sample intervals. It can clearly be seen from Fig. 5.11 that the signals B_t and X_t oscillated after the set-point change. A digital simulation was sufficient to detect the cause of the oscillation - it was found that the simulated system behaved in a similar fashion when the process time-delay k was less than 1 and the gain K greater at $Q^d = 4.2$ than

TABLE 5.5. AUTOREGRESSIVE PARAMETERS : RUN II

	Number of Samples N			
	65	70	75	80
$\hat{r}_1$	0.18 $\pm$ 0.28	0.19 $\pm$ 0.26	0.19 $\pm$ 0.24	0.21 $\pm$ 0.24
$\hat{r}_2$	0.19 $\pm$ 0.28	0.18 $\pm$ 0.26	0.19 $\pm$ 0.26	0.18 $\pm$ 0.24
$\hat{r}_3$	0.04 $\pm$ 0.30	- 0.03 $\pm$ 0.28	- 0.06 $\pm$ 0.26	- 0.06 $\pm$ 0.26
$\hat{r}_4$	- 0.02 $\pm$ 0.30	- 0.03 $\pm$ 0.28	- 0.02 $\pm$ 0.26	- 0.01 $\pm$ 0.26
$\hat{r}_5$	- 0.21 $\pm$ 0.30	- 0.23 $\pm$ 0.28	- 0.22 $\pm$ 0.26	- 0.23 $\pm$ 0.26
$\hat{r}_6$	0.32 $\pm$ 0.28	0.34 $\pm$ 0.26	0.36 $\pm$ 0.26	0.35 $\pm$ 0.24
$\hat{r}_7$	0.02 $\pm$ 0.28	- 0.00 $\pm$ 0.28	0.01 $\pm$ 0.26	- 0.00 $\pm$ 0.24
$\hat{r}_8$	- 0.10 $\pm$ 0.28	- 0.13 $\pm$ 0.28	- 0.14 $\pm$ 0.26	- 0.12 $\pm$ 0.24
$\hat{r}_9$	- 0.01 $\pm$ 0.30	0.02 $\pm$ 0.28	0.00 $\pm$ 0.26	0.02 $\pm$ 0.26
$\hat{r}_{10}$	- 0.05 $\pm$ 0.28	- 0.09 $\pm$ 0.28	- 0.08 $\pm$ 0.26	- 0.10 $\pm$ 0.24
$\hat{r}_{11}$	0.02 $\pm$ 0.28	0.02 $\pm$ 0.26	0.02 $\pm$ 0.26	0.01 $\pm$ 0.24

at $Q^d = 3$ units. A substantial improvement was shown in the simulated system when the process-model time-delay was reduced. It was therefore decided to decrease the constant α in eqn.(5.2.1) still further to a value in the range of eqn.(5.4.4); the chosen value was, in fact, $\alpha' = 0.40$.

Run II was continued with the reduced model time-delay until a serious process upset caused a plant shutdown. This second part of Run II also showed a marked improvement over manual control although the control was again less effective at quality set-points significantly different from 3 units. It therefore appeared essential to compensate for the changes in process gain by the inclusion of the non-linear characteristic (5.4.5). Before proceeding with this further system modification, however, a second identification analysis was found to be possible; i.e. Run II had produced a set of data lasting 80 sample intervals where both the set-point and process throughput had remained constant. The separation technique was applied to this data using the computational procedure of section 5.3.

The autoregressive parameters $\hat{r}_i$ of the models for $\{\Delta X_t, t = 1, 2, \dots, N\}$, $N = 65, 70, 75, 80$ are shown in Table 5.5. Once again the relatively large estimation errors precluded the use of the reduction procedure of section 3.4. However the parameters $\hat{r}_3$ and $\hat{r}_4$ in Table 5.5 were sufficiently close to zero to justify the tentative assumption that separation occurred at the 2nd sample giving $m = 2$, $\beta = 5$. The updating equations (3.4.4) and (3.4.5) could then be written;

$$\hat{q}_i = \hat{r}_i, i = 1, 2$$

$$-\hat{R}_T(z^{-1}) \underline{\Omega} H(z^{-1}) [1 - \hat{Q}(z^{-1})][z^{1-k}P(z^{-1})W(z^{-1}) - 1] \quad (5.4.8)$$

In this case the process-model time-delay l was 6 sample intervals and $H(z^{-1})$ the combination of filter and predictor, i.e.

$$H(z^{-1}) = \frac{0.63 z^{-6}}{1 - 0.37z^{-1}}$$

TABLE 5.6. UPDATED PROCESS-MODEL PARAMETERS : RUN II

	Number of Samples N			
	65	70	75	80
$\hat{p}_0$	0.22	0.24	0.23	0.24
$\hat{p}_1$	0.44	0.43	0.41	0.43
$\hat{p}_2$	0.41	0.43	0.40	0.42
$\hat{p}_3$	0.36	0.38	0.36	0.36
$\hat{p}_4$	0.22	0.22	0.24	0.22
$\hat{p}_5$	0.20	0.25	0.25	0.25
$\hat{p}_6$	0.09	0.11	0.11	0.12

Fig.5.12 STEP RESPONSES OF UPDATED MODELS

Hence, from eqn.(5.4.8),

$$\hat{R}_T(z^{-1}) \underline{\Omega} 0.63z^{-6}[1-0.37z^{-1}]^{-1}\left[1-\sum_{i=1}^2 \hat{r}_i z^{-i}\right][z^{6-k}P(z^{-1})W(z^{-1})-1]$$

where $[\cdot]_T$ is the 5th-order truncation operator. It follows that

$$z^{-k}P(z^{-1}) = \frac{0.63[1-\hat{r}_1 z^{-1}-\hat{r}_2 z^{-2}]z^{-6}-\sum_{i=5}^{11} \hat{r}_i z^{-i}}{0.63[1-\hat{r}_1 z^{-1}-\hat{r}_2 z^{-2}]W(z^{-1})}$$

This equation has been used to calculate the updated process-model parameters $\hat{p}_i$, $i = 0, 1, \dots, 6$ for the 4 polynomials $\hat{R}(z^{-1})$ in Table 5.5. The results are shown in Table 5.6. As before the average of each parameter has been used to compute the updated process-model step response shown in Fig. 5.12 - also shown is the step response derived from the Run I updating analysis where the gain has been adjusted to be equal to the Run II value. The only significant difference between these two responses is a small time-delay - about 0.4 of a sample interval.

It should be noted that the estimation errors of the updated process-model parameters were considerably larger for Run II than for Run I. This can be established from the analysis of section 3.4, which concludes that the estimation errors are particularly dependent on the magnitude of $(h_0 w_0)^{-1}$. Since $(h_0 w_0)^{-1}$ was 0.47 in Run I as compared with 1.03 in Run II it could be expected that the estimation errors would be approximately twice as large in the second updating analysis - i.e. for the same length of data. For this reason the analysis of Run II was restricted to larger samples than those used in Run I - however the record was of insufficient length to compensate fully for the increased errors.

In conclusion the second identification analysis, although less reliable than the first, was sufficient to confirm the results of Run I and provided further information concerning the non-linear gain characteristic, i.e.

TABLE 5.7. SYSTEM PARAMETERS : RUN III

Desired quality value	Q_0^d	4.8 units of quality
Prevailing process input	B_0	5.5 units of B
Process transportation lag	k_0	6.4 sample intervals
Quality measurement lag	f_0	0.75 sample intervals
Process-model time-delay	l	8 sample intervals
Process-model gain	K	$0.6Q_t^d + 0.3$
Predictor pulse-transfer function	$z^{-l}H(z^{-1})$	1
Filter pulse-transfer function	$F(z^{-1})$	$\frac{0.63}{1 - 0.37z^{-1}}$
Controller pulse-transfer function	$W(z^{-1})$	$K[0.74 - 0.55z^{-1} - 0.05z^{-2} + 0.04z^{-3}]$

Fig.5.13 PROCESS RECORD OF EXPERIMENTAL RUN III

from Run I, at $Q^d = 4.2$, $K' = 2.8$

from Run II, at $Q^d = 3.0$, $K' = 2.1$

where K' refers to the gain of the updated process-model. On these two results the non-linear gain relation (5.4.5) could be further updated to

$$K = 0.6 Q^d + 0.3 \quad (5.4.9)$$

The next experimental phase of the investigation was therefore concerned with the implementation of the non-linear gain. After a further set of satisfactory simulation studies a third experimental run was started with the control system parameters shown in Table 5.7. A record of part of Run III is shown in Fig. 5.13. Unfortunately it was disrupted by a series of large process upsets - an example can be seen in Fig. 5.13, from the 20th sample onwards. However, by comparing the records of quality $\{Q_t\}$ and estimated disturbance $\{X_t\}$, it can be seen that the control system quickly reduced the effect of large process disturbances.

It has so far been shown that a significant reduction in quality variations was achieved by the implementation of a Box & Jenkins control system without an optimum predictor - in fact, an estimated 50% reduction in quality variance, or equivalently a 30% reduction in amplitude. The question then remained of the possibility of decreasing the variations still further by the use of an optimum predictor. Unfortunately it has had to go unanswered, since further system development has been prevented by circumstances outside the authors control. However, it is thought that the identification results from Runs I and II had produced sufficient information on which to base the development of a predictor. The design procedure might have been as follows:- consider the average of the several disturbance models obtained from the two analyses,

$$\text{Run I : } \Delta x_t = \frac{1}{1-0.30z^{-1}-0.26z^{-2}+0.16z^{-3}} \zeta_t, \text{ var}\{\zeta_t\} = \lambda = 2.8 \quad (5.4.10)$$

$$\text{Run II : } \Delta x_t = \frac{1}{1-0.19z^{-1}-0.18z^{-2}} \zeta_t, \text{ var}\{\zeta_t\} = \lambda = 2.2 \quad (5.4.11)$$

As might have been expected the two disturbance models of eqns. (5.4.10), (5.4.11) are distinctly different - although the estimation errors were such that no definite statement can be made concerning the time-variation of the disturbance dynamics. A predictor could be based on a disturbance model similar to both (5.4.10) and (5.4.11); for instance, consider the model

$$\Delta x_t = \frac{1}{1-0.25z^{-1}-0.18z^{-2}} \zeta_t, \text{ var}\{\zeta_t\} = \lambda = 2.5$$

$$\text{then } x_t = \frac{1}{(1-z^{-1})(1-0.25z^{-1}-0.18z^{-2})} \zeta_t \quad (5.4.12)$$

