THE UNIVERSITY OF ALBERTA

PARAMETER ESTIMATION AND MULTIRATE ADAPTIVE CONTROL

by

© Weiping Lu

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

IN

PROCESS CONTROL

Department of Chemical Engineering

EDMONTON, ALBERTA

SPRING 1989

ISBN 0-315-52968-7

University of Alberta | Inter-departmental Correspondence

to: Faculty of Graduate Studies and Research

date: April 13, 1989

our file:

from: Professors S.L. Shah
and D.G. Fisher

your file:

subject:

Ph.D. Thesis by Mr. Weiping Lu

The PhD research work done by Mr. Lu resulted in several publications on which we were shown as co-authors. This letter is to confirm that we agree to the inclusion of this material in his thesis and hereby provide a release to the University as follows:

We grant permission to the University of Alberta Library to reproduce single copies of Mr. Lu's PhD thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

SIGNED ______________________
Professor D. G. Fisher

____
Professor S. L. Shah

Date: April 13, 1989

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Weiping Lu

TITLE OF THESIS PARAMETER ESTIMATION AND MULTIRATE ADAPTIVE CONTROL

DEGREE FOR WHICH THESIS WAS PRESENTED DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED SPRING 1989

(SIGNED) Weiping Lu..........

PERMANENT ADDRESS:

#202 11140 - 68 AV...........

Edmonton Alberta......

...Canada.................

DATED April 12..........1989

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled PARAMETER ESTIMATION AND MULTIRATE ADAPTIVE CONTROL submitted by Weiping Lu in partial fulfilment of the requirements for the degree of DOCTOR OF PHILOSOPHY in PROCESS CONTROL.

....
Professor D. G. Fisher
Supervisor

....
Professor S. L. Shah

....
Professor R. K. Wood

....
Associate Professor
P. J. Crickmore

....
Professor
V. G. Gourishankar

....
Professor
P. K. Kokotovic
External Examiner

Date...March 9, 1989......

To my parents,

to Zhili and

to Linshuang and the next one

ABSTRACT

This thesis includes a number of theoretical contributions that are complete in themselves but are also related by the area of application and/or the theoretical approach used. The first area includes parameter identification, output estimation, adaptive control and adaptive inferential control of multirate systems. The work on parameter convergence is then extended to include some general results for systems where the identified model is overparameterized.

A multirate system is defined as one in which the output is sampled at an integer, J, multiple of the input sample period T, i.e. the data collected are $\{y(kJT),[u((k-1)JT+iT),i=1,2,...,J],\ k=0,1,2,...\}$. Three main contributions are made with respect to multirate systems: (i) formulation and convergence analysis of parameter and output estimation algorithms, which produce process parameter estimates whenever new measured outputs are available and output estimates at every input sampling interval, i.e. every interval T, (ii) formulation and convergence analysis of a multirate adaptive control algorithm, which uses the output estimates calculated at every input sampling interval as a closed loop feedback signal, (iii) formulation and convergence analysis of a multirate inferential estimation algorithm, which makes use of a secondary output measurement to improve the intersampling estimates and is a first step toward the

extension of the multirate system study to multi-input, multi-output systems.

Overparameterized models result in formulations of multirate systems (discussed above) plus in many adaptive systems in which there is uncertainty about the order of the actual plant and/or the disturbances. This work uses the concept of quotient space to prove that for a broad class of system signals (cf. excitation), the parameters converge to an equivalent class set in the parameter vector space, or if the parameter estimates are constrained, to the intersection of the equivalent class set and the constrained region. The general results are applied specifically to the ordinary and extended least squares algorithms.

Acknowledgement

First, I would like to express my gratitude to my supervisor Dr. D. Grant Fisher for his enthusiastic supervision, help and guidance leading to this thesis.

I would like also thank Dr. A. J. Morris for providing some of the initial motivation and direction for Chapter 5 on the multirate inferential estimation and control.

I owe much to discussions with professors, research staff members and graduate students in the Process Control group. It is with great pleasure that I take this opportunity to thank them.

Special thanks are also due to Mr. Robert L. Barton and Mr. Henry Sit of the DACS centre for their never-ending patience in answering questions related to computing facilities.

Finally, I would like to express my love and thanks to my wife, Zhili, for her love, sacrifice and support throughout the four years of my Ph. D program.

Table of Contents

List of Figures

Figure Page

1. INTRODUCTION

This thesis deals with parameter identification, output estimation and adaptive control of linear time invariant systems with unknown parameters and two sampling rates: the slow rate for the output and the fast rate for the input. It also includes a theoretical study of parameter convergence with overparameterized models, which was originally undertaken to solve the parameter convergence problem associated with the identification of the multirate systems described above. However the results obtained are far more general than required for multirate systems and represent an important contribution to the identification area.

1.1 Multirate Adaptive Systems

In many discrete control applications the appropriate control interval is T but practical restrictions, such as the cycle time of a discrete measurement transducer, mean that the output values, y(t), are measured only at every J^{th} interval. For example, in chemical processing applications, composition analyzers such as gas chromatographs have a cycle time of several minutes compared to a desired control interval of say 30 seconds. If the control interval is increased to JT to match the availability of measurements then control performance deteriorates. It is not a trivial problem to maintain a control interval of T when the only available output measurements are at intervals of JT. However if the process can be represented by a continuous, SISO, linear, time-invariant system with a zero-order-hold of sampling period T for the input, the problem can be stated as follows: given the measurement sequence $\{u(kT), y(kJT), k=1,2,...\}$, where J is a positive integer and T is the basic sampling period, reconstruct the output values $\{y(kT), k=0,1,2,...\}$. If the output y(kT) can be accurately estimated at every (input) sampling instant then the control algorithm can be implemented using the reconstructed information with an interval T instead of an interval JT.

This thesis provides a systematic solution to the problem of modelling and adaptive control of the multirate systems described above. The scope includes model formulation, parameter identification, output estimation and adaptive control. The proposed algorithms are based on a solid theoretical framework and their convergence analysis is included.

Studies of non-adaptive multirate systems in areas such as computer control and signal processing date back to the 1950s. A comprehensive treatment is included in the book by Crochiere and Rabina (1983) and a recent paper by Arki and Yamamoto (1986) contains several important references which defines the current state of non-adaptive multirate systems. However, a complete theoretical study for the identification, estimation and adaptive control of the multirate system described above has not been presented in the literature. Soderstrom (1980) did some work on this topic, but his work was restricted to first order systems and the intersampling values of the output were reconstructed by approximation. Zhang and Tomizuka (1988) proposed a multirate adaptive control algorithm, which is also restricted to first order systems and does not guarantee an asymptotically zero output estimation error or parameter convergence. In a context of inferential control Guilandoust et al. (1986, 1987a, 1987b, 1988) also assumed that the main plant output is sampled slower than the input and showed that its intersampling behaviour can be estimated with the aid of a secondary output, the measurement of which is available at the faster (input) sampling rate. Young and Mellichamp (1987) proposed an algorithm to identify the parameters and estimate the output of an approximate model of the multirate system using a virtual 'effective input'. The recent work of these two groups of authors (Guilandoust et al. 1986, 1987a, 1987b, 1988) Young and Mellichamp 1987) focused more on applications and did not include a complete theoretical foundation. Recently Scattolini (1988) presented a study of the self-tuning control of multirate systems. In his approach the issue of output intersampling estimation is not addressed and therefore the control performance index is defined only at the (slow) output sampling instants. Stability and convergence properties were given and proven only for the reduced case of equal rate systems.

Note that most of the multirate system work presented in this thesis was completed before most of the above cited publications appeared in the literature.

It is worth noting that the multirate systems dealt with in this thesis are different from those treated by the hybrid techniques of Gawthrop (1980), Elliott (1982) and Narendra et al. (1985). In this thesis the output is measured at a slower rate than the input and the parameter estimates of the system are updated at the slow output sampling rate. The hybrid techniques require that the input and output be measured simultaneously, i.e. at a same rate. The associated parameter identification and/or control calculation may be done at the same or slower rate. Note however that when the calculations are done at the slow interval, all the intersample values of the output (at the fast interval) are available from the measured data. In the approach used in this thesis it is necessary to estimate the intersample output values and prove that the estimates converge to the true (intersampled) outputs.

1.2 Parameter Convergence with Overparameterized Models

Parameter convergence is an important consideration in any process identification study. The parameter convergence problem when the identified model is not a minimal order representation of the plant has not been solved even for simple cases such as overparameterized deterministic autoregressive moving average (DARMA) models. Neverthless, it is common practice in identification applications to use nonminimal or overparameterized models when only an upper bound of the system order or delay is known, or when some specific structure is required for an estimation or control strategy.

One significant study which does apply to overparameterized systems is that by Ljung (1987). He showed that for prediction error type algorithms the criterion minimizing arguments converge, if the system signals are **informative**, to a set, each point of which results in the same input-output relations as that of the plant. (To use this result it is necessary to show that the estimated parameters for a particular prediction type algorithm asymptotically converge to the limiting set of the criterion minimizing arguments. This can be done for the recursive least squares algorithm (RLS) by applying the ordinary differential equation (ODE) technique (Ljung 1977, Ljung and Soderstrom 1983) subject to some regularity conditions.) However, Ljung's results are based on the very restrictive assumption of quasi-stationary data (A signal u(t) is quasi-stationary only if $\lim_{T\to\infty} \sum_{t=1}^{T} u^2(t)/T$ **exists** and is **finite** (Ljung 1987).), which does not apply to most signals encountered in applications and theoretical studies. Indeed an argument based on the quasi-stationary data assumption is more an identifiability than a convergence study, and is very close to arguing that 'if the parameter estimates converge to something (which cannot be a priori verified for non-quasi-stationary data) then they must converge to the true values.'