From the results of section 2.5 the optimum predictor is given by

$$H(z^{-1}) = \{ [(1-z^{-1})(1-0.25z^{-1}-0.18z^{-2})]^{-1} \}_T (1-z^{-1})(1-0.25z^{-1}-0.18z^{-2})$$

and the optimum k-step ahead prediction error by

$$e_r(k) = \left[1 + \sum_{i=1}^{k-1} s_i z^{-i} \right] \zeta_{r+k}$$

$$\text{where } 1 + S(z^{-1}) = 1 + \sum_{i=1}^{\infty} s_i z^{-i} = [(1-z^{-1})(1-0.25z^{-1}-0.18z^{-2})]^{-1}$$

$$\text{Also } \text{var}\{e_r(k)\} = \left[1 + \sum_{i=1}^{k-1} s_i^2 \right] \lambda$$

$$\text{For instance, } \text{var}\{e_r(4)\} = 7.31 \lambda = 18.3$$

$$\text{var}\{e_r(5)\} = 10.06 \lambda = 26.5$$

$$\text{var}\{e_r(6)\} = 12.95 \lambda = 32.4$$

Now $\text{var}\{X_t\}$ in Runs I and II was approximately 40 so that $\text{var}\{x_t\}$ could be expected to be in the range 25-35. For these values it is clear that little, if any, improvement can be achieved by predicting more than about 4 steps ahead. Furthermore if the disturbance model is, in fact, substantially different from (5.4.12) and if the

process-model is in error then the output error variance will be significantly greater than the optimum values shown above. It does not, therefore, appear possible in practice to achieve the extra 25% reduction in quality variance required to meet the estimate of the feasibility study. The initial estimate of a 75% reduction in quality variance was based on the type of disturbance shown in the uncontrolled variations of Fig. 5.2. However subsequent records showed the disturbance to be of substantially higher-frequency and therefore less easy to predict.

It cannot definitely be stated that prediction is of no practical value in this investigation. The above results were obtained from only two short records and it was thought that they contained less than the average amount of the slowly-varying drift component of disturbance - e.g. the record of Run III shows a more predominant drift. More representative results may have been obtained from a longer record of $\{X_t\}$ - i.e. about 200 samples. It may then have been possible to design a predictor, based on the drift component, which would have yielded a further significant improvement in quality control.

5.5. Summary and Conclusions

The investigation achieved its major objective of providing an automatic control system capable of significantly reducing quality variations. However, there remains some scope for development, particularly as far as the predictor is concerned.

Of equal importance to any quantitative results has been the experience gained in the identification and control of this type of process. Firstly, the separation technique has proved to be a practically viable method of simultaneous identification and control. Consistent results were obtained even though the updating analyses had to be based on relatively short sets of data.

Secondly, it has been shown that it is possible to design a workable control system on the basis of readily available empirical

data. In retrospect it would have been advisable to have included a low-pass filter in the Run I control scheme - simulation studies confirmed that the system would then have been stable.

Thirdly, the results indicate that the major benefit from automatic control comes from the deterministic part of the system rather than from prediction of the random disturbances. It is evident that an optimum linear predictor, as defined in section 2.5, has limited application to this type of process - even though the disturbances are of particularly large magnitude.

Finally, this investigation has suggested several improvements that are necessary to make methods of simultaneous identification and control more readily applicable; e.g.

- (i) the removal of the restriction that a record of $\{X_t\}$ must be obtained when both the set-point and the process time-delay are constant;
- (ii) extension to processes with any magnitude of time-delay - in effect this means an alternative approach to separation.

An attempt will be made in the next chapter to formulate and solve some of the associated theoretical problems.

CHAPTER VIFURTHER METHODS OF SIMULTANEOUS IDENTIFICATION AND CONTROLIntroduction

The methods of simultaneous identification and control developed so far all suffer from restrictions on the type of system to which they can be applied. In particular, the proposed methods of process-model and predictor updating require a stationary disturbance and the separation technique only applies to processes with large time-delays. Also, they all can only estimate the impulse response of the process and an autoregressive model for the disturbance.

This chapter will attempt to extend the previous theories to a more general class of problems. The first section will present a generalisation of the updating equation by the use of Astrom's representation of linear dynamic systems⁷. It will then be shown, in section 6.2, that the separation effect can be employed to estimate mixed models for both the disturbance and the process, given similar conditions to those of section 3.4. In sections 6.3 and 6.4 the generalised updating equation will be used to form a simultaneous identification and control algorithm where no condition is placed on the magnitude of the process time-delay. Finally section 6.5 will discuss some aspects of simultaneous identification and control which still present difficulties.

6.1. Generalisation of the updating equation

Consider the representation of a linear dynamical system proposed by Astrom^{7,8} :- let the noise-free process be described by

$$y_t = \frac{B'(z^{-1})}{1 + A'(z^{-1})} c_{t-k} \quad (6.1.1)$$

and the disturbance by

$$x_t = \frac{1 + S'(z^{-1})}{1 - Q'(z^{-1})} \zeta_t \quad (6.1.2)$$

Fig.6.1. THE TWO MODELS OF A LINEAR SYSTEM

Fig.6.2. THE MIXED-MODEL CONTROL SCHEME

where $\{\zeta_t\}$ is a sequence of independent random variables with a zero mean and variance λ . The polynomials are given by

$$\left. \begin{aligned} A(z^{-1}) &= a'_1 z^{-1} + a'_2 z^{-2} + \dots + a'_n z^{-n} \\ B(z^{-1}) &= b'_0 + b'_1 z^{-1} + \dots + b'_n z^{-n} \\ S(z^{-1}) &= s'_1 z^{-1} + s'_2 z^{-2} + \dots + s'_n z^{-n} \\ Q(z^{-1}) &= q'_1 z^{-1} + q'_2 z^{-2} + \dots + q'_n z^{-n} \end{aligned} \right\} \quad (6.1.3)$$

where $b'_0 \neq 0$, but any of the coefficients a'_i, b'_i, s'_i, q'_i , $i = 1, \dots, n$ can be zero.

Thus the measurable system output $\{E_t\}$ is given by

$$E_t = \frac{B(z^{-1})}{1 + A(z^{-1})} c_{t-k} + \frac{1 + S(z^{-1})}{1 - Q(z^{-1})} \zeta_t \quad (6.1.4)$$

or rearranging

$$[1 + A(z^{-1})] E_t = B(z^{-1}) c_{t-k} + [1 + S(z^{-1})] \zeta_t \quad (6.1.5)$$

$$\left. \begin{aligned} \text{where} \quad 1 + A(z^{-1}) &= [1 + A'(z^{-1})][1 - Q(z^{-1})] \\ B(z^{-1}) &= B'(z^{-1})[1 - Q(z^{-1})] \\ 1 + S(z^{-1}) &= [1 + S'(z^{-1})][1 + A'(z^{-1})] \end{aligned} \right\} \quad (6.1.6)$$

and where the polynomials on the left-hand side of eqns.(6.1.6) are of degree $(2n+1)$. The two equivalent representations (6.1.4) and (6.1.5) are shown in Fig.6.1.

Now consider the noisy process (6.1.5) in the Box & Jenkins feedback control system of Fig.6.2. Let the process-model be represented by the pulse-transfer function

$$\frac{z^{-l} \nu(1 + \beta(z^{-1}))}{1 + \alpha(z^{-1})} \quad (6.1.7)$$

where l is the number of samples in the pure time-delay, and

$$\left. \begin{aligned} \beta(z^{-1}) &= \beta_1 z^{-1} + \beta_2 z^{-2} + \dots + \beta_m z^{-m} \\ \alpha(z^{-1}) &= \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots + \alpha_m z^{-m} \end{aligned} \right\} \quad (6.1.8)$$

Also, let the predictor be represented by the pulse transfer function

$$\frac{z^l \mu(z^{-1})}{1 + \eta(z^{-1})} \quad (6.1.9)$$

where

$$\left. \begin{aligned} \mu(z^{-1}) &= \mu_0 z^{-1} + \mu_1 z^{-1-1} + \dots + \mu_m z^{-1-m} \\ \eta(z^{-1}) &= \eta_1 z^{-1} + \eta_2 z^{-2} + \dots + \eta_m z^{-m} \end{aligned} \right\} \quad (6.1.10)$$

and where any of the coefficients $\beta_1, \beta_2, \dots, \beta_m, \alpha_1, \dots, \alpha_m, \mu_1, \dots, \mu_m, \eta_1, \dots, \eta_m$ can be zero.

It can then be shown that the estimated disturbance $\{x_t\}$ is given by

$$x_t = \frac{[1+\beta(z^{-1})][1+\eta(z^{-1})][1+\Lambda(z^{-1})]}{[1+\beta(z^{-1})][1+\Lambda(z^{-1})][1+\eta(z^{-1})-\mu(z^{-1})]+z^j B(z^{-1})\mu(z^{-1})v^{-1}[1+\alpha(z^{-1})]} \cdot x_t \quad (6.1.11)$$

where $j = 1 - k$. But from eqns. (6.1.2), (6.1.6)

$$x_t = \frac{1 + S(z^{-1})}{1 + \Lambda(z^{-1})} \zeta_t \quad (6.1.12)$$

and substituting for $\{x_t\}$ in eqn. (6.1.11) gives

$$x_t = \frac{[1+\beta(z^{-1})][1+\eta(z^{-1})][1+S(z^{-1})]}{[1+\beta(z^{-1})][1+\Lambda(z^{-1})][1+\eta(z^{-1})-\mu(z^{-1})]+z^j B(z^{-1})\mu(z^{-1})v^{-1}[1+\alpha(z^{-1})]} \zeta_t \quad (6.1.13)$$

It is noted that the numerator of eqn. (6.1.13) is a $(2m+n+1)^{\text{th}}$ -degree polynomial in z^{-1} and the denominator is a $(2m+n+j+1)^{\text{th}}$ -degree polynomial.

Now if the process is operated using the Box & Jenkins control system of Fig. 6.2 over the interval $0 < t \leq N$ a set of data $\{X_t, t = 1, 2, \dots, N\}$ can be recorded. This data can then be used to estimate a mixed model for $\{X_t\}$ of the form

$$X_t = \frac{1 + \hat{U}(z^{-1})}{1 + \hat{R}(z^{-1})} a_t \quad (6.1.14)$$

$$\text{where } \hat{U}(z^{-1}) = \sum_{i=1}^I \hat{u}_i z^{-i}, \quad \hat{R}(z^{-1}) = \sum_{i=1}^J \hat{r}_i z^{-i}$$

and where $\{a_t\}$ is a sequence of residuals corresponding to an estimate of $\{\zeta_t\}$. The parameters $\hat{u}_i, \hat{r}_i$ can be computed by the maximum likelihood method discussed in section 2.4. Then, since the numerator and denominator polynomials of eqn. (6.1.13) will in general have no common

factors, eqns. (6.1.13) and (6.1.14) can be equated to give

$$1 + \hat{U}(z^{-1}) \underline{\Omega} [1+\beta(z^{-1})][1+\eta(z^{-1})][1+\alpha(z^{-1})] \quad (6.1.15)$$

$$1 + \hat{R}(z^{-1}) \underline{\Omega} [1+\beta(z^{-1})][1+A(z^{-1})][1+\eta(z^{-1})-\mu(z^{-1})] \\ + z^j B(z^{-1}) \mu(z^{-1}) v^{-1} [1+\alpha(z^{-1})] \quad (6.1.16)$$

These two equations are, therefore, generalisations of the updating equation which was developed in section 3.1. where the process and disturbance models were restricted to an impulse-response and an autoregressive model respectively.