Recently parameter convergence with overparameterized models has received a great deal of attention in the literature. This includes the work of Goodwin at al (1985), Janecki (1987), Xia et al. (1987) and Heymann (1988). To ensure parameter convergence, all of these results implicitly require that the degree of overparameterization be known or that the common factor polynomial be uniquely fixed by the estimated model structure.

The general results obtained in this thesis apply not only to the identification of multirate system, but also to a very wide class of identification applications. The results do not need the quasi-stationary data assumption nor any a priori knowledge of the degree of overparametrization, and represent an essentially complete solution to the problem of parameter convergence with overparameterized models.

1.3 Outline

In this thesis each chapter is essentially a paper prepared for publication in a control journal. While focusing on the main theme of the thesis each chapter is self-contained and contributes at least one independent and significant result. Simulation examples are included to illustrate the algorithms and the theoretical results obtained.

The thesis consist of two major parts: multirate systems (Chapter 2-5) and parameter convergence (Chapter 6-7). Chapter 8 includes overall conclusions and recommendations for future work.

The first part of this thesis deals with the model formulation, open loop parameter identification, output estimation and adaptive control of multirate systems. Starting with a conventional discretized continuous model an equivalent multirate model is derived by two approaches, one of which is straightforward and mathematically simple (Chapter 3) while the other is of a more physical nature (Chapter 2). This equivalent model is convenient for parameter identification and for the estimation of the outputs $\{y(kT),\ k=01,2,\ldots\}$ using the available data $\{u(kT),y(kJT),\ k=0,1,2,\ldots\}$. Two algorithms, projection and least squares, are proposed. The convergence properties of the identification and estimation signals at the basic sampling instants are not available in the literature but are shown and formally proven in this thesis. The projection algorithm is not usually implemented in practice due to its poor convergence rate. However, since it is easier to analyze, results are first given for this algorithm (Chapter 2) and then extended to its least squares counterpart (Chapter 3). With the parameter and output estimates, $\hat{\theta}(kT)$ and $\hat{y}(kT)$, available at every basic sampling instants, little effort is needed to incorporate these estimates into most existing adaptive controllers. The first demonstration example (Chapter 4) is constrained adaptive control of a multirate system. Stability and convergence are formally proven for this closed loop system. The second example (Chapter 5) is adaptive inferential control. The original inferential estimation algorithm was presented by Guilandoust et al. (1986, 1987a, 1987b, 1988) The inferential problem is not the same as the multirate system described above since additional information, i.e. the secondary output $\{v(kT), k=0,1,2...\}$, is assumed available as measured data. This thesis provides a much clearer perspective than the previous work of Guilandoust et al. (1986, 1987a, 1987b, 1988) by rigorously reformulating the model and the estimation algorithm based on a theoretical framework. In addition, convergence properties of the inferential algorithm are also studied and a practical, simplified algorithm is proposed.

The second part of this thesis is concerned with parameter convergence with overparameterized models. Results are first presented (Chapter 6) for the simple but fundamental case of RLS parameter convergence with overparameterized models. These results are then extended (Chapter 7) to a broad class of identification algorithms for which a Lyapunov type function of parameter estimates can be found. The general results cover that of parameter convergence in the identification of multirate systems as well as those in the recent literature (e.g. Goodwin et al.

1985, Janecki 1987, Heymann 1988).

1.4 References

Araki M. and Yamamoto K., "Multivariable Multirate Sampled-Data Systems: State-Space Description, Transfer Characteristics, and Nyquist Criterion," IEEE Trans. Automat. Contr., vol. AC-31, pp. 145-154, Feb. 1986.

Crochiere R. E. and Rabina L. R., Multirate Digital Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1983.

Elliott H., "Hybrid Adaptive Control of Continuous Time Systems", IEEE Trans. Automat. Contr., vol. AC-27, pp. 419-426, April 1982.

Gawthrop P. J., "Hybrid Self-tuning Control", Proc. Inst. Elec. Eng., vol 127, part D, no. 5, pp. 973-998 Sept. 1980.

Goodwin G. C., Norton J. P., and Viswanathan M. N., "Persistency of Excitation for Nominimal Models of Systems Having Purely Deterministic Disturbance," IEEE Trans. Automat. Contr., vol. AC-30, pp. 589-594 June 1985.

Guilandoust M. T., Morris A. J. and Tham M. T., "Estimation and Control of Distillation Product Composition Using Tray Temperature Measurments," IFAC Symposium on Dynamics and Control of Chemical Reactors and Distillation Columns, pp. 203-208, Bournemouth, UK, 1986.

Guilandoust M. T., Morris A. J. and Tham M. T., "Adaptive Estimation in Inferential Process Control," Proc. American Control Conf., pp1743-1748, 1987a.

Guilandoust M. T., Morris A. J. and Tham M. T., "Adaptive Inferential Control," IEE Proc. vol. 134, Pt. D, NO. 3, pp. 171-179, May 1987b.

Guilandoust M. T., Morris A. J. and Tham M. T., "An Adaptive Estimation Algorithm for Inferential Control," Ind. Eng. Chem. Process Des Dev., 27, pp. 1658-1664, 1988.

Heymann M., "Persistency of Excitation Results for Structured Nonminimal Models," IEEE Trans. Automat. Contr., vol. AC-33, pp. 112-116, 1988.

Janecki D., "Remarks on Persistency of Excitation for Nonminimal Models of Systems Having Purely Deterministic Disturbances," IEEE Trans. Automat. Contr., vol. AC-32, pp. 1123-1125, 1987.

Ljung L., "Analysis of Recursive Stochastic Algorithms," IEEE Trans. Automat. Contr., vol. AC-22, pp. 551-575, 1977.

Ljung L., System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

Ljung L. and Soderstrom, Theory and Practice of Recursive Identification. Cambridge, MA: M.I.T. Press, 1983.

Narendra K. S., Khalifa I. H. and Annaswamy A. M., "Error Models for Stable Hybrid Adaptive Systems", IEEE Tran. Automat. Contr., vol. AC-30. pp. 339-347, Apr. 1985.

Scattolini R., "Self-tuning Control of Systems with Infrequent and Delayed Output Sampling," IEE Proc. vol. 135, Pt. D, No. 4, pp. 213-221, July 1988.

Soderstrom T., Methods and Applications in Adaptive Control, Springer Verlag, 1980.

Xia L., Moore J. B., and Gevers M., "On adaptive estimation and pole assignment of overparameterized system", Int. J. of Adaptive Contr. and Signal Processing, vol. 1, pp. 143-160, 1987.

Young R. E. and Mellichamp D. A., "Multirate Sampled Data Models Applied to Parameter Estimation and Adaptive Control," AICHE Annual Meeting, New York, NY, Nov. 1987.

Zhang Q. and Tomizuka M., "Multirate Sampling Adaptive Control and Its Application to Thermal Mixing Systems," Int. J. Control, vol. 47 No. 3, pp. 135-144, 1988.

2. OUTPUT ESTIMATION WITH MULTIRATE SAMPLING: PROJECTION ALGORITHM[1]

2.1 Introduction

In many discrete control applications the appropriate control interval is T but restrictions, such as the cycle time of a discrete measurement transducer, mean that the output values, y(t), are measured only at every J^{th} interval. For example, in chemical processing applications, composition analyzers such as gas chromatographs have a cycle time of several minutes compared to a desired control interval of say 30 seconds. If the control interval is increased to JT to match the availability of measurements then control performance deteriorates. Various approaches, such as inferential estimation (Guilandoust et al. 1986, 1987a, 1987b, 1988) which use measurements of secondary output values v(t) related to y(t), have been proposed to handle this problem. However, in this chapter an output estimation scheme is presented which produces estimates of the output, $\hat{y}(kT)$, based directly on measurements of input u(kT) and output y(kJT), where k=0,1,2,... . Convergence is proven for integer values of J and simulation results are presented which demonstrate convergence for J=10.

---

[1]A version of this chapter has been published: Lu W. and Fisher D. G., Int. J. Control, 1988, vol. 48, NO. 1, pp. 149-160.

2.1.1 Overview of Problem Formulation

There is a very large number of parameter estimation schemes plus output estimators, observers and filters which can be used to produce estimates of the process output y(t). However, most of these approaches assume that all the input/output data are available at every sampling instant.

The problem to be solved in this chapter is illustrated in Figure 2.1. A time series of known input values {u(kT),u(kT-T),...} and output values {y(kT),y(kT-JT),y(kT-2JT),...} are sent to a parameter estimation algorithm which produces estimates of the parameter vector, θ_0, of the assumed model. Based on the known values of the input and the estimated parameters, an output estimation algorithm produces estimates of the process output y(kT) and/or predicted values of the output. These estimated output values are shown to converge to the actual output and can be used for operator information, process characterization, adaptive control or time delay compensation. However, this chapter deals only with the formulation of the output estimation problem and the proof of its convergence properties.