6.2. The separation effect for mixed models.

Consider eqn. (6.1.16) and rewrite it as follows

$$1 + \hat{R}(z^{-1}) \underline{\Omega} [1+\beta(z^{-1})][1+A(z^{-1})][1+\eta(z^{-1})] \\ + \mu(z^{-1}) \{ z^j B(z^{-1}) v^{-1} [1+\alpha(z^{-1})] - [1+\beta(z^{-1})][1+A(z^{-1})] \} \quad (6.2.1)$$

From eqn. (6.1.10) it can be seen that $\mu(z^{-1})$ contains coefficients of z^{-i} , $i > 1$ only. Thus the second term on the right-hand side of eqn. (6.2.1) contains coefficients of z^{-i} , $i > \min(1, k)$ only. Then if the degree of the first term on the right-hand side of eqn. (6.2.1) is less than $\min(1, k)$ the two terms separate. That is, if $2(m+n)+1 < \min(1, k)$

$$1 + \sum_{i=1}^{2(m+n)+1} \hat{r}_i z^{-i} \underline{\Omega} [1+\beta(z^{-1})][1+A(z^{-1})][1+\eta(z^{-1})] \quad (6.2.2)$$

$$\text{and } \hat{R}_T(z^{-1}) \underline{\Omega} \mu(z^{-1}) \{ z^j B(z^{-1}) v^{-1} [1+\alpha(z^{-1})] - [1+\beta(z^{-1})][1+A(z^{-1})] \} \quad (6.2.3)$$

Since $\beta(z^{-1})$, $\eta(z^{-1})$, $\alpha(z^{-1})$ and v are known the polynomial $A(z^{-1})$ can be estimated from eqn. (6.2.2), and then $z^j B(z^{-1})$ can be estimated from eqn. (6.2.3). That is,

$$1 + \hat{A}(z^{-1}) = \frac{1 + \sum_{i=1}^{2(m+n)+1} \hat{r}_i z^{-i}}{[1+\beta(z^{-1})][1+\eta(z^{-1})]} \quad (6.2.4)$$

$$\text{and } \hat{z}^j \hat{B}(z^{-1}) = \frac{\mu(z^{-1}) [1 + \beta(z^{-1})] [1 + \hat{A}(z^{-1})] + \hat{R}_T(z^{-1})}{v^{-1} [1 + \alpha(z^{-1})]} \quad (6.2.5)$$

Also from eqn.(6.1.15) $S(z^{-1})$ can be estimated by

$$\hat{S}(z^{-1}) = \frac{1 + \hat{U}(z^{-1})}{[1 + \beta(z^{-1})] [1 + \eta(z^{-1})]} \quad (6.2.6)$$

The three equations (6.2.4), (6.2.5), (6.2.6) give estimates of all the parameters of the dynamic model of the process (6.1.5). Then the control system can be updated by replacing the process-model and controller in Fig.6.2 by the estimated process $\hat{z}^{-k} \hat{B}(z^{-1}) [1 + \hat{A}(z^{-1})]^{-1}$ and its inverse respectively. Also an optimum predictor can be calculated from eqn.(2.5.10) as

$$\hat{H}(z^{-1}) = \{ [1 + \hat{S}(z^{-1})] [1 + \hat{A}(z^{-1})]^{-1} \}_T \{ [1 + \hat{S}(z^{-1})]^{-1} [1 + \hat{A}(z^{-1})] \} \quad (6.2.7)$$

Unfortunately the condition that $2(m+n)+1 < \min(1,k)$ restricts the application of this technique to processes with very large time-delays. For instance if $\hat{A}(z^{-1})$, $\hat{Q}(z^{-1})$, $\beta(z^{-1})$, $\eta(z^{-1})$ are all 2nd-degree polynomials in z^{-1} then $2(m+n)+1 = 5$ so that the process time-delay must be at least 6 sample intervals if separation is to occur.

The only advantages that the above technique has over the simpler method of section 3.4 are as follows:-

- (i) the polynomial $\hat{R}(z^{-1})$ need not be truncated on the right since its degree is known to be finite;
- (ii) the estimated process and disturbance models are in a more compact form than could be expected with an impulse-response and autoregressive model - i.e. the system is described by fewer parameters.

These advantages appear to have little significance in practice and are clearly outweighed by the following disadvantages:-

- (i) the orders I, J of the moving-average and autoregressive polynomials in eqn.(6.1.14) will be relatively large; this presents

many problems in the computation of the parameters u_i, r_i , since the maximum likelihood method uses a hill-climbing routine in a parameter space of $(I+J)$ dimensions.

- (ii) the restriction that $2(m+1)+1 < \min(1,k)$ is more severe than its equivalent in the method of section 3.4.

6.3. A general method of simultaneous identification and control

This section will exploit the particular characteristics of the second updating equation (6.1.16) to produce an identification algorithm which applies to processes with any magnitude of time-delay.

Firstly, it will be convenient to represent the noise-free process by

$$y_t = \frac{B(z^{-1})}{1 + A(z^{-1})} c_t$$

where
$$B(z^{-1}) = b_1 z^{-1} + b_2 z^{-2} + \dots + b_{2n} z^{-n} \quad (6.3.1)$$

Then if the process has a pure time-delay of k sample intervals the coefficients $b_1, b_2, \dots, b_{k-1}$ will be zero and b_k will be non-zero. The updating equations (6.1.15) and (6.1.16) can then be rewritten as

$$1 + \hat{U}(z^{-1}) \underline{\Omega} [1 + \beta(z^{-1})] [1 + \eta(z^{-1})] [1 + S(z^{-1})] \quad (6.3.2)$$

and
$$1 + \hat{R}(z^{-1}) \underline{\Omega} [1 + \beta(z^{-1})] [1 + A(z^{-1})] [1 + \eta(z^{-1}) - \mu(z^{-1})] + z^{-1} B(z^{-1}) \mu(z^{-1}) \nu^{-1} [1 + \alpha(z^{-1})] \quad (6.3.3)$$

As before the parameters $\hat{s}_i, i = 1, \dots, I$ of $\hat{S}(z^{-1})$ can be estimated from eqn.(6.3.2) since all the coefficients $\hat{u}_i, \beta_i, \eta_i$ are known - i.e.

$$\hat{S}(z^{-1}) = \frac{1 + \hat{U}(z^{-1})}{[1 + \beta(z^{-1})] [1 + \eta(z^{-1})]} \quad (6.3.4)$$

Now consider eqn.(6.3.3) and expand the right-hand side,

$$1 + \sum_{i=1}^J \hat{r}_i z^{-i} \underline{\Omega} \left\{ 1 + \sum_{i=1}^m \beta_i z^{-i} \right\} \left\{ 1 + \sum_{i=1}^{2n} a_i z^{-i} \right\} \left\{ 1 + \sum_{i=1}^m \eta_i z^{-i} - \sum_{i=0}^m \mu_i z^{-1-i} \right\} + z^{-1} \left\{ \sum_{i=1}^{2n} b_i z^{-i} \right\} \left\{ \sum_{i=0}^m \mu_i z^{-1-i} \right\} \nu^{-1} \left\{ 1 + \sum_{i=1}^m \alpha_i z^{-i} \right\} \quad (6.3.5)$$

$$\text{Define } V(z^{-1}) = \sum_{i=0}^M v_i z^{-i} = \left\{ 1 + \sum_{i=1}^m \beta_i z^{-i} \right\} \left\{ 1 + \sum_{i=1}^m \eta_i z^{-i} - \sum_{i=0}^m \mu_i z^{-1-i} \right\} \quad (6.3.6)$$

$$\text{and } G(z^{-1}) = \sum_{i=0}^M g_i z^{-i} = v^{-1} \left\{ \sum_{i=0}^m \mu_i z^{-i} \right\} \left\{ 1 + \sum_{i=1}^m \alpha_i z^{-i} \right\} \quad (6.3.7)$$

where $v_0 = 1$ and $M = 2m+1$. The choice of M as the degree of both polynomials $V(z^{-1})$ and $G(z^{-1})$ is a matter of convenience - there is no loss of generality since any of the coefficients g_i can be chosen to be zero. Also if the polynomials $A(z^{-1})$ and $B(z^{-1})$ are extended to M terms, eqns.(6.3.5) can be written as

$$1 + \sum_{i=1}^J \hat{r}_i z^{-i} - \frac{\Omega}{\sum_{i=0}^M a_i z^{-i} \sum_{j=0}^M v_j z^{-j}} + \frac{\sum_{i=0}^M b_i z^{-i} \sum_{j=0}^M g_j z^{-j}}{\sum_{i=0}^M a_i z^{-i} \sum_{j=0}^M v_j z^{-j}} \quad (6.3.8)$$

where $a_0 = 1$ and $b_0 = 0$. Eqn.(6.3.8) can be expanded in powers of z^{-1} as follows -

$$\begin{aligned} 1 + \sum_{i=1}^J \hat{r}_i z^{-i} - \frac{\Omega}{\sum_{i=0}^M \sum_{j=0}^i a_j v_{i-j} z^{-i}} + \frac{\sum_{i=1}^M \sum_{j=i}^M a_j v_{M+i-j} z^{-i-M}}{\sum_{i=0}^M \sum_{j=0}^i a_j v_{i-j} z^{-i}} \\ + \frac{\sum_{i=0}^M \sum_{j=0}^i b_j g_{i-j} z^{-i}}{\sum_{i=1}^M \sum_{j=1}^M b_j g_{M+i-j} z^{-i-M}} \end{aligned}$$

Thus if the parameters $\hat{a}_i, b_i$ are replaced by their estimates $\hat{a}_i, \hat{b}_i$,

$$\hat{r}_i = \sum_{j=0}^i \{ \hat{a}_j v_{i-j} + \hat{b}_j g_{i-j} \} \quad \text{for } i = 1, 2, \dots, M \quad (6.3.9)$$

$$\text{and } \hat{r}_{i+M} = \sum_{j=i}^M \{ \hat{a}_j v_{M+i-j} + \hat{b}_j g_{M+i-j} \} \quad \text{for } i = 1, 2, \dots, M \quad (6.3.10)$$

Since the coefficients of $V(z^{-1})$ and $G(z^{-1})$ can be calculated from the known parameters of the control system, and since the coefficients of $\hat{R}(z^{-1})$ can be calculated from a record of $\{x_t\}$, eqns.(6.3.9) and (6.3.10) can be used to compute estimates $\hat{A}(z^{-1}), \hat{B}(z^{-1})$ of the process dynamics.

The control system can then be updated by replacing the process-model by $\hat{B}(z^{-1})[1+\hat{A}(z^{-1})]^{-1}$ and calculating an optimum predictor from the disturbance model $[1+\hat{S}(z^{-1})][1+\hat{A}(z^{-1})]^{-1}$.

Computation of the estimates $\hat{a}_i, \hat{b}_i$ is simply a matter of solving a set of simultaneous linear algebraic equations; this can clearly be seen by writing eqns.(6.3.9) and (6.3.10) in matrix form as

$$\underline{\hat{r}} = \underline{\Psi} \underline{\hat{\theta}} + \underline{v} \quad (6.3.11)$$

where

$$\underline{\hat{r}} = (\hat{r}_1 \ \hat{r}_2 \ \cdot \ \cdot \ \hat{r}_M \ \hat{r}_{M+1} \ \cdot \ \cdot \ \hat{r}_{i+M})^T$$

$$\underline{\hat{\theta}} = (\hat{a}_1 \ \hat{a}_2 \ \cdot \ \cdot \ \hat{a}_M \ \hat{b}_1 \ \cdot \ \cdot \ \hat{b}_M)^T$$

$$\underline{v} = (v_1 \ v_2 \ \cdot \ \cdot \ v_M \ 0 \ \cdot \ \cdot \ 0)^T$$

and
 $\underline{\Psi} =$

1	0	0	·	·	·	0	g_0	0	0	·	·	·	0
v_1	1	0	·	·	·	0	g_1	g_0	0	·	·	·	0
v_2	v_1	1	·	·	·	0	g_2	g_1	g_0	·	·	·	0
·	·						·	·					·
·	·						·	·					·
·	·						·	·					·
v_{M-2}	v_{M-3}	·	·	·	1	0	g_{M-2}	g_{M-3}	·	·	·	g_0	0
v_{M-1}	v_{M-2}	·	·	·	v_1	1	g_{M-1}	g_{M-2}	·	·	·	g_1	g_0
v_M	v_{M-1}	·	·	·	v_2	v_1	g_M	g_{M-1}	·	·	·	g_2	g_1
0	v_M	·	·	·	v_3	v_2	0	g_M	·	·	·	g_3	g_2
·	·				·	·	·	·				·	·
·	·				·	·	·	·				·	·
·	·				·	·	·	·				·	·
					v_M	v_{M-1}						g_M	g_{M-1}
0	0	·	·	·	0	v_M	0	0	·	·	·	0	g_M

Thus the parameter vector $\underline{\hat{\theta}}$ can be calculated from

$$\underline{\hat{\theta}} = \underline{\Psi}^{-1} (\underline{\hat{r}} - \underline{v}) \quad (6.3.12)$$

It should be noted that the matrix $\underline{\Psi}$ is square and, in general, non-singular when the degree of the polynomials $\hat{A}(z^{-1}), \hat{B}(z^{-1})$ is the

same as the degree of the polynomials $V(z^{-1})$ and $G(z^{-1})$. However in the general case the matrix $\underline{\Psi}$ will only be square and non-singular under certain conditions. Let

(i) the polynomial $V(z^{-1})$ contain M_1 non-zero coefficients $v_1, v_2, \dots, v_{M_1}$;

(ii) the polynomial $G(z^{-1})$ contain M_2 non-zero coefficients $g_1, g_2, \dots, g_{M_2}$;

(iii) the polynomial $\hat{R}(z^{-1})$ contain J coefficients $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_J$.

Further let it be required to estimate n_1 parameters $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{n_1}$ and n_2 parameters $\hat{b}_1, \hat{b}_2, \dots, \hat{b}_{n_2}$. Then a necessary condition for the matrix

$\underline{\Psi}$ to be square is that $J = n_1 + n_2$. But it can also be shown that

$$J = \max \{ (n_1 + M_1), (n_2 + M_2) \}$$

Thus if $J = (n_1 + M_1)$, it follows that $n_2 = M_1$

and if $J = (n_2 + M_2)$, it follows that $n_1 = M_2$.

That is, a second necessary condition for $\underline{\Psi}$ to be square and non-singular is that either $n_1 = M_2$ or $n_2 = M_1$.

Example 6.3.1

$$\text{let } V(z^{-1}) = 1 + v_1 z^{-1}, \text{ i.e. } M_1 = 1$$

$$\text{and } G(z^{-1}) = g_0 + g_1 z^{-1} + g_2 z^{-2}, \text{ i.e. } M_2 = 2$$

$$\text{Also let } \hat{R}(z^{-1}) = \hat{r}_1 z^{-1} + \hat{r}_2 z^{-2} + \hat{r}_3 z^{-3} + \hat{r}_4 z^{-4}, \text{ i.e. } J = 4.$$

Thus we can either estimate $\hat{b}_1$ ($n_2 = M_1 = 1$), $\hat{a}_1, \hat{a}_2, \dots$ or $\hat{a}_1, \hat{a}_2$ ($n_1 = M_2 = 2$), $\hat{b}_1, \hat{b}_2, \dots$. For instance if we choose to estimate $\hat{b}_1, \hat{a}_1, \hat{a}_2, \hat{a}_3$ then the matrix equation (6.3.11) becomes

$$\begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \hat{r}_3 \\ \hat{r}_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & | & g_0 \\ v_1 & 1 & 0 & | & g_1 \\ 0 & v_1 & 1 & | & g_2 \\ 0 & 0 & v_1 & | & 0 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{b}_1 \end{bmatrix} + \begin{bmatrix} v_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In general the matrix $\underline{\Psi}$ can be partitioned as follows:-

$$\begin{bmatrix} (n_1 \times n_1) & (n_1 \times n_2) \\ (n_2 \times n_1) & (n_2 \times n_2) \end{bmatrix}$$

In some cases it may be possible to calculate more parameters $\hat{a}_i, \hat{b}_i$ than above by allowing $\underline{\Psi}$ to be rectangular or singular. The inverse of $\underline{\Psi}$ in (6.3.12) would then be replaced by the psuedo-inverse³¹.

It should also be noted that the required estimates of $\hat{A}(z^{-1})$ and $\hat{B}(z^{-1})$ will only be obtained if all the significant parameters $\hat{a}_i, \hat{b}_i$ are included in the matrix equation (6.3.11). Thus the computational procedure could be as follows:-

- (i) assume n_1 non-zero parameters $a_i, i = 1, 2, \dots, n_1$ and n_2 non-zero parameters $b_i, i = 1, 2, \dots, n_2$ where n_1, n_2 are the largest expected values of the degrees of the polynomials $A(z^{-1})$ and $B(z^{-1})$ respectively;
- (ii) calculate $J = n_1 + n_2$ estimates $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_J$;
- (iii) form the matrix equation (6.3.11) and solve for $\underline{\theta}$ using eqn.(6.3.12);
- (iv) if the estimates $\hat{a}_{n_1}$ and/or $\hat{b}_{n_2}$ are insignificantly different from zero reform the matrix equation with $n'_1 = n_1 - 1$ and/or $n'_2 = n_2 - 1$ and re-solve for $\underline{\hat{\theta}}$;
- (v) continue reducing the values of n_1 and n_2 until all the estimated parameters are significantly different from zero.

Example 6.3.2

Consider the initial control system which has a process-model

$$\text{given by } \frac{v(1 + \beta(z^{-1}))}{1 + \alpha(z^{-1})} = \frac{1}{1 - 0.5z^{-1}}$$

$$\text{and a predictor } \frac{\mu(z^{-1})}{1 + \eta(z^{-1})} = \frac{0.8z^{-1}}{1 - 0.2z^{-1}}$$

Then from eqns.(6.3.6) and (6.3.7),

$$\begin{aligned} V(z^{-1}) &= 1 - z^{-1} \\ G(z^{-1}) &= 0.8 - 0.4z^{-1} \end{aligned}$$

Let the estimate of a third-order autoregressive polynomial $R(z^{-1})$ be

$$\hat{R}(z^{-1}) = -1.45z^{-1} + 0.65z^{-2} - 0.05z^{-3}$$

Thus if we choose to estimate $\hat{b}_1, \hat{a}_1, \hat{a}_2$ the matrix equation (6.3.11)

$$\text{becomes } \begin{bmatrix} -1.45 \\ 0.65 \\ -0.05 \end{bmatrix} = \begin{bmatrix} 1.0 & 0 & 0.8 \\ -1.0 & 1.0 & -0.4 \\ 0 & -1.0 & 0 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{b}_1 \end{bmatrix} + \begin{bmatrix} -1.0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{giving } \underline{\hat{\theta}} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} -0.75 \\ 0.05 \\ 0.375 \end{bmatrix}$$

Then if $\hat{a}_2$ is assumed to be insignificantly different from zero the matrix equation can be reformed as

$$\begin{bmatrix} -1.45 \\ 0.65 \end{bmatrix} = \begin{bmatrix} 1.0 & 0.8 \\ -1.0 & -0.4 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} + \begin{bmatrix} -1.0 \\ 0 \end{bmatrix}$$

$$\text{giving } \underline{\hat{\theta}} = \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} -0.85 \\ 0.50 \end{bmatrix}$$

It is advantageous to choose the initial control system such that the polynomials $V(z^{-1})$ and $G(z^{-1})$ contains as many parameters as possible. The restrictions on the choice of n_1 and n_2 are then less severe and it may even arise that more equations of the form (6.3.9), (6.3.10) are available than the number of parameter estimates required. That is the matrix equation could be

$$\underline{\hat{r}} = \underline{\Psi} \underline{\hat{\theta}} + \underline{v}$$

where $\underline{\hat{r}}$ and $\underline{v}$ are J -vectors, $\underline{\hat{\theta}}$ is an $(n_1 + n_2)$ -vector, $\underline{\Psi}$ is a $J \times (n_1 + n_2)$ matrix such that $J > n_1 + n_2$ and $\underline{\Psi}$ contains no row or column with all zero elements. The estimate $\underline{\hat{\theta}}$ could then be obtained from the regression equation

$$\hat{\underline{\theta}} = (\underline{\psi}^T \underline{\psi})^{-1} \underline{\psi}^T (\hat{\underline{r}} - \underline{v}) \quad (6.3.13)$$