A conventional, discrete input-output model is shown in Figures 2.2a with the input, u(k), and the output, y(k) both sampled at every interval, T. Here let T=1 to simplify the notation. As shown later, it is always possible to find an equivalent input-output model with the form shown in

Figure 2.2b where $B_J(q^{-1})$ is a polynomial in q^{-1}, and $A_J(q^{-J})$is a polynomial in q^{-J}. The available input-output measurement data $\{u(k),u(k-1),...,y(k),y(k-J),...\}$ are sufficient to identify $B_J(q^{-1})$, $A_J(q^{-J})$. The estimates of the output $\hat{y}(k)$ are then obtained by passing u(k) through a nonlinear filter which uses estimated parameters as shown in Figure 2.2c and 2.1.

Almost any identification algorithm can be used to identify the parameters in the model shown in Figure 2.2b. The proposed multirate output estimation algorithm has been formulated using the output error formulation (Landau 1979, Goodwin and Sin 1984). When the output error method is used the estimates of y(k) can be easily generated at every sampling instant without affecting the properties of the parameter estimates. As proven later, the error between the estimated output $\hat{y}(k)$ and y(k) asymptotically approaches zero or is at least bounded under very weak conditions.

The equivalence of the models shown in Figure 2.2a and 2b is shown in section 2.2 using a standard state space model approach. In section 2.3 the properties of the output estimation system based on an output error, projection approach are stated as theorems and formally proven. This is followed by a simulated example and the conclusions.

2.2 Process Model

For simplicity it is assumed that the process can be adequately represented by the continuous, single input and single output (SISO), linear, time-invariant system:

$$\dot{x}(t)=A_c x(t)+B_c u(t) \tag{2.1}$$

$$y(t)=C_c x(t) \tag{2.2}$$

The dimensions of matrices A_c, B_c and C_c are nxn, nx1 and 1xn respectively.

If the input u(t) is sampled with the unity rate, i.e. T=1, and a zero-order-hold is used, the input can be represented by

$$u(t)=u(k), \quad (k \le t < k+1), \quad (k=0,1,\ldots) \tag{2.3}$$

By integrating (2.1) from kJ+i-J to kJ+i,

$$x(kJ+i)=\exp[A_c J]x(kJ+i-J)+\int_{kJ+i-J}^{kJ+i}\exp[A_c(kJ+i-\tau)]B_c u(\tau)d\tau$$

$$(i=1,2,\ldots,J) \tag{2.4}$$

Making the variable change $t=\tau-(kJ+i-J)$ and defining

$$B^s:=\int_{s-1}^{s}\exp[A_c(J-t)]B_c dt, \quad (s=1,2,\ldots,J) \tag{2.5}$$

equation (2.4) can be written as

$$x(kJ+i)=\exp[A_c J]x(kJ+i-J)+B^J u(kJ+i-1)+$$
$$\ldots+B^1 u(kJ+i-J), \quad (i=1,2,\ldots,J) \tag{2.6}$$

By combining (2.2) and (2.6),

$$y(kJ+i)=C_c[q^J I-\exp(A_c J)]^{-1}[B^J u(kJ+i+J-1)+\ldots+B^1 u(kJ+i)]$$
$$=[B_J^J(q^{-J})/A_J^J(q^{-J})]u(kJ+i+J-1)+\ldots+$$
$$[B_J^1(q^{-J})/A_J^J(q^{-J})]u(kJ+i)$$
$$(i=1,2,\ldots,J) \tag{2.7}$$

where $B_J^s(q^{-J})/A_J^J(q^{-J}):=C_c[q^J I-\exp(A_c J)]^{-1}B^s$ $(s=1,2,\ldots,J)$. The order of the polynomials in q^{-J}, A_J^J, $B_J^J, \ldots, B_J^1$ is n. Since $[q^J I-\exp(A_c J]^{-1}$ is strictly proper, the first coefficients, i.e. the constant terms of $B_J^1, \ldots, B_J^J$ are zero.

Thus, equation (2.7) can be written as

$$A_J(q^{-J})y(kJ+i)=B_J(q^{-1})u(kJ+i) \qquad (i=1,2,\ldots,J) \qquad (2.8)$$

where $A_J(q^{-J})$ is an n^{th} order polynomial in q^{-J}, $B_J(q^{-1})$ is an $(nxJ)^{th}$ order polynomial in q^{-1} with its constant term being zero.

In summary, the resulting multirate model in Figure 2.2b is

$$A_J(q^{-J})y(k)=B_J(q^{-1})u(k) \qquad (2.9)$$

which is equivalent to the more commonly used equal rate model

$$A_1(q^{-1})y(k)=B_1(q^{-1})u(k) \qquad (2.10)$$

in the sense that both models describe the input/output relationship of the same state space model (2.1) and (2.2) with u(t) given in the form of (2.3).

Note that equation (2.9) can also be obtained by multiplying both sides of equation (2.10) by a polynomial $P(q^{-1})$ (Crochiere and Rabina 1983).

2.3 Estimation System

2.3.1 Process Model

Based on the derivation in section 2.2, the actual process (2.1), (2.2) and (2.3) can be represented by (dropping the subscript J for simplicity)

$$A(q^{-J})y(k)=B(q^{-1})u(k) \tag{2.11}$$

where

$$A(q^{-J})=1+a_1q^{-J}+a_2q^{-2J}+...+a_nq^{-nJ} \tag{2.12}$$

$$B(q^{-1})=b_1q^{-1}+b_2q^{-2}+...+b_mq^{-m} \tag{2.13}$$

$$m=nxJ$$

Alternatively, the input/output relationship can be described by

$$y(k)=\phi(k-1)^T\theta_0 \tag{2.14}$$

where

$$\phi(k-1)^T=[-y(k-J),-y(k-2J),...,-y(k-nJ), \quad u(k-1),u(k-2),...,u(k-m)] \tag{2.15}$$

and

$$\theta_0=[a_1,...,a_n,b_1,...,b_m]^T \tag{2.16}$$

2.3.2 Estimation Model

$$\hat{A}(k,q^{-J})\bar{y}(k)=\hat{B}(k,q^{-1})u(k) \tag{2.17}$$

where

$$\hat{A}(k,q^{-J})=1+\hat{a}_1(k)q^{-J}+...+\hat{a}_n(k)q^{-nJ} \tag{2.18}$$

$$\hat{B}(k,q^{-1})=\hat{b}_1(k)q^{-1}+...+\hat{b}_m(k)q^{-m} \tag{2.19}$$

Note that the number of parameters to be estimated in $\hat{A}$ and $\hat{B}$ is $n+m=n+nxJ$.

2.3.3 Estimation System

Define the following:

a) **a posteriori model output**

$$\bar{y}(k):=\bar{\phi}(k-1)^T\hat{\theta}(k) \tag{2.20}$$

where

$$\bar{\phi}(k-1)^T=[-\bar{y}(k-J),-\bar{y}(k-2J),\ldots,-\bar{y}(k-nJ),$$
$$u(k-1),u(k-2),\ldots,u(k-m)] \quad (2.21)$$

$$\hat{\theta}(k)=[\hat{a}_1(k),\ldots,\hat{a}_n(k),\hat{b}_1(k),\ldots,\hat{b}_m(k)]^T \quad (2.22)$$

Note that (2.20)–(2.22) are equivalent to (2.17)–(2.19).

b)a postericri model output error

$$\eta(k)=y(k)-\bar{y}(k) \quad (2.23)$$

c)a priori model output

$$\hat{y}(k):=\bar{\phi}(k-1)^T\hat{\theta}(k-1) \quad (2.24)$$

d) a priori model output error

$$e(k):=y(k)-\hat{y}(k) \quad (2.25)$$

e) generalized a posteriori output error

$$\bar{\eta}(k):=D(q^{-J})\eta(k) \quad (2.26)$$

where

$$D(q^{-J})=1+d_1q^{-J}+d_2q^{-2J}+\ldots+d_lq^{-lJ} \quad (2.27)$$

is a fixed moving average filter.

f) generalized a priori output error

$$\bar{v}(k):=e(k)+[D(q^{-J})-1]\eta(k) \quad (2.28)$$

The parameter identification algorithm is defined by

$$\hat{\theta}(kJ)=\hat{\theta}(kJ-1)+[\bar{\phi}(kJ-1)/(1+\bar{\phi}(kJ-1)^T\bar{\phi}(kJ-1))]\bar{v}(kJ) \quad (2.29)$$

$$\hat{\theta}(kJ+i)=\hat{\theta}(kJ), \qquad (i=1,2,\ldots,J-1) \quad (2.30)$$

$$\hat{\theta}(0)=\hat{\theta}_0 \quad (2.31)$$

The regressor $\bar{\phi}(k)$ has been defined in (2.21). Its initial value can be set by

$$\bar{\phi}(-1)=\phi(-1) \quad (2.32)$$

$$\bar{\phi}(-1+i)=\text{arbitrary}, \qquad i=1,2,\ldots,J-1 \quad (2.33)$$

2.3.4 Convergence at the Output Sampling Instants

The convergence properties of the signals generated in the estimation system are given in Theorem 2.1. These convergence properties are given only at the larger sampling instants. For example, it will be seen that the signal sequence $\{\hat{y}(kJ),\hat{y}(kJ+J),\hat{y}(kJ+2J),\ldots\}$ asymptotically converges to the process output sequence $\{y(kJ),y(kJ+J),y(kJ+2J),\ldots\}$.