Example 6.3.3

Consider an initial control system with a process-model

given by
$$\frac{v(1+\beta(z^{-1}))}{1+\alpha(z^{-1})} = \frac{1}{1-0.5z^{-1}}$$

and a predictor
$$\frac{\mu(z^{-1})}{1+\eta(z^{-1})} = \frac{0.4z^{-1} + 0.4z^{-2}}{1-0.2z^{-1}}$$

Thus
$$V(z^{-1}) = 1 - 0.6z^{-1} - 0.4z^{-2}$$

$$G(z^{-1}) = 0.4 + 0.2z^{-1} - 0.2z^{-2}$$

Let the estimated $R(z^{-1})$ be

$$\hat{R}(z^{-1}) = -1.2z^{-1} + 0.2z^{-1} + 0.2z^{-2}$$

n.b. this $\hat{R}(z^{-1})$ was calculated from eqn.(6.3.5) where the process was given by

$$\frac{B(z^{-1})}{1+A(z^{-1})} = \frac{0.5z^{-1}}{1-0.8z^{-1}}$$

and small estimation errors have been introduced - i.e. the true polynomial $R(z^{-1})$ is given by

$$R(z^{-1}) = -1.2z^{-1} + 0.18z^{-2} - 0.22z^{-3}$$

If it is chosen to estimate $\hat{a}_1$ and $\hat{b}_1$ only, the matrix equation becomes

$$\begin{bmatrix} -1.2 \\ 0.2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ -0.6 & 0.2 \\ -0.4 & -0.2 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} + \begin{bmatrix} -0.6 \\ -0.4 \\ 0 \end{bmatrix}$$

Hence the estimates $\hat{a}_1, \hat{b}_1$ calculated from the regression equation (6.3.13) are

$$\hat{\underline{\theta}} = \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} -0.82 \\ 0.56 \end{bmatrix}$$

such that the updated process-model is given by

$$\frac{\hat{B}(z^{-1})}{1 + \hat{A}(z^{-1})} = \frac{0.56z^{-1}}{1 - 0.82z^{-1}}$$

At present there are two major problems in the application of this generalised updating technique. Firstly the method requires a reliable estimation procedure for mixed moving-average-autoregressive models; i.e. to calculate $\hat{U}(z^{-1})$ and $\hat{R}(z^{-1})$ in eqn.(6.1.14). In practice it can be expected that at least 2 moving-average and 3 autoregressive parameters will be required. Unfortunately the available estimation algorithms - e.g. Astrom's maximum likelihood method - are difficult to apply when a large number of parameters is involved. In particular, techniques relying on a hill-climbing procedure often produce misleading answers due to the appearance of multiple peaks in the cost function due to an insufficiency of data³².

The second problem concerns the sensitivity of the updated parameters to estimation errors in $\hat{R}(z^{-1})$.

6.4. Sensitivity of the general method to estimation errors

Consider the matrix equation (6.3.11).

$$\hat{\underline{r}} = \underline{\Psi} \hat{\underline{\theta}} + \underline{v} \quad (6.4.1)$$

and its equivalent if no estimation errors are present,

$$\underline{r} = \underline{\Psi} \underline{\theta} + \underline{v} \quad (6.4.2)$$

Define an error vector $\underline{\delta}$ as the difference between the estimate $\hat{\underline{r}}$ and its true value $\underline{r}$,

$$\underline{\delta} = \hat{\underline{r}} - \underline{r} \quad (6.4.3)$$

Then from eqns.(6.4.1), (6.4.2),

$$\hat{\underline{r}} - \underline{r} = \underline{\Psi} (\hat{\underline{\theta}} - \underline{\theta})$$

$$\text{hence} \quad \hat{\underline{\theta}} - \underline{\theta} = \underline{\Psi}^{-1} \underline{\delta} \quad (6.4.4)$$

which is an expression for the error in the estimated parameter vector.

In particular $\epsilon\{\hat{\underline{\theta}} - \underline{\theta}\} = 0$ since $\epsilon\{\underline{\delta}\} = 0$ implying that $\hat{\underline{\theta}}$ is an

Fig.6.3. SOLUTION TO EXAMPLE 6.4.1.

asymptotically unbiased estimate of $\underline{\theta}$. Also if the covariance matrix of $\underline{\hat{\theta}}$ is $\epsilon\{\underline{\hat{\theta}} - \underline{\theta}\{\underline{\hat{\theta}} - \underline{\theta}\}^T = \underline{\Phi}$ then

$$\text{var}\{\underline{\hat{\theta}}\} = \epsilon\{\underline{\hat{\theta}} - \underline{\theta}\{\underline{\hat{\theta}} - \underline{\theta}\}^T = \underline{\Psi}^{-1} \underline{\Phi} \underline{\Psi}^{-1}$$

Thus for a square, non-singular matrix $\underline{\Psi}$ the error in $\underline{\hat{\theta}}$ is particularly dependent on the magnitude of the determinant of $\underline{\Psi}$.

Example 6.4.1

Consider the case of example 6.3.2 where

$$V(\underline{z}^{-1}) = 1 - \underline{z}^{-1}$$

$$G(\underline{z}^{-1}) = 0.8 - 0.4\underline{z}^{-1}$$

Then if $\underline{\hat{\theta}}^T = [\hat{a}_1 \hat{b}_1]$, the matrix $\underline{\Psi}$ is given by

$$\underline{\Psi} = \begin{bmatrix} 1.0 & -0.8 \\ -1.0 & -0.4 \end{bmatrix}$$

and from eqn.(6.4.4),

$$\begin{bmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{bmatrix} = \frac{1}{0.4} \begin{bmatrix} -0.4 & -0.8 \\ 1.0 & 1.0 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$$

giving

$$\begin{aligned} \hat{a}_1 - a_1 &= -\delta_1 - 2\delta_2 \\ \hat{b}_1 - b_1 &= 2.5(\delta_1 + \delta_2) \end{aligned}$$

Thus if δ_1 and δ_2 have the same sign the errors on $\hat{a}_1$ and $\hat{b}_1$ could be considerable. This can clearly be seen in a graph of $\hat{a}_1$ against $\hat{b}_1$ - from eqn.(6.4.1),

$$(i) \quad \hat{r}_1 - v_1 = \hat{a}_1 + g_0 \hat{b}_1$$

$$(ii) \quad \hat{r}_2 = v_1 \hat{a}_1 + g_1 \hat{b}_1$$

and for this example,

$$(i) \quad -0.45 = \hat{a}_1 + 0.8\hat{b}_1$$

$$(ii) \quad 0.65 = -\hat{a}_1 - 0.4\hat{b}_1$$

These two lines are shown in Fig. 6.3. A positive error δ_1 in $\hat{r}_1$ moves line (i) in the negative $\hat{a}_1$ direction whereas a positive error δ_2 in $\hat{r}_2$ moves line (ii) in the positive $\hat{a}_1$ direction. Since the two lines

have similar slopes it can be seen from Fig.6.3 that a small error δ causing a shift of one line in the $\hat{a}_1$ direction leads to a considerable shift in the point of intersection (i.e. the solution point). An example is shown by the dotted line in Fig.6.3 which corresponds to an error $\delta_1 = -0.15$ and a shift of line (i) to

$$-0.30 = \hat{a}_1 + 0.8b_1$$

The results of the above example are typical of many estimation situations which will occur in practice. However, the fact that the updated parameter estimation errors depend on the form of the matrix $\underline{\Psi}$, and in particular its determinant, suggests that it may be possible to design the initial control system such that the errors are minimised. For instance, in example 6.4.1 it would be desirable to choose g_0 , g_1 and v_1 such that the two lines in the parameter space are approximately orthogonal.

Alternatively it may be possible to form another linear algebraic equation in the parameters θ_i which will reduce the estimation errors. This is equivalent to forming a line in parametric space approximately orthogonal to the lines derived from the matrix equation (6.4.1). It has been found that such a line represents the steady-state gain of the process, i.e.

$$\left. \frac{B(z^{-1})}{1 + A(z^{-1})} \right|_{z=1} = K$$

or

$$\sum_{i=1}^n b_i = K \left[1 + \sum_{i=1}^n a_i \right] \quad (6.4.4)$$

Since the steady-state gain K is a most important parameter in the design of a control policy - whether control be manual or automatic - it is probable that, in practice, a good estimate will be available prior to identification. Alternatively an estimate of K may

be obtained by the method proposed in Appendix IV; this only applies, however, when the disturbance $\{x_t\}$ has a zero mean.

The inclusion of the gain equation (6.4.4) requires a modification to the matrix equation (6.4.1), i.e.

$$\begin{bmatrix} \hat{\underline{r}} \\ -\hat{\underline{K}} \end{bmatrix} = \begin{bmatrix} \underline{\Psi} \\ \hat{\underline{K}} \end{bmatrix} \begin{bmatrix} \hat{\underline{\theta}} \end{bmatrix} + \begin{bmatrix} \underline{v} \\ 0 \end{bmatrix} \quad (6.4.5)$$

where $\hat{\underline{K}}$ is an estimate of $\underline{K}$ and the $(n_1 + n_2)$ row vector $\hat{\underline{K}}$ is given by

$$\hat{\underline{K}} = \begin{bmatrix} -\hat{\underline{K}} & -\hat{\underline{K}} & . & . & -\hat{\underline{K}} & | & 1 & 1 & . & . & 1 \end{bmatrix}$$

and the partition occurs at the n_1^{th} element. Eqn.(6.4.5) can be solved for $\hat{\underline{\theta}}$ by the regression algorithm (6.3.13).

Example 6.4.2

Consider again the initial control system of examples 6.3.2 and 6.4.1, and let $\hat{r}_1 = -1.30$, $\hat{r}_2 = 0.70$. Then the original matrix equation (6.4.1) is

$$\begin{bmatrix} -1.30 \\ 0.70 \end{bmatrix} = \begin{bmatrix} 1.0 & 0.8 \\ -1.0 & -0.4 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} + \begin{bmatrix} -1.0 \\ 0 \end{bmatrix} \quad (6.4.6)$$

Let an estimate of $\underline{K}$ (true value 3.33') be 3.0; then the modified gain equation (6.4.5) is

$$\begin{bmatrix} -1.3 \\ 0.7 \\ 3.0 \end{bmatrix} = \begin{bmatrix} 1.0 & 0.8 \\ -1.0 & -0.4 \\ -3.0 & 1.0 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} + \begin{bmatrix} -1.0 \\ 0 \\ 0 \end{bmatrix}$$

and from (6.3.13) the estimated parameter vector is

$$\begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} -0.82 \\ 0.56 \end{bmatrix}$$

This estimate compares favourably with the estimate from eqn.(6.4.6) of

Fig. 6.4. MODIFIED SOLUTION IN EXAMPLE 6.4.2.