Theorem 2.1:

Consider the algorithm (2.29)-(2.31) applied to process model (2.11); then, provided that the system $H(q^{-J})=D(q^{-J})/A(q^{-J})$ is very strictly passive:

(i)
$$\lim_{N\to\infty} \sum_{k=1}^{N} \eta(kJ)^2<\infty \tag{2.34}$$

which implies

$$\lim_{N\to\infty} \sum_{k=1}^{N} \bar{\eta}(kJ)^2<\infty \tag{2.35}$$

(ii)
$$\lim_{N\to\infty} \sum_{k=1}^{N} \bar{\phi}(kJ-1)^T\bar{\phi}(kJ-1)\bar{\eta}(kJ)^2<\infty \tag{2.36}$$

which implies

$$\lim_{N\to\infty} \sum_{k=1}^{N} \|\hat{\theta}(k)-\hat{\theta}(k-s)\|^2<\infty \tag{2.37}$$

where s is any finite integer.

(iii) If $\{u(k)\}$ is bounded, then

$$\lim_{k\to\infty} \bar{v}(kJ)=0 \tag{2.38}$$

$$\lim_{k\to\infty} |y(kJ)-\hat{y}(kJ)|=0 \tag{2.39}$$

Proof:

Careful examination of the estimation algorithm (2.29)–(2.31), shows that $\bar{y}(kJ),\bar{y}(kJ+1),\ldots,\bar{y}(kJ+J-1)$ do not interact with each other. In other words, only $\{\bar{y}(kJ),\bar{y}(kJ+J),\bar{y}(kJ+2J),\ldots\}$ are used in the parameter identification; and $\bar{y}(kJ+i),\bar{y}(kJ+i+J),\ \bar{y}(kJ+i+2J),\ldots\ (i=1,2,\ldots,J-1)$ are used in (2.24) only for the purpose of generating $\hat{y}(k)$ between the sampling time instants. Therefore, Theorem 2.1 is equivalent to Theorem 3.5.1 of Goodwin and Sin (1984). ∎

2.3.5 Convergence at the Input Sampling Instants

The convergence properties of the signals are now proven in Theorem 2.2 not only at the larger but also at the smaller sampling instants. The result is that $|\hat{y}(k)-y(k)|$ is at least bounded by a value proportional to the parameter estimation error and also by another value proportional to the increment of the input, and it approaches zero if the parameter estimation error or the input difference approaches zero.

Theorem 2.2:

Consider the algorithm (2.29)–(2.33) applied to process model (2.11); then; provided that the system $H(q^{-J})=D(q^{-J})/A(q^{-J})$ is very strictly passive:

(i) $$\|\hat{\theta}(k)-\theta_0\|^2 \le \|\hat{\theta}(0)-\theta_0\|^2 \qquad \forall k>0 \qquad (2.40)$$

(ii) There exists a positive number ϵ such that if $\{u(k)\}$ is bounded then $\|\hat{\theta}(0)-\theta_0\|^2<\epsilon$ implies:

(a) $$|y(k)-\hat{y}(k)| \le \delta_\eta \limsup_{k\to\infty} \|\hat{\theta}(k)-\theta_0\| + \Delta_\eta(k) \qquad \forall k>0 \qquad (2.41)$$

where $0<\delta_\eta<\infty$, $\Delta_\eta(k)$ is some sequence satisfying $\lim_{k\to\infty} \Delta_\eta(k)=0$.

(b) $$|y(k)-\hat{y}(k)| \le \delta_u \limsup_{k\to\infty} |u(k)-u(k-1)| + \Delta_u(k) \qquad \forall k>0 \qquad (2.42)$$

where $0<\delta_u<\infty$, $\Delta_u(k)$ is some sequence satisfying $\lim_{k\to\infty} \Delta_u(k)=0$.

(c) $$\lim_{k\to\infty} |y(k)-\hat{y}(k)|=0 \qquad (2.43)$$

provided that $\lim_{k\to\infty} \hat{\theta}(k)=\theta_0$ or $\lim_{k\to\infty} |u(k)-u(k-1)|=0$.

<u>Proof:</u>

Note that k only has discrete values and the sampling instant is T for u and JT for y. For simplicity of notation T=1 has already been assumed.

Step 1:

Define

$$b(k):=-\bar{\phi}(k-1)^T\tilde{\theta}(k) \qquad (2.44)$$

where

$$\tilde{\theta}(k):=\hat{\theta}(k)-\theta_0 \qquad (2.45)$$

From equations (2.11) and (2.17)

$$A(q^{-J})y(k)=B(q^{-1})u(k) \qquad (2.46)$$

$$\hat{A}(k,q^{-J})\bar{y}(k)-\hat{B}(k,q^{-1})u(k)=0 \qquad (2.47)$$

By combining (2.46) and (2.47) and introducing $A(q^{-J})\bar{y}(k)$,

$$A(q^{-J})[y(k)-\bar{y}(k)]=B(q^{-1})u(k)-A(q^{-J})\bar{y}(k)+$$
$$\hat{A}(k,q^{-J})\bar{y}(k)-\hat{B}(k,q^{-1})u(k)$$
$$=-\bar{\phi}(k-1)^{T}\tilde{\theta}(k) \qquad (2.48)$$

or

$$A(q^{-J})\eta(k)=b(k) \qquad (2.49)$$

Thus, by using the definition of $\bar{\eta}(k)$ (2.26)

$$A(q^{-J})\bar{\eta}(k)=D(q^{-J})b(k) \qquad (2.50)$$

Specifically

$$\bar{\eta}(kJ)+a_1\bar{\eta}(kJ-J)+\ldots+a_n\bar{\eta}(kJ-nJ)$$
$$=b(kJ)+d_1b(kJ-J)+\ldots+d_lb(kJ-lJ) \qquad (2.51)$$

where the input is b(0), b(J), b(2J),..., while the output is $\bar{\eta}(0)$, $\bar{\eta}(J)$, $\bar{\eta}(2J)$,..., and the input-output relation is very strictly passive by the assumption for $H(q^{-J})=D(q^{-J})/A(q^{-J})$.

Letting

$$V(k)=\tilde{\theta}(k)^{T}\tilde{\theta}(k) \qquad (2.52)$$

It follows after Goodwin and Sin (1984) that

$$V(kJ)=V(kJ-1)-2b(kJ)\bar{\eta}(kJ)-\bar{\phi}(kJ-1)^{T}\bar{\phi}(kJ-1)\bar{\eta}(kJ)^2 \qquad (2.53)$$

From (2.53), by noting V(kJ-1)=V(kJ-J)

$$V(kJ)=V(0)-2\sum_{j=1}^{k} b(jJ)\bar{\eta}(jJ)-\sum_{j=1}^{k} \bar{\phi}(jJ-1)^{T}\bar{\phi}(jJ-1)\bar{\eta}(jJ)^2 \qquad (2.54)$$

Since the system (2.51) with its zero initial states implied in (2.32)–(2.33) is very strictly passive,

$$\sum_{j=1}^{k} b(jJ)\bar{\eta}(jJ)\geq 0 \qquad \forall k>0 \qquad (2.55)$$

Thus from (2.52), (2.54) and (2.55)

$$0 \le V(kJ) \le V(0) \qquad \forall k>0$$

This gives (i) by considering (2.30).

Step 2:

By the very strictly passive assumption, the system (2.11) determined by θ_0 is asymptotically stable. Thus it is concluded that there exists $\epsilon_1>0$ such that: $\|\hat{\theta}(0)-\theta_0\|^2 \le \epsilon_1$ implies that the system formed by $\hat{\theta}(0)$ has all its eigenvalues inside the unit circle. So if $\|\hat{\theta}(0)-\theta_0\|^2 \le \epsilon_1$, by (i) it is deduced that the systems formed by $\hat{\theta}(k)$ ($\forall k>0$) have uniformly all their eigenvalues strictly inside the unit circle, and system (2.17) is uniformly asymptotically stable. Therefore, bounded $\{u(k)\}$ implies bounded $\{\bar{y}(k)\}$. This means that $\|\bar{\phi}(k)\| \le M_\phi$ ($\forall k \ge 0$) for some $0 \le M_\phi < \infty$. From (2.44) and (i)

$$|b(k)| \le \|\bar{\phi}(k-1)\| \|\tilde{\theta}(k)\| \le M_\phi \limsup_{k\to\infty} \|\tilde{\theta}(k)\| + \Delta_c(k) \tag{2.56}$$

where

$$\limsup_{k\to\infty} \|\tilde{\theta}(k)\| < \infty, \qquad \lim_{k\to\infty} \Delta_c(k) = 0$$

Since $A(q^{-J})$ is asymptotically stable, from (2.49) and (2.56)

$$|\eta(k)| \le \delta_\eta \limsup_{k\to\infty} \|\tilde{\theta}(k)\| + \Delta_\eta(k) \tag{2.57}$$

where

$$0 < \delta_\eta < \infty, \qquad \lim_{k\to\infty} \Delta_\eta(k) = 0$$

By considering (2.30) and (2.57)

$$|y(kJ+i)-\hat{y}(kJ+i)| = |y(kJ+i)-\bar{\phi}(kJ+i-1)^T\hat{\theta}(kJ+i-1)|$$

$$= |y(kJ+i)-\bar{\phi}(kJ+i-1)^T\hat{\theta}(kJ)|$$

$$=|y(kJ+i)-\bar{\phi}(kJ+i-1)^T\hat{\theta}(kJ+i)|$$

$$=|y(kJ+i)-\bar{y}(kJ+i)|$$

$$=|\eta(kJ+i)|$$

$$\le\delta_\eta \lim_{k\to\infty} \sup\|\tilde{\theta}(k)\|+\Delta_\eta(kJ+i) \quad (i=1,2,...,J-1) \quad (2.58)$$

(2.58) and (2.39) together give (a) of (ii).

Step 3:

From the definition of $\bar{\phi}(k)$ in (2.21):

$$\bar{\phi}(kJ-1)=[I-M_1(kJ)q^{-J}-...-M_n(kJ)q^{-nJ}]^{-1}M(q^{-1})u(kJ-1) \quad (2.59)$$

where $M_i(kJ)$ $(i=1,2,...,n)$ are (nxm)x(nxm) matrices with $-\hat{\theta}(kJ-iJ)^T$ as its i^{th} row and zero as its j^{th} $(1\le j\le(nxm)$ and $j\ne i)$ row.

$$M(q^{-1})=[0, \cdot \cdot \cdot, 0, 1, q^{-1}, \cdot \cdot \cdot, q^{-m+1}]^T \quad (2.60)$$

which is of dimension (nxm)x1. From (2.44)

$$b(kJ)=-\tilde{\theta}(kJ)^T\bar{\phi}(kJ-1)=P(kJ,q^{-1})u(kJ-1) \quad (2.61)$$

where

$$P(kJ,q^{-1})=-\tilde{\theta}(kJ)^T[I-M_1(kJ)q^{-J}-...-M_n(kJ)q^{-nJ}]^{-1}M(q^{-1})$$

$$:=P_n(kJ,q^{-1})/P_J(kJ,q^{-d}) \quad (2.62)$$

$P_n(kJ,q^{-1})$ is a polynomial in q^{-1}.

$$P_d(kJ,q^{-J})=1+P_{d1}(kJ)q^{-J}+P_{d2}(kJ)q^{-2J}+...+P_{dn}(kJ)q^{-nJ} \quad (2.63)$$

The parameters of $P(kJ,q^{-1})$ are functions of $\hat{\theta}(kJ),\hat{\theta}(kJ-J),...,\hat{\theta}(kJ-nJ)$.

By considering (2.30), it can be derived in general that

$$b(kJ+i)=P(kJ,q^{-1})u(kJ+i-1) \quad (i=0,1,...,J-1) \quad (2.64)$$

or

$$P_d(kJ,q^{-J})b(kJ+i)=P_n(kJ,q^{-1})u(kJ+i-1)\,(i=0,1,\ldots,J-1) \quad (2.65)$$

Step 4:

By referring to the proof of Theorem 3.5.1 in Goodwin and Sin (1984),

$$\lim_{k\to\infty} b(kJ)=0 \qquad \forall\{u(k)\} \quad (2.66)$$

Note that (2.66) was obtained without imposing any restriction on the initial values of $b(kJ)$. Therefore it can be easily concluded that for any $\{u(k)\}$ the solution of the following time variant system

$$P_J(kJ,q^{-J})b^*(kJ)=0 \quad (2.67)$$

asymptotically approaches zero, ie.

$$\lim_{k\to\infty} b^*(kJ)=0 \quad (2.68)$$

for any initial values.

Specifically the time invariant system

$$P_d^0(q^{-J})b^*(kJ)=0 \quad (2.69)$$

is asymptotically stable and has all its eigenvalues inside the unit circle, where $P_d^0(q^{-J})$ is obtained by replacing $\hat{\theta}$ in $P_d(kJ,q^{-J})$ with θ_0. This is true since $\lim_{k\to\infty}\hat{\theta}(k)=\theta_0$ also implies $\lim_{k\to\infty} b(kJ)=0$.

Then from (i) there exists $\epsilon_2>0$ such that $\|\hat{\theta}(0)-\theta_0\|^2\le\epsilon_2$ implies that for any $\{u(k)\}$ the system (2.67) has all its eigenvalues strictly inside the unit circle at any time instant k.

Thus system (2.67) is uniformly asymptotically stable.

This implies that in system (2.65) a bounded input produces a bounded output. Since $b(kJ+i)-b(kJ)$ $(i=0,1,...,J-1)$ is the output of system (2.65) if the input is $u(kJ+i-1)-u(kJ-1)$ $(i=0,1,...,J-1)$, considering $\lim_{k\to\infty} b(kJ)=0$ and the boundedness assumption of $u(k)$

$$|b(k)|=\delta_b \lim_{k\to\infty} \sup|u(k)-u(k-1)|+\Delta_b \tag{2.70}$$

where

$$0<\delta_b<\infty, \quad \lim_{k\to\infty} \sup|u(k)-u(k-1)|<\infty, \qquad \lim_{k\to\infty} \Delta_b(k)=0$$

Then an argument similar to step 2 can be used to give (b) of (ii). (c) of (ii) is obvious from (a) and (b).

Finally note that in (ii) $\epsilon=\min[\epsilon_1,\epsilon_2]$. ■

2.3.6 Remarks

1) In (ii) of Theorem 2.2 it is required that the initial parameter estimates, $\hat{\theta}(0)$, lie sufficiently close to the true parameters θ_0. This condition can easily be met by running the parameter estimation algorithm by itself prior to starting the output estimation. If the input $u(k)$ is persistently exciting for sufficient time during this initial period the desired $\hat{\theta}(0)$ can be obtained in a finite time. Once suitable starting parameters, $\hat{\theta}(0)$, are available then the output estimation can be started and it is no longer necessary to provide continuous persistent excitation. The value of ϵ in theorem 2.2-(ii) depends on the eigenvalue locations of the system in equation (2.11).

2) The results in (ii) of Theorem 2.2 are important from

both a practical and a theoretical point of view since they show that the output estimation error is bounded by the smaller value of the right sides of (2.41) and (2.42). If the excitation is rich then the parameter estimation error will become small and (2.41) will give a tight bound to the output estimation error. If the excitation is poor then the input u(k) is probably changing slowly and (2.42) will provide the tight bound instead.

3) The initial conditions (2.32) and (2.33) are used in order to guarantee property (i) of Theorem 2.2 under the weakest possible conditions. If $\hat{\theta}(k)$ converges to θ_0 then these initial conditions are not required.

4) To predict y(k) one step ahead, the a priori output model (2.24) can be used to give

$$\hat{y}(k+1)=\bar{\phi}^T(k)\hat{\theta}(k) \qquad (2.71)$$

since both $\bar{\phi}(k)$ and $\hat{\theta}(k)$ are available at time instant k when u(k) is known. In general, if {u(k)} is available then the output can be predicted d-steps ahead using

$$y_e(k+d)=\phi_e^T(k-1+d)\hat{\theta}(k) \qquad (2.72)$$

where

$$\phi_e^T(k-1+d)=(-\bar{y}_e(k-J+d),-\bar{y}_e(k-2J+d),\dots,-\bar{y}_e(k-nJ+d), u(k-1+d),u(k-2+d)\dots)$$

$$\bar{y}_e(\tau)=\bar{y}(\tau) \qquad \text{if } \tau\le k$$

$$=y_e(\tau) \qquad \text{if } \tau>k \qquad (2.73)$$

The convergence properties of y_e follow directly from those of $\bar{y}$, $\hat{y}$ and $\hat{\theta}$. Note that y_e gives the predicted values of y at every interval T and not just at the output sampling intervals JT. These predicted output values can be used in applications such as time delay compensation.

5) Since $\hat{\theta}(k)$, $\hat{y}(k)$ and the predicted values $y_e(k+d)$ are available at every u(k) sampling interval it should be relatively straightforward to formulate adaptive controllers with a control interval T. However, it would still be necessary to prove the closed loop stability of the adaptive system based on the properties given in Theorem 2.1 and 2.2.

6) The equation error method variant of the algorithm (2.29)-(2.31) can be easily derived based on the idea that only y(kJ), y(kJ+J), y(kJ+2J)...(not $\bar{y}(kJ)$, $\bar{y}(kJ+2J)$,...) are used in the parameter identification and $\bar{y}(kJ+i)$, $\bar{y}(kJ+i+J)$,... (i=1,2,...,J-1) are used for the purpose of generating estimates of y(k) at the smaller sampling instants. It is easy to prove that $\lim_{k\to\infty} |\hat{y}(k)-y(k)|=0$ if the persistent excitation condition is provided. Simulation examples (not presented here) verify this result.

2.4 Examples

Figure 2.3 illustrates the results obtained when the multirate output estimation scheme is applied to a second order system. The system was simulated by the equation:

$$(1-1.5q^{-1}+0.7q^{-2})y(k)=(q^{-1}+0.5q^{-2})u(k)$$

and u(k) was excited by a PRBS with magnitude 0.5. The output y(k) was sampled with a period J=10 times that of the input signal and, as shown in Figure 2.3, the estimated value $\hat{y}(k)$ converges to the true values after approximately 1000 intervals. Note however that this represents only 100 measured values of y(kJT) since J=10.