$\hat{a}_1 = -1.10$, $\hat{b}_1 = 1.00$. Fig. 6.4 shows the affect of introducing the gain equation; lines (i) and (ii) represent the two equations (6.4.6) giving a solution at S_1 . The gain equation is line (iii) and the modified solution is at S_2 . The true solution point is represented as point S.

The above example is typical of the modified results obtained by including the gain equation. In general it is found that the modified estimates represent a point in parametric space on or close to the line representing the gain equation. It follows that although the individual estimates $\hat{a}_i$, $\hat{b}_i$ are improved the updated value of K is little different from its original estimate.

The method of simultaneous identification and control presented in this and the previous sections is as yet just a proposal. Due to the inadequacies of existing estimation techniques for mixed models and to doubts about the accuracy of the method, it is not as proposed sufficiently well-developed to be widely applicable to practical problems. It does, however, offer a general solution to the simultaneous identification and control problem for a single-input, single-output, linear dynamic system and may with further research prove to be a valuable technique for application to industrial process control. Meanwhile there remain several problems which restrict the application of all the techniques derived in this thesis.

6.5. Some problems associated with simultaneous identification and control.

For a technique of simultaneous identification and control to be applicable to all practical problems it must cater for the following characteristics of the process:-

- (i) a time-variable set-point;
- (ii) time-variable process dynamics and in particular a variable time-delay;

- (iii) a non-stationary disturbance;
- (iv) multivariable systems;
- (v) non-linear dynamics.

All the techniques derived in this thesis have assumed a constant set-point during the period of identification. In practice it may be possible to allow for a limited number of set-point changes as follows : from eqn.(3.1.1)

$$X_t = \frac{M(z^{-1})x_t + [z^{-k}P(z^{-1}) - z^{-1}M(z^{-1})]E_t^d}{M(z^{-1})[1-H(z^{-1})] + z^jP(z^{-1})H(z^{-1})}$$

A change in $\{E_t^d\}$ is followed by a transient in $\{X_t\}$. If, however, $z^{-1}M(z^{-1}) \ll z^{-k}P(z^{-1})$ the transient is small and if its duration is short compared with the total experimentation time then it can often be neglected. The effect on $\{X_t\}$ of a set-point change can then be regarded as caused by the disturbance. In practice it has been found that one or two set-point changes can be allowed in the identification experiment without causing significant errors in the estimated parameters.

The second problem concerns the time-variation of the process dynamics. If the variation is small during the identification period it can often be neglected. However, faster variations require an adaptive approach to identification and control. In this context the recent work of Peterka^{12,30} may form the basis of a practical solution. This work employs an iterative least-squares approach to estimation and control and updates the control algorithm after every sample. However, as yet it has only been shown that the adaptive algorithm can optimise a time-invariant system. Peterka³⁰ also notes that it is not possible, in general, to identify the process and disturbance dynamics when the process is under feedback control - his work requires estimates of the parameters in the control algorithm only and as such does not identify the process. However the proposal of section 6.3 has shown that in theory it is possible to identify and control simultaneously.

A particularly common form of time-variation concerns the process time-delay k ; e.g. in flow processes where the time-delay is directly related to flow. In many cases changes in k will be steps and, in general, will not be an integral number of sample intervals. For instance if the time-delay changes from k to k' sample intervals then the process changes from $z^{-k}P(z^{-1})$ to $z^{-k'}P'(z^{-1})$ where, in general, $P'(z^{-1}) \neq P(z^{-1})$. Since the time-delay must be represented by an integral number of sample intervals (i.e. in the z -transform representation) the new process will be $z^{-l}P''(z^{-1})$ where $l = \text{Integer}(k')$ and $P''(z^{-1}) = z^{-k'+l}P'(z^{-1})$. Thus even if $P'(z^{-1}) = P(z^{-1})$ the new representation of the process $P''(z^{-1})$ is, in general, different from the original $P(z^{-1})$. Since this change can occur in one sample interval it presents a particularly difficult problem for an identification technique. It may, however, be possible to use the modified z -transform representation²⁴ in this case although the exact nature of the time-delay change would have to be known.

Thirdly there are the problems associated with a non-stationary disturbance. It is usual to represent such a disturbance by an integrated model (see section 1.2) where it is assumed that an l^{th} -difference of the disturbance is stationary. However, a predictor based on the integrated model although it can adapt to variations in the mean, variance, etc. of the disturbance⁹ does not actually predict these variations. A more general approach is allowed using Kalman-filtering techniques⁴ although this requires a description of the time variation.

The fourth problem is the extension of simultaneous identification and control techniques to multivariable systems. It is not, at present, clear exactly how the proposals of this thesis can be extended in this way. The major problem is that for an m -input, n -output system ($m > n$) there will be only n -signals $\{X_t\}$ and therefore insufficient information unless $m = n$. However there is no necessity, in

theory, for basing the updating on $\{X_t\}$ since similar information is contained in the output $\{E_t\}$ and the input $\{c_t\}$. It may, therefore, be possible in the multivariable case to base updating on the m inputs $\{c_t\}$. It is more likely, however, that the more general representation of the state-variable approach will have to be adopted. In this context it would be valuable to find the analogue of the techniques of this thesis in the state-space.

Finally there are the many difficulties associated with non-linearities in the process dynamics. Although some procedures have been developed for identifying non-linear dynamics (e.g. Detchmendy and Sridhar³³) they assume a priori a particular structure for the non-linearity. For the general case a good deal more research is needed into the structural representation of non-linearities for identification and control purposes.

CHAPTER VII

CONCLUSIONS

7.1. Simultaneous Identification and Control as a Practical Proposition

There are many practical advantages of identifying a process while it is under feedback control. The most important are as follows:-

- (i) identification is carried out while the process is operated in a similar mode to that of the resulting optimum control system;
- (ii) the input perturbation is supplied by the inherent process disturbances and not from an external source;
- (iii) closed-loop operation can, in general, decrease the output variance from the open-loop value with a corresponding economic saving.

The techniques developed in this thesis have all the above advantages over the established methods of identification, which assume that identification and control can be treated independently. The several proposals of Chapter III, however, are limited to specific classes of systems. The methods of process-model and predictor updating are restricted to systems where the disturbance is known a priori to be stationary and the separation technique is restricted to processes with large time-delays. On the other hand, the computational algorithms are simple and are easily programmed for on-line use. The separation technique in particular appears to be of considerable practical importance - in this case no a priori assumptions need be made concerning the disturbance. Finally the methods of Chapter VI offer a solution to a more general class of simultaneous identification and control problems although some further research needs to be conducted into the estimation of mixed models.

7.2. Applications

An important part of the research work of this thesis has

been concerned with the industrial investigation of Chapter V. It was found at an early stage that none of the established identification techniques could be applied to this process due to various restraints on experimentation. In this situation the only alternative was to attempt to apply a method of simultaneous identification and control and in particular the separation technique. It was shown in Chapter V that the separation technique produced consistent estimates of the process dynamics for relatively small samples of data. As a consequence of the two updating analyses it was possible to design an automatic control system which substantially reduced the variations in the output. These results provided a justification of the hypothesis that the separation technique is a viable method of identification and is particularly insensitive to the non-ideal conditions encountered in practice.

The control system did not, however, include an optimum predictor as originally planned due to doubts concerning the disturbance model. Although the separation technique produced estimates of the disturbance model these were inconsistent and did not suggest that optimum prediction could significantly reduce the output variations for the order of time-delay to be expected. It was concluded that the major benefits from automatic control would be the elimination of drift in the output; in these circumstances a predictor would be required to forecast the drift and not the higher-frequency components of the disturbance.

7.3. Topics of further research

It is clear that, at the present time, the theory of stochastic control is limited to ideal systems which will rarely be encountered in practice². The major practical problems are concerned with the representation of non-stationary disturbances and non-linear, time-varying dynamics and until these can be solved it is unlikely that stochastic control theory will find general application in the chemical

process industry. The industrial investigation of Chapter V adequately demonstrated these problems and showed that although some allowance could be made for non-linearities in the process the complete problem of time-varying, non-linear and non-stationary dynamics was far from being solved.

It is thought that this thesis has presented a new approach to the identification problem which is liable to find extensive application in the chemical process industry. However if the object of identification is to be the implementation of a stochastic control system the theories must be extended to cater for a more general class of processes. A satisfactory solution of the identification problem is of particular importance since the effectiveness of the resulting control system is to a great extent dependent on the amount of information present in the dynamical model of the process.

APPENDIX ITHE FRONT-END TRUNCATION OPERATOR $[.]_T$ Definition

The k^{th} -order truncation operator $[.]_T$ acting on a polynomial in z^{-1} , eliminates all the terms in z^{-i} for $i < k$. Hence if

$$R(z^{-1}) = r_0 + \sum_{i=1}^{\infty} r_i z^{-i} \quad (\text{A.1.1})$$

$$\text{then } [R(z^{-1})]_T = \sum_{i=k}^{\infty} r_i z^{-i} \text{ for } k > 0 \quad (\text{A.1.2})$$

As a convention $[R(z^{-1})]_T$ will be written as $R_T(z^{-1})$.

Properties

Consider the series $1 + R(z^{-1})$ where $R(z^{-1})$ is as defined in eqn.(A.1.1) except that $r_0 = 0$. Then it follows that

$$(i) \quad [1 + R(z^{-1})]_T = R_T(z^{-1}) \text{ since } k > 0 \quad (\text{A.1.3})$$

$$(ii) \quad [R_T(z^{-1})]_T = R_T(z^{-1}) \quad (\text{A.1.4})$$

$$(iii) \quad R(z^{-1}) - R_T(z^{-1}) = \sum_{i=1}^{k-1} r_i z^{-i} \quad (\text{A.1.5})$$

$$(iv) \quad \{[1 + R_T(z^{-1})]^{-1}\}_T = - \frac{R_T(z^{-1})}{1 + R_T(z^{-1})} \quad (\text{A.1.6})$$

This property is derived by expanding the left-hand side of eqn.