In most parameter estimation applications least squares algorithms give faster convergence than projection type algorithms. Figure 2.4 shows the result using the least squares variant of the algorithm (2.29)-(2.31). The output estimation error is eliminated after approximately 200 intervals, i.e. after 20 measured values of the output. In Chapter 3 a formal proof of convergence is completed for the least squares type algorithm.

2.5 Conclusions

1) It is proven that the output y(t) can be estimated at every control interval T even if y(t) is measured only at every JT instant, where J is an integer, and that the output estimation error $|\hat{y}(kT)-y(kT)|$ is at least bounded and normally approaches zero.

2) The output estimation $\hat{y}(kT)$ and the parameters $\hat{\theta}(kT)$ are available at every input sampling interval for use in adaptive controllers, time delay compensators etc.

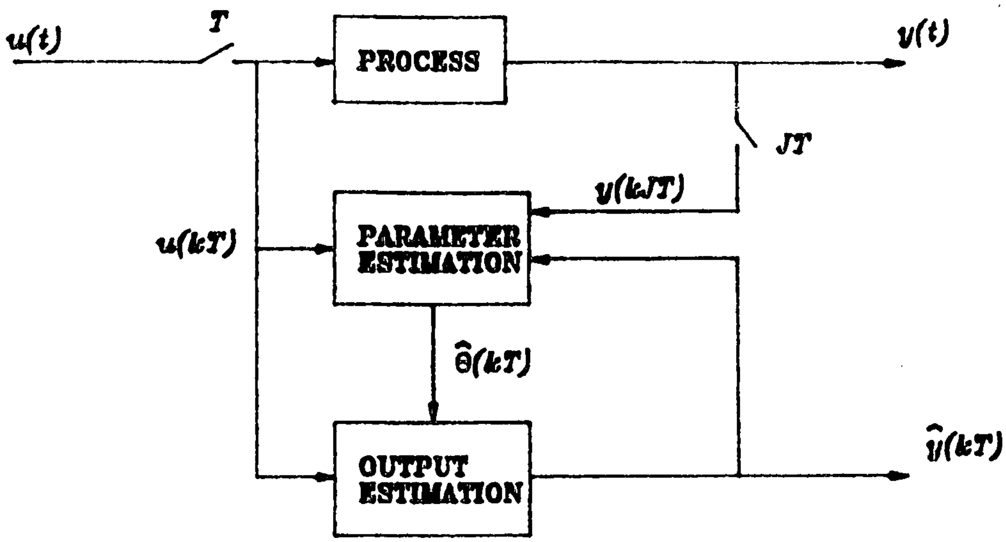

Figure 2.1: Block diagram representation of the multirate output estimation algorithm.

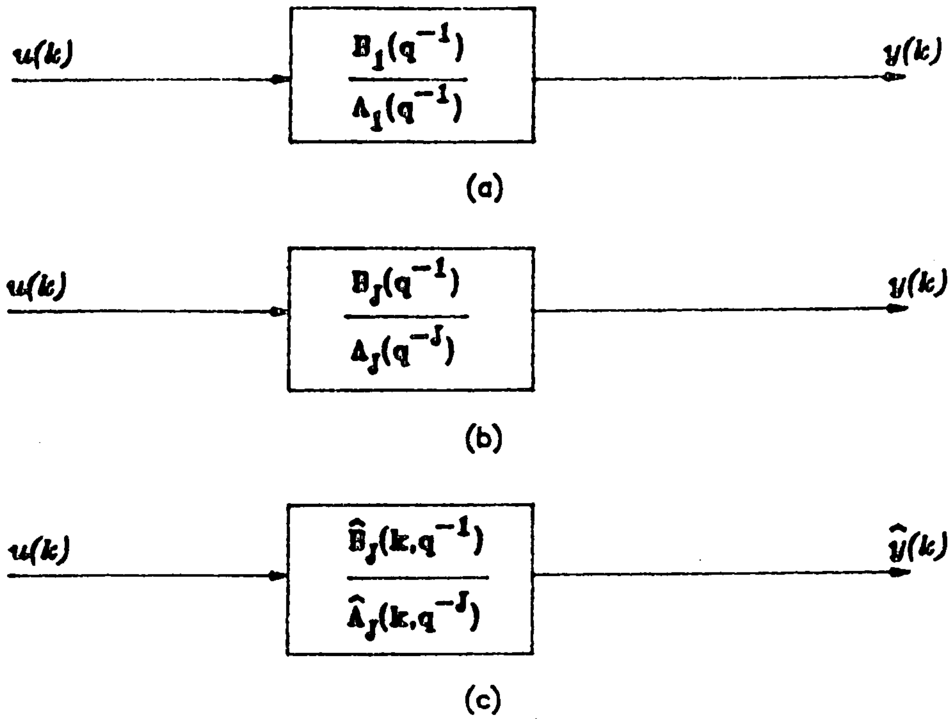

Figure 2.2: Output estimation models: (a) conventional discrete model, (b) equivalent model, (c) adaptive estimation model.

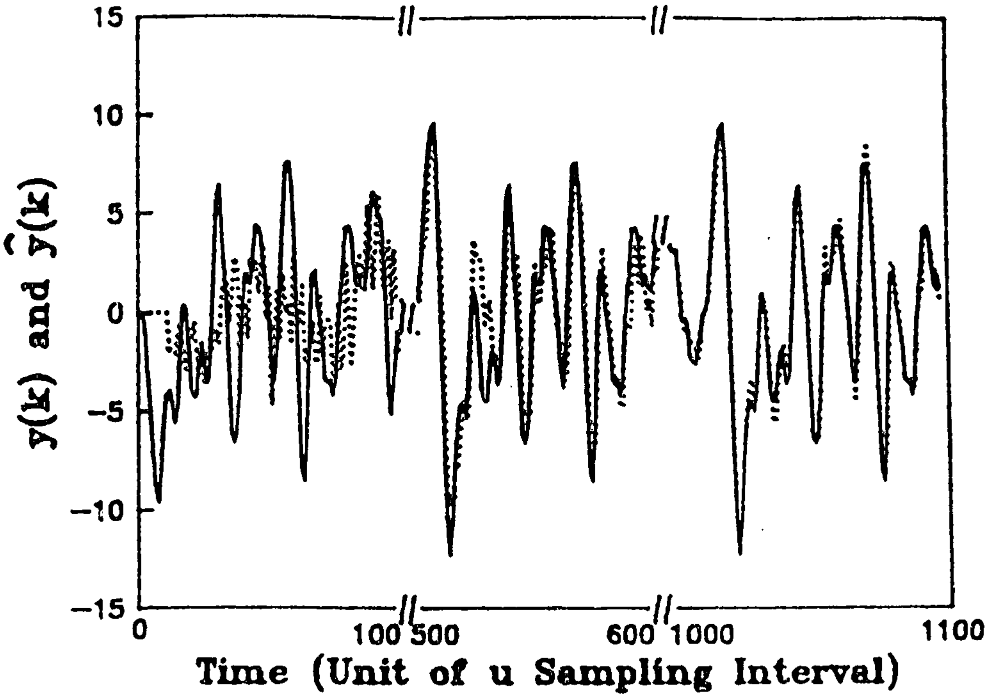

Figure 2.3: Output estimation with J=10 using a projection algorithm (solid line=y(k), dotted line=$\hat{y}(k)$).

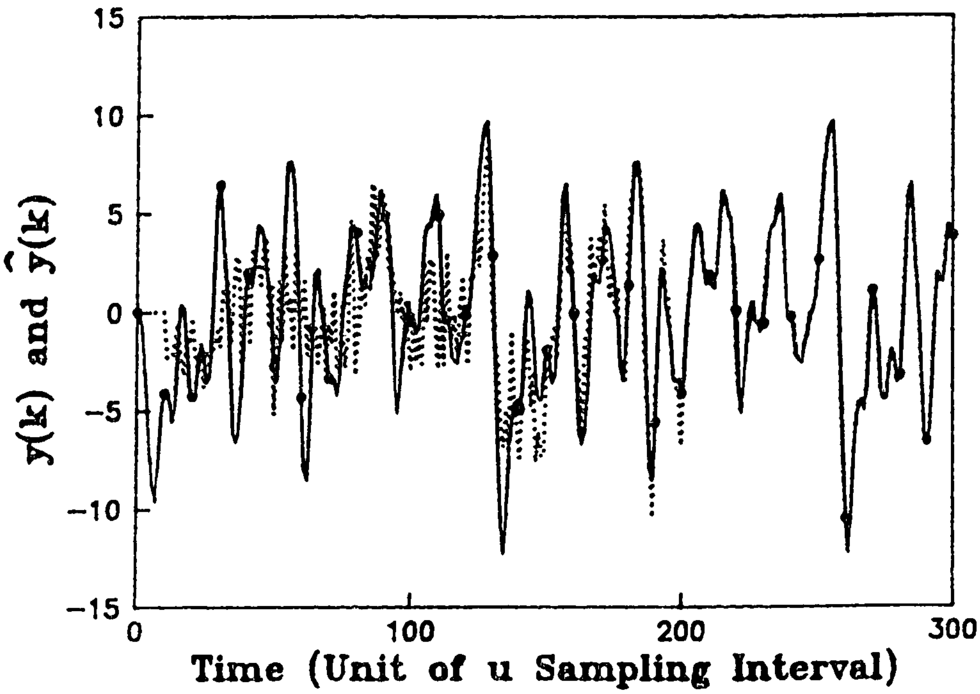

Figure 2.4: Output estimation with J=10 using a least squares algorithm (solid line=y(k), dotted line=$\hat{y}(k)$, circles=measured y(kJ).)

2.6 References

Crochiere R. E. and Rabina L. R., Multirate Digital Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1983.

Goodwin, G. C. and Sin K. S., Adaptive Filtering, Prediction and Control, Prentice Hall, New Jersey, 1984.

Guilandoust M. T., Morris A. J. and Tham M. T., "Estimation and Control of Distillation Product Composition Using Tray Temperature Measurments," IFAC Symposium on Dynamics and Control of Chemical Reactors and Distillation Columns, pp. 203-208, Bournemouth, UK, 1986.

Guilandoust M. T., Morris A. J. and Tham M. T., "Adaptive Estimation in Inferential Process Control," Proc. American Control Conf., pp1743-1748, 1987a.

Guilandoust M. T., Morris A. J. and Tham M. T., "Adaptive Inferential Control," IEE Proc. vol. 134, Pt. D, NO. 3, pp. 171-179, May 1987b.

Guilandoust M. T., Morris A. J. and Tham M. T., "An Adaptive Estimation Algorithm for Inferential Control," Ind. Eng. Chem. Process Des Dev., 27, pp. 1658-1664, 1988.

3. OUTPUT ESTIMATION WITH MULTIRATE SAMPLING: LEAST SQUARES ALGORITHM[2]

3.1 Introduction

This chapter deals with output estimation in applications where the output, $y(kJT)$, is sampled at a slower rate than the input, $u(kT)$, and where output estimates, $\hat{y}(kT)$, are required at the faster rate. This problem was studied in Chapter 2 using a projection algorithm. However simulation studies showed that the projection algorithm has a very slow convergence rate and hence is not recommended for practical application. In this chapter a least squares algorithm which has a much faster convergence rate is used to solve the same problem and the relative output estimation convergence properties are proven. These results are an extension of those developed by Landau (1979) and Goodwin and Sin (1984) for the case of equal sampling rates (i.e. for $J=1$). Essentially no additional assumptions or limitations are imposed for the solution with $J>1$ although the algorithm is more complex.

Assume that a linear time invariant process model is represented by $A_1(q^{-1})y(kT)=B_1(q^{-1})u(kT)$ (3.1)

where $A_1(q^{-1})=1+a_{11}q^{-1}+a_{12}q^{-2}+\ldots+a_{1n}q^{-n}=\prod_{i=1}^{n}[1-(\lambda_i q)^{-1}]$ (3.2)

and $B_1(q^{-1})=b_{11}q^{-1}+\ldots+b_{1n}q^{-n}$. (3.3)

---

[2] A version of this chapter has been accepted for publication: Lu W. and Fisher D. G., to appear, IEEE Trans. Automat. Contr., June, 1989.

The available online measurement data are u(kT) and y(kJT) (note that {y(kJT+iT), i=1,2,...,J-1} are <u>not available</u> from the measurement data), where k=0,1,2,... . T is the basic sampling period and J is any finite positive integer. Note that for notational simplicity T=1 in the following discussion.

The first step is to transform the process model (3.1) into a form, which can be identified from the available measurement data sequence. To do so, multiply both sides of (3.1) by (Crochiere and Rabina 1983)

$\prod_{i=1}^{n} [1+(\lambda_i q)^{-1}+\ldots+(\lambda_i q)^{2-J}+(\lambda_i q)^{1-J}]$ to obtain

$$A(q^{-J})y(k)=B(q^{-1})u(k) \tag{3.4}$$

where $A(q^{-J})=[1-(\lambda_1 q)^{-J}][1-(\lambda_2 q)^{-J}]\ldots[1-(\lambda_n q)^{-J}]$

$$=1+a_1 q^{-J}+a_2 q^{-2J}+\ldots+a_n q^{-nJ} \tag{3.5}$$

and $B(q^{-1})=b_1 q^{-1}+b_2 q^{-2}+\ldots+b_m q^{-m}$, (3.6)

$$m=Jxn. \tag{3.7}$$

<u>Remarks:</u>

1) If the coefficients of $A_1(q^{-1})$ and $B_1(q^{-1})$ are all real so are those of $A(q^{-J})$ and $B(q^{-1})$, since if λ is a root of A_1 so is its conjugate.

2) If A_1 is stable so is A.

3) In Chapter 2 an alternative method of transforming (3.1) to (3.4) was presented which provides greater physical

insight.

3.2 Estimation System

Assume that the process can be represented by (3.4) which can also be written as $y(k)=\phi(k-1)^T\theta_0$ (3.8)

where $\phi(k-1)^T=[-y(k-J),-y(k-2J),\ldots,-y(k-nJ),$

$$u(k-1),u(k-2),\ldots,u(k-m)] \quad (3.9)$$

and $\theta_0=[a_1,\ldots,a_n,b_1,\ldots,b_m]^T$. (3.10)

The estimation model is given by

$$\hat{A}(k,q^{-J})\bar{y}(k)=\hat{B}(k,q^{-1})u(k) \quad (3.11)$$

where $\hat{A}(k,q^{-J})=1+\hat{a}_1(k)q^{-J}+\hat{a}_2(k)q^{-2J}+\ldots+\hat{a}_n(k)q^{-nJ}$ (3.12)

and $\hat{B}(q^{-1})=\hat{b}_1(k)q^{-1}+\hat{b}_2(k)q^{-2}+\ldots+\hat{b}_m(k)q^{-m}$. (3.13)

Alternatively, write $\bar{y}(k)=\bar{\phi}(k-1)^T\hat{\theta}(k)$ (3.14)

where $\bar{\phi}(k-1)^T=[-\bar{y}(k-J),-\bar{y}(k-2J),\ldots,-\bar{y}(k-nJ),$

$$u(k-1),u(k-2),\ldots,u(k-m)] \quad (3.15)$$

and $\hat{\theta}(k)=[\hat{a}_1(k),\ldots,\hat{a}_n(k),\hat{b}_1(k),\ldots,\hat{b}_m(k)]^T$. (3.16)

Let

$$\eta(k)=y(k)-\bar{y}(k), \quad (3.17)$$

$$\hat{y}(k)=\bar{\phi}(k-1)^T\hat{\theta}(k-1), \quad (3.18)$$

$$e(k)=y(k)-\hat{y}(k) \quad (3.19)$$

$$\bar{\eta}(k)=D(q^{-J})\eta(k) \quad (3.20)$$

where $D(q^{-J})=1+d_1q^{-J}+d_2q^{-2J}+\ldots+d_lq^{-lJ}$ (3.21)

is a fixed moving averaging filter,

and $\bar{v}(k)=e(k)+[D(q^{-J})-1]\eta(k)$. (3.22)

3.2.1 Output Error Method

The parameter identification algorithm is given by

$$\hat{\theta}(kJ)=\hat{\theta}(kJ-J)+$$

$$P(kJ-2)\bar{\phi}(kJ-1)\bar{v}(kJ)/[1+\bar{\phi}(kJ-1)^T P(kJ-2)\bar{\phi}(kJ-1)] \quad (3.23)$$

$$\hat{\theta}(kJ+i)=\hat{\theta}(kJ), \qquad i=1,2,...,J-1 \quad (3.24)$$

$$P[(k+1)J-2]=P(kJ-2)-$$

$$P(kJ-2)\bar{\phi}(kJ-1)\bar{\phi}(kJ-1)^T P(kJ-2)$$

$$/[1+\bar{\phi}(kJ-1)^T P(kJ-2)\bar{\phi}(kJ-1)];$$

$$P(-2)>0 \quad (3.25)$$

$$\hat{\theta}(0)=\text{arbitrary}. \quad (3.26)$$

Note that between output sampling instants $\hat{\theta}(k)$ is not updated and it is not necessary to calculate $P(k)$. The initial values of $\bar{\phi}(k)$ can be set by

$$\bar{\phi}(-1)=\phi(-1) \quad (3.27)$$

$$\bar{\phi}(-1+i)=\text{arbitrary}, \qquad i=1,2,...,J-1 \quad (3.28)$$

where $\phi(-1)$ is available from the measurement data.

<u>Theorem 3.1:</u>

Consider the algorithm (3.23)-(3.28) applied to process model (3.4); then provided that the system $H(q^{-J})=[D(q^{-J})/A(q^{-J})-1/2]$ is very strictly passive:

(i) $\|\hat{\theta}(kJ)-\theta_0\|^2 \le \kappa_1 \|\hat{\theta}(0)-\theta_0\|^2 \qquad \forall k>0$ (3.29)

where $\kappa_1=\lambda_{max}[P(-2)^{-1}+\phi(-1)\phi(-1)^T]/\lambda_{min}[P(-2)^{-1}]$, and $\lambda_{max}(\cdot)$ and $\lambda_{min}(\cdot)$ represent maximum and minimum eigenvalues respectively.

(ii) There exists a positive number ϵ such that if $\{u(k)\}$ is bounded then $\|\hat{\theta}(0)-\theta_0\|^2<\epsilon$ implies:

(a) $|y(k)-\hat{y}(k)| \le$

$$\delta[\limsup_{k\to\infty}\|\hat{\theta}(kJ)-\theta_0\|][\limsup_{k\to\infty}|u(k)-u(k-1)|]+\Delta(k)$$

$$\forall k>0 \qquad (3.30)$$

where $\Delta(k)$ is a sequence satisfying $\lim_{k\to\infty} \Delta(k)=0$, δ and the two superior limits are finite positive numbers.