(A.1.6) by the binomial theorem and then truncating, i.e.

$$\begin{aligned} \{[1 + R_T(z^{-1})]^{-1}\}_T &= \{1 - R_T(z^{-1}) + (R_T(z^{-1}))^2 - \dots\}_T \\ &= -R_T(z^{-1})[1 - R_T(z^{-1}) + (R_T(z^{-1}))^2 - \dots] \end{aligned}$$

$$(v) \quad \text{if } Q(z^{-1}) = \sum_{i=0}^{\infty} q_i z^{-i} \text{ then}$$

$$[Q(z^{-1})R_T(z^{-1})]_T = Q(z^{-1})R_T(z^{-1}) \quad (\text{A.1.7})$$

APPENDIX IISYSTEM RESPONSE TO A VARIABLE-MEAN DISTURBANCE

Consider the linear dynamic system represented by

$$E_t = \sum_{i=0}^{\infty} k_i x_{t-i} = K(z^{-1})x_t \quad (\text{A.2.1})$$

where
$$\sum_{i=0}^{\infty} |k_i| < \infty$$

The sequences $\{E_t, t = 0, 1, \dots\}$, $\{x_t, t = 0, 1, \dots\}$ represent the system output and a random disturbance respectively, while $K(z^{-1})$ is the impulse response of a stable linear process.

Then we can form the following relationship between the ensemble averages of $\{E_t\}$ and $\{x_t\}$,

$$\mu_E(t) = \sum_{i=0}^{\infty} k_i \mu_x(t-i) \quad (\text{A.2.2})$$

where $\mu(t)$ refers to an ensemble average at time t . Let $\mu_x(t)$ vary over an arbitrary interval $0 < t < \tau$ according to

$$\mu_x(t) = \mu_x(t-1) + \delta_t \quad (\text{A.2.3})$$

where the δ_i are arbitrarily small numbers. Further define the initial and final conditions as

$$\mu_x(t) = \mu_x(0) \quad \text{for } t < 0 \quad (\text{A.2.4})$$

$$\mu_x(t) = \mu_x(\tau) + \mu_x(0) \quad \text{for } t > \tau \quad (\text{A.2.5})$$

where $\mu_x(0)$, $\mu_x(\tau)$ are constants.

Then eqn.(A.2.2) can be expanded using (A.2.4), (A.2.5) to

give

$$\mu_E(t) = \sum_{i=t}^{\infty} k_i \mu_x(0) + \sum_{i=t-\tau+1}^{t-1} k_i \mu_x(t-i) + \sum_{i=0}^{t-\tau} k_i [\mu_x(\tau) + \mu_x(0)]$$

and from eqn.(A.2.3)

$$\mu_E(t) = \sum_{i=0}^{\infty} k_i \mu_x(0) + \sum_{i=0}^{t-\tau} k_i \mu_x(\tau) + \sum_{i=t-\tau+1}^{t-1} k_i \sum_{j=1}^{t-i} \delta_j \quad (\text{A.2.6})$$

Then in the steady-state

$$\lim_{t \rightarrow \infty} \mu_E(t) = \sum_{i=0}^{\infty} k_i \mu_x(0) + \sum_{i=0}^{\infty} k_i \mu_x(\tau) + \lim_{t \rightarrow \infty} \sum_{i=t-\tau+1}^{t-1} k_i \sum_{j=1}^{t-i} \delta_j \quad (\text{A.2.7})$$

and since the stability condition

$$\sum_{i=0}^{\infty} |k_i| < \infty \text{ implies } \lim_{i \rightarrow \infty} k_i = 0$$

it follows that

$$\lim_{t \rightarrow \infty} \sum_{i=t-\tau+1}^{t-1} k_i \sum_{j=1}^{t-i} \delta_j = 0 \text{ if } \sum_{j=1}^N \delta_j < \infty \text{ for all } N \quad (\text{A.2.8})$$

The condition in eqn.(A.2.8), that the sum of any N consecutive δ 's should be bounded, is exactly equivalent to requiring that $\mu_x(\tau)$ be finite for all τ .

We can now state the following theorem:-

THEOREM A.2.

For a linear dynamic system defined by eqn.(A.2.1) subjected to a bounded, time-variable mean, disturbance the ensemble average of the output $\{E_t\}$ is zero in the steady-state if and only if,

$$K(z^{-1})|_{z=1} = 0 \quad (\text{A.2.9})$$

PROOF

A time-variable mean is equivalent to a time variation of the ensemble average. For a zero-mean stationary random signal $\{x_t\}$ the ensemble average $\mu_x(t)$ is zero for all t. However $\mu_x(t)$ may vary if the random signal $\{x_t\}$ is non-stationary. Theorem A.2 applies to a random signal $\{x_t\}$ which is non-stationary over an arbitrary interval $0 < t < \tau$.

It has been shown that eqn.(A.2.7) holds for a signal $\{x_t\}$

whose ensemble average is a constant $\mu_x(0)$ for $t \leq 0$ and remains constant at $\mu_x(0) + \mu_x(\tau)$ for $t \geq \tau$. Then if the change in level $\mu_x(\tau)$ is finite for any τ , eqn.(A.2.8) shows that eqn.(A.2.7) reduces to

$$\lim_{t \rightarrow \infty} \mu_E(t) = [\mu_x(0) + \mu_x(\tau)] K(z^{-1})|_{z=1} \quad (\text{A.2.10})$$

Then $\mu_E(t)$ is zero in the steady-state, for any bounded $\mu_x(\tau)$, if and only if

$$K(z^{-1})|_{z=1} = 0 \quad , \quad \text{Q.E.D.}$$

COROLLARY

From eqn.(A.2.10) it can be seen that the steady-state ensemble average of $\{E_t\}$ is equal to the steady-state ensemble average of $\{x_t\}$ if and only if

$$K(z^{-1})|_{z=1} = 1$$

APPENDIX IIIPREDICTORS FOR AUTOREGRESSIVE MODELS

We are concerned here with a random process $\{x_t\}$ represented by an n^{th} -order autoregressive model

$$x_t = \frac{1}{1 - Q(z^{-1})} \zeta_t \quad (\text{A.3.1})$$

$$\text{where } Q(z^{-1}) = \sum_{i=1}^n q_i z^{-i}$$

and $\{\zeta_t\}$ is a white noise sequence.

THEOREM A.3.1

An optimum k^{th} -step ahead linear predictor for an n^{th} -order autoregressive process has n non-zero coefficients for all permissible k , i.e.

$$H_T(z^{-1}) = \sum_{i=0}^{n-1} h_i z^{-k-i} \quad (\text{A.3.2})$$

PROOF

$$\text{Let } [1 - Q(z^{-1})]^{-1} = 1 + \sum_{i=1}^{\infty} S_i z^{-i} \quad (\text{A.3.3})$$

Then by the binomial theorem, with $S_0 = 1$,

$$\left. \begin{aligned} S_i &= \sum_{j=1}^i q_j S_{i-j} \quad \text{for } i < n \\ S_i &= \sum_{j=1}^n q_j S_{i-j} \quad \text{for } i \geq n \end{aligned} \right\} \quad (\text{A.3.4})$$

The optimum k^{th} -step ahead linear predictor for the process represented by eqn.(A.3.1) is given by eqn.(2.5.10) as

$$H_T(z^{-1}) = \{[1 - Q(z^{-1})]^{-1}\}_T (1 - Q(z^{-1}))$$

and from eqn.(A.3.3)

$$H_T(z^{-1}) = \sum_{i=k}^{\infty} S_i z^{-i} [1 - Q(z^{-1})] \quad (A.3.5)$$

Expanding this equation we have

$$\begin{aligned} H_T(z^{-1}) &= S_k z^{-k} + \sum_{i=k+1}^{k+n} [S_i - \sum_{j=k}^{i-1} S_j q_{i-j}] z^{-i} \\ &+ \sum_{i=k+n+1}^{\infty} S_i z^{-i} - \sum_{i=1}^n q_i z^{-i} \sum_{j=k+n+1}^n S_j z^{-j} \end{aligned} \quad (A.3.6)$$

It is easily seen by comparison with eqns.(A.3.4) that the first n coefficients of z^{-i} in eqn.(A.3.6) are zero for $|q_i| > 0$, $i \leq n$.

However, consider any term in z^{-k-n-l} for $l > 0$,

$$h_{n+l} = S_{k+n+l} - \sum_{j=1}^n q_j S_{k+n+l-j}$$

but from eqn.(A.3.4),

$$S_{k+n+l} = \sum_{j=1}^n q_j S_{k+n+l-j} \quad \text{for } (k+l) > 0$$

Hence

$$h_{n+l} = 0 \quad \text{for } l > 0$$

and the k -step ahead predictor of eqn.(A.3.6) reduces to

$$H_T(z^{-1}) = \sum_{i=0}^{n-1} h_i z^{-k-i} \quad \text{Q.E.D.}$$

THEOREM A.3.2.

The k^{th} -step ahead predictor $H_k(z^{-1}) = H_T(z^{-1})$ for an autoregressive process $\{x_t\}$ can be calculated from

$$H_k(z^{-1}) = \sum_{i=1}^{k-1} q_i H_{k-i}(z^{-1}) z^{-i} + \sum_{i=k}^n q_i z^{-i} \quad (A.3.7)$$

where $0 < k < n$ and $H_{k-i}(z^{-1})$ is the $(k-i)^{\text{th}}$ -step ahead optimum predictor.

Also

$$H_k(z^{-1}) = \sum_{i=1}^n q_i H_{k-1}(z^{-1}) z^{-i} \quad \text{for } k \geq n \quad (A.3.8)$$

PROOF

The definition of $H_{k-i}(z^{-1})$, $i < k$, is given in eqn.(A.3.5)

as

$$H_{k-i}(z^{-1}) = \sum_{j=k-i}^{\infty} S_j [1 - Q(z^{-1})]$$

Then if eqn.(A.3.7) holds for $k < n$, we have

$$\sum_{j=k}^{\infty} S_j z^{-j} [1 - Q(z^{-1})] = \sum_{i=1}^{k-1} q_i z^{-i} \sum_{j=k-i}^{\infty} S_j z^{-j} [1 - Q(z^{-1})] + \sum_{i=k}^n q_i z^{-i}$$

and dividing through by $[1 - Q(z^{-1})] = [1 + S(z^{-1})]^{-1}$,

$$\sum_{j=k}^{\infty} S_j z^{-j} = \sum_{i=1}^{k-1} q_i z^{-i} \sum_{j=k-i}^{\infty} S_j z^{-j} + \sum_{i=k}^n q_i z^{-i} [1 + \sum_{j=1}^{\infty} S_j z^{-j}]$$

or

$$\sum_{j=k}^{\infty} S_j z^{-j} = \sum_{i=1}^{k-1} \sum_{j=k-i}^{\infty} q_i S_j z^{-i-j} + \sum_{i=k}^n q_i z^{-i} + \sum_{i=k}^n \sum_{j=1}^{\infty} q_i S_j z^{-i-j} \quad (\text{A.3.9})$$

Then equating coefficients of z^{-k-1} , $0 < l < n$,

$$S_{k+1} = \sum_{i=1}^{k-1} q_i S_{k+1-i} + q_{k+1} + \sum_{i=k}^{k+1-1} q_i S_{k+1+i}$$

or

$$S_{k+1} = q_{k+1} + \sum_{i=1}^{k+1-1} q_i S_{k+1-i}$$

which is true by the definition of $S(z^{-1})$ in eqn.(A.3.4).

Also for $l > n$,

$$S_{k+1} = \sum_{i=1}^n q_i S_{k+1-i}$$

which also holds due to the relation (A.3.4). Hence eqn.(A.3.7)

holds for all $0 < k < n$.

Similarly for $k > n$, eqn.(A.3.8) holds if

$$\sum_{j=k}^{\infty} S_j z^{-j} [1 - Q(z^{-1})] = \sum_{i=1}^n q_i z^{-i} \sum_{j=k-i}^{\infty} S_j z^{-j} [1 - Q(z^{-1})]$$

i.e. if

$$\sum_{j=k}^{\infty} S_j z^{-j} = \sum_{i=1}^n \sum_{j=k-i}^{\infty} q_i S_j z^{-i-j} \quad (\text{A.3.10})$$

Equating coefficients of z^{-k-l} , $l > 0$,

$$S_{k+l} = \sum_{i=1}^n q_i S_{k+l-i}$$

which is true from eqn.(A.3.4) for $k \geq n$ and $l > 0$. This proves that eqn.(A.3.8) holds for all $k \geq n$.

APPENDIX IVESTIMATION OF THE PROCESS STEADY-STATE GAIN

Consider the equation for the estimated disturbance $\{X_t\}$ in the Box & Jenkins control system, i.e.

$$X_t = \frac{M(z^{-1})x_t + [z^{-k}P(z^{-1}) - z^{-1}M(z^{-1})]E_t^d}{M(z^{-1})[1-H(z^{-1})] + z^jP(z^{-1})H(z^{-1})} \quad (A.4.1)$$

If $E_t^d = E^d$ is constant then eqn.(A.4.1) can be written

$$X_t = \frac{M(z^{-1})}{M(z^{-1})[1-H(z^{-1})] + z^jP(z^{-1})H(z^{-1})} \cdot x_t + \bar{X} \quad (A.4.2)$$

$$\text{where } \bar{X} = \left. \frac{z^{-k}P(z^{-1}) - z^{-1}M(z^{-1})}{M(z^{-1})[1-H(z^{-1})] + z^jP(z^{-1})H(z^{-1})} \right|_{z=1} \cdot E^d \quad (A.4.3)$$

and $\bar{X}$ is a constant. In practice the level $\bar{X}$ builds up from $t = 0$ to a constant value after about 20 sample intervals - this transient follows the impulse response of the function

$$\frac{z^{-k}P(z^{-1}) - z^{-1}M(z^{-1})}{M(z^{-1})[1-H(z^{-1})] + z^jP(z^{-1})H(z^{-1})}$$

If $\{x_t\}$ is a zero-mean random signal then, after the transient has settled, $\bar{X}$ becomes the mean of the signal $\{X_t\}$, i.e.

$$E\{X_t\} = \bar{X} \quad \text{if} \quad E\{x_t\} = 0$$

Thus if a record of $\{X_t, t = I + 1, \dots, N\}$ is available, the mean $\bar{X}$ can be estimated from

$$\bar{X} = \frac{1}{N - I} \sum_{t=I+1}^N X_t$$

Then from eqn.(A.4.3)

$$\bar{X} = \frac{P(1) - M(1)}{M(1)[1-H(1)] + P(1)H(1)} E^d$$

where, for instance, $M(1) = M(z^{-1})|_{z=1}$

and rearranging

$$P(1) = \frac{M(1)\{E^d + \bar{X}[1-H(1)]\}}{E^d - H(1)\bar{X}} \quad (A.4.4)$$

Since all the terms on the right-hand side of this equation are known, an estimate of $P(1)$ - the steady-state gain of the process - can be calculated. In particular if $H(1) = 1$, as will be the case in most practical examples, eqn.(A.4.4) reduces to

$$P(1) = \frac{M(1)E^d}{E^d - \bar{X}} \quad (A.4.5)$$

In summary, if the disturbance $\{x_t\}$ has a zero-mean and if the set-point E_t^d is constant then an estimate of the process steady-state gain can be calculated given the average $\bar{X}$ of the signal $\{X_t\}$.

REFERENCES

1. BALAKRISHNAN, A.V., PETERKA, V. : Identification in Automatic Control Systems. IFAC Proceedings of the Fourth Congress, Warsaw, 1969
2. WONHAM, W.M. : Optimal Stochastic Control. Automatica, Vol 5, pp. 113-118, 1969
3. TOU, J.T. : Optimum Design of Digital Control Systems. Academic Press, 1963
4. KALMAN, R.E. : A New Approach to Linear Filtering and Prediction Problems. Trans. ASME, Series D, Journal of Basis Engineering, vol. 82, 1960, pp. 35-45
5. BOX, G.E.P., JENKINS, G.M. : Models for Prediction and control. Technical Reports 1 - 11, University of Lancaster, 1966
6. WHITTLE, P. : Prediction and Regulation The English Universities Press, 1963
7. ASTROM, K.J. : Computer Control of a Paper Machine - an Application of Linear Stochastic Control Theory. IBM Journal of Research and Development, Vol.11, 1967, No.4 (July), pp. 389-405
8. ASTROM, K.J., BOHLIN, T. : Numerical Identification of Linear Dynamic Systems from Normal Operating Records. Proc. IFAC Symposium on Self-Adaptive Control Systems, Teddington, Sept., 1965

9. BOX, G.E.P., JENKINS, G.M. : Further Contributions to Adaptive Quality Control : Simultaneous Estimation of Dynamics : Non-zero Costs.
Bulletin of the International Statistical Institute, 34th Session, 1963, Ottawa, Canada
10. GUSTAVSSON, I. : Parametric Identification on Multiple Input, Single Output Linear Dynamic Systems.
Report 6907, Lund Institute of Technology, July 1969
11. EYKHOFF, P. : Process Parameter Estimation.
Progress in Control Engineering, Vol. 2, 1963
12. PETERKA, V., SMUK, K. : On-line Estimation of Dynamic Model Parameters from Input-Output Data.
IFAC Proceedings of the Fourth Congress, Warsaw, 1969
13. PANUSKA, V. : A Stochastic Approximation Method for Identification of Linear Systems using Adaptive Filtering.
UKAC, Advances in Computer Control, Leicester, April 1968
14. SAKRISON, D.J. : The Use of Stochastic Approximation to Solve the System Identification Problem.
IEEE Trans. Automatic Control, AC-12, 1967, 5 (October), pp. 563-567
15. EYKHOFF, P. : Process Parameter and State Estimation.
IFAC Symposium on Identification in Automatic Control Systems, Prague, 1967, p. 2
16. ASTROM, K.J. : Lectures on the Identification Problem - The Least Squares Method.
Report 6806, Lund Institute of Technology, Sept. 1968

17. FLORENTIN, J.J., HAINSWORTH, B.D., RESWICK, J.B., WESTCOTT, J.H.
Correlation Analysis of a Heat Exchanger.
Joint Symposium. Instrumentation and Computation,
Institution of Chemical Engineers, 1959
18. PRIESTLEY, M.B. : Estimation of Transfer Functions in Closed
Loop Stochastic Systems.
Automatica, Vol 5, No. 5, pp. 623-632, Sept. 1969
19. BRIGGS, P.A.N., GODFREY, K.R., HAMMOND, P.H. : Estimation of
Process Dynamic Characteristics by Correlation
Methods using Psuedo Random Signals.
IFAC Symposium on Identification in Automatic
Control Systems, Prague, 1967, p. 3.10
20. CLARKE, D.W. : Generalized Least Squares Estimation of the
Parameters of a Dynamic Model.
IFAC Symposium on Identification in Automatic
Control Systems, Prague, 1967, p. 3.17
21. CLARKE, D.W. : The Accuracy of Industrial Identification Techniques.
IEE Conference on Industrial Applications of
Dynamic Modelling, Conf.Pub.No.57, pp. 31-41,
Sept. 1969
22. BOX, G.E.P., JENKINS, G.M. : Mathematical Models for Adaptive
Control and Optimization.
Tech. Report No.45, University of Wisconsin, May 1965
23. BULLAS, D.O. : Models of a Continuous Industrial Polymerisation
Process.
IEE Conference on Industrial Applications of
Dynamic Modelling, Conf.Pub.No.57, pp.72-80,
Sept. 1969

24. LINDORFF, D.P. : Theory of Sampled-Data Control Systems.
John Wiley & Sons, 1965
25. ZADEH, L.A., DESOER, C.A. : Linear System Theory, The State-Space Approach.
McGraw-Hill, 1963
26. BOX, G.E.P., JENKINS, G.M. : Some Statistical Aspects of Adaptive Optimization and Control.
Jnl. R. Statis. Soc., B, 24, pp. 297-343, 1962.
27. PLACKETT, R.S. : Applied Regression Analysis.
Oxford Press, 1960
28. KOOPMANS, T.C. : Statistical Inference in Dynamic Economic Models
Cowles Commission for Research in Economics
John Wiley & Sons, 1950
29. DURBIN, J. L The Fitting of Time Series Models
Rev. Int. Inst. Stat., 28, 233-243, 1960
30. PETERKA, V. : Adaptive Digital Regulation of Noisy Systems.
IFAC Symposium on Identification and Process
Parameter Estimation, Prague, June 1970
31. AOKI, M. : Optimisation of Stochastic Systems
Academic Press 1967
32. BOHLIN, T. : On the Problem of Ambiguities in Maximum Likelihood Identification.
IFAC Symposium on Identification and Process
Parameter Estimation, Prague, June 1970
33. DETCHMENDY & S. RIDHAR : Sequential Estimation of States and Parameters in Noisy Non-Linear Dynamical Systems.
Proc. JACC, pp. 55-63, 1965

PUBLICATIONS

1. PHILLIPSON, P.H., TURTLE, D.P. : Updating the Box & Jenkins
Control Scheme.
I.E.E. Conference on Industrial Applications
of Dynamic Modelling, Conf. Pub. No.57, Sept.1969.

2. PHILLIPSON, P.H., TURTLE, D.P. : Simultaneous Identification and
Control for Processes with Time-Delays.
IFAC Symposium on Identification and Process
Parameter Estimation, Prague, June 1970, paper 6.1.