(b)
$$\lim_{k\to\infty} |y(k)-\hat{y}(k)|=0 \qquad (3.31)$$

provided that $\lim_{k\to\infty} \hat{\theta}(kJ)=\theta_0$ or $\lim_{k\to\infty} [u(k)-u(k-1)]=0$.

Proof:

Since some intermediary steps of this proof follow Goodwin and Sin (1984) they are omitted in this presentation but a full proof is included in Chapter 5, where a more general case is considered.

(i) The proof of (i) follows those of Corollary 3.5.1 and Lemma 3.3.6 in Goodwin and Sin (1984).

(ii) Define $b(k)=-\bar{\phi}(k-1)^T\tilde{\theta}(k)$ (3.32)

where $\tilde{\theta}(k)=\hat{\theta}(k)-\theta_0$. (3.33)

Combining (3.4) and (3.11) gives $A(q^{-J})\eta(k)=b(k)$ (3.34)

or equivalently

$$A(q^{-J})\eta(kJ+i)=-\bar{\phi}(kJ+i-1)^T\tilde{\theta}(kJ) \quad i=0,1,2,\dots,J-1. \qquad (3.35)$$

Here (3.24) is used. From (3.35)

$$A(q^{-J})[\eta(kJ+i)-\eta(kJ)]=-[\bar{\phi}(kJ+i-1)-\bar{\phi}(kJ-1)]^T\tilde{\theta}(kJ)$$
$$i=1,2,..,J-1. \qquad (3.36)$$

But $\bar{\phi}(kJ+i-1)^T=$

$$[F(kJ-J,q^{-1})q^{-J},F(kJ-2J,q^{-1})q^{-2J},\dots,F(kJ-nJ,q^{-1})q^{-nJ},$$
$$q^{-1},q^{-2},\dots,q^{-m}]u(kJ+i) \quad i=0,1,2,\dots,J-1 \qquad (3.37)$$

where $F(k,q^{-1})=\hat{B}(k,q^{-1})/\hat{A}(k,q^{-J})$.

By the very strictly passive assumption, $A(q^{-J})$ is asymptotically stable. Thus there exists $\epsilon_1>0$ such that: if

$\|\hat{\theta}(0)-\theta_0\|^2<\epsilon_1$, $\hat{A}(0,q^{-J})$ is also asymptotically stable. From (i) it is concluded that $\hat{A}(kJ,q^{-J})$ $\forall k\geq 0$ have uniformly all their eigenvalues inside a circle with its radius strictly less than one if letting $\epsilon=\epsilon_1/\kappa_1$. By a similar way and a result on the parameter estimates (Goodwin and Sin, 1984) it is not difficult to show that ϵ can be such that the system $F(k,q^{-1})$ is slowly time varing and global uniform exponentially stable uniformly in its parameters. Therefore by (3.37) bounded $\{u(k)\}$ implies bounded $\{\bar{\phi}(k)\}$ and that there exists $0<M<\infty$ such that

$$\|\bar{\phi}(kJ+i-1)-\bar{\phi}(kJ-1)\| \leq M|u(kJ+i)-u(kJ)| \qquad i=1,2,\ldots,J-1 \tag{3.38}$$

since $\bar{\phi}(kJ+i-1)-\bar{\phi}(kJ-1)$ versus $u(kJ+i)-u(kJ)$ $i=1,\ldots,J-1$ also satisfy (3.37).

Considering that $\lim_{k\to\infty}\eta(kJ)=0$ (easily obtained from Corollary 3.5.1 of Goodwin and Sin, 1984) and $A(q^{-J})$ is asymptotically stable, it is obtained from (3.36) and (3.38)

$$|\eta(kJ+i)|\leq\delta_1[\limsup_{k\to\infty}\|\tilde{\theta}(kJ)\|][\limsup_{k\to\infty}|u(kJ+i)-u(kJ)|]+\Delta(kJ+i) \qquad i=1,2,\ldots,J-1 \tag{3.39}$$

where $\Delta(kJ+i)$ $i=1,2,\ldots,J-1$ are some sequences satisfying $\lim_{k\to\infty}\Delta(kJ+i)=0$, δ_1 and the two superior limits are finite positive numbers. Note that for $i=1,2,\ldots,J-1$

$\limsup_{k\to\infty}|u(kJ+i)-u(kJ)| \leq (J-1)\limsup_{k\to\infty}|u(k)-u(k-1)|$.

Letting $\Delta(kJ)=|y(kJ)-\hat{y}(kJ)|$, $\delta=(J-1)\delta_1$ and noting $\lim_{k\to\infty}|y(kJ)-\hat{y}(kJ)|=0$ (also easily obtained from Corollary 3.3.6 of Goodwin and Sin, 1984) and

$$\eta(kJ+i)=y(kJ+i)-\bar{y}(kJ+i)=y(kJ+i)-\bar{\phi}(kJ+i-1)\tilde{\theta}(kJ+i)$$

$$=y(kJ+i)-\bar{\phi}(kJ+i-1)^T\tilde{\theta}(kJ)$$

$$=y(kJ+i)-\hat{y}(kJ+i) \qquad i=1,2,\ldots,J-1 \qquad (3.40)$$

it is obtained from (3.39) part (a) of (ii). Part (b) of (ii) is obvious. ∎

3.2.2 Remarks and Example

1) Replacing $\bar{\phi}(\cdot)$ by $\phi(\cdot)$ and $\bar{v}(kJ)$ by $y(kJ)-\phi(kJ-1)^T\hat{\theta}(kJ-J)$ in (3.23-3.24) results in the equation error variant of algorithm (3.23-3.26). Its convergence properties are similar to those of the output error method but are stated as: "there exists a positive number ϵ such that if $\{u(k)\}$ is bounded then $\|\hat{\theta}(0)-\theta_0\|^2\leq\epsilon$ implies:

$$|y(k)-\hat{y}(k)| \leq \delta\lim_{k\to\infty}\sup\|\hat{\theta}(kJ)-\theta_0\|+\Delta(k) \qquad \forall k>0$$

where $\Delta(k)$ is a sequence satisfying $\lim_{k\to\infty}\Delta(k)=0$, δ and the superior limit are finite positive numbers." The proof is essentially the same as that of Theorem 3.1, except that (3.35) is used instead of (3.36) since the $\lim_{k\to\infty}\eta(kJ)=0$ may not hold here.

2) In (3.30), the output estimation error is bounded by the product of the parameter estimation error and the input difference in the limit sense. This is a very meaningful result, since it shows that for the output error method $|y(k)-\hat{y}(k)|$ will almost always tend to zero since rich excitation of $u(k)$ implies small parameter estimation error $\|\tilde{\theta}(kJ)\|$ and lack of excitation via $u(k)$ normally (but not

necessarily) implies small $|u(k)-u(k-1)|$, i.e. in most cases at least one of the two terms in (3.30) will be very small.

3) Theorem 3.1 is a local rather than global result because the proof requires that $\|\tilde{\theta}(0)\|$ be "sufficiently small". This condition does not appear to be critical and can be achieved in practice by running the parameter estimation algorithm for a finite time with rich excitation. The output estimation can then be started and the requirement for continuous rich excitation dropped.

4) The equivalent plant model (3.4) is structuraly overparameterized in the sense that $A(q^{J})$ and $B(q^{-1})$ contain a monic common factor. Persistent excitation, u(k), of order 2m=2(Jxn) should be sufficient to guarantee parameter convergence if the original plant model (3.1) is irreducible. A formal proof of parameter convergence is presented in Chapter 4 and 7.

5) The very strictly passive condition is not due to the multirate sampling. It is inherent in the output error identification method. The conditions of Theorem 3.1 essentially demand nothing more or less than those of Corollary 3.5.1 in Goodwin and Sin (1984), which is applied to the single sampling rate case. Discussion of the relation between the very strictly passive condition and the original plant is included in (Landau 1979) and (Goodwin and Sin

1984).

6) It is worth noting that the output estimation technique proposed in this section is different from the Hybrid Techniques of Gawthrop (1980), Elliott (1982) and Narendra et al. (1985) which require that the input and output be measured simultaneously even though their parameter adaptation can be done at the same or slower sampling rate.

An Example:

Figure 3.1 illustrates the result obtained when the output error algorithm (3.23)-(3.28) is applied to the following system

$$(1-1.5q^{-1}+0.7q^{-2})y(k)=(q^{-1}+0.5q^{-2})u(k)$$

where u(k) is excited by a PRBS. The output y(k) was sampled with a period J=30 times that of the input signal. Figure 3.1 shows that the output estimation error has essentially been eliminated after time 3900T (T=1). Note that this involves only 130 measured values of y(kJT). The parameter estimates also converge to the true parameters of the equivalent model (3.4). Use of the equation error algorithm gave comparable results (not included), but the use of a projection algorithm for parameter estimation gave significantly slower convergence (Chapter 2).

3.3 Conclusions

It was shown that, even when the output is sampled J times slower than the input, the output intersampling estimation error is bounded and 'almost always' converges to zero. This formulation and proof provides a basis for multirate sampling applications in time-delay compensation, inferential and adaptive control.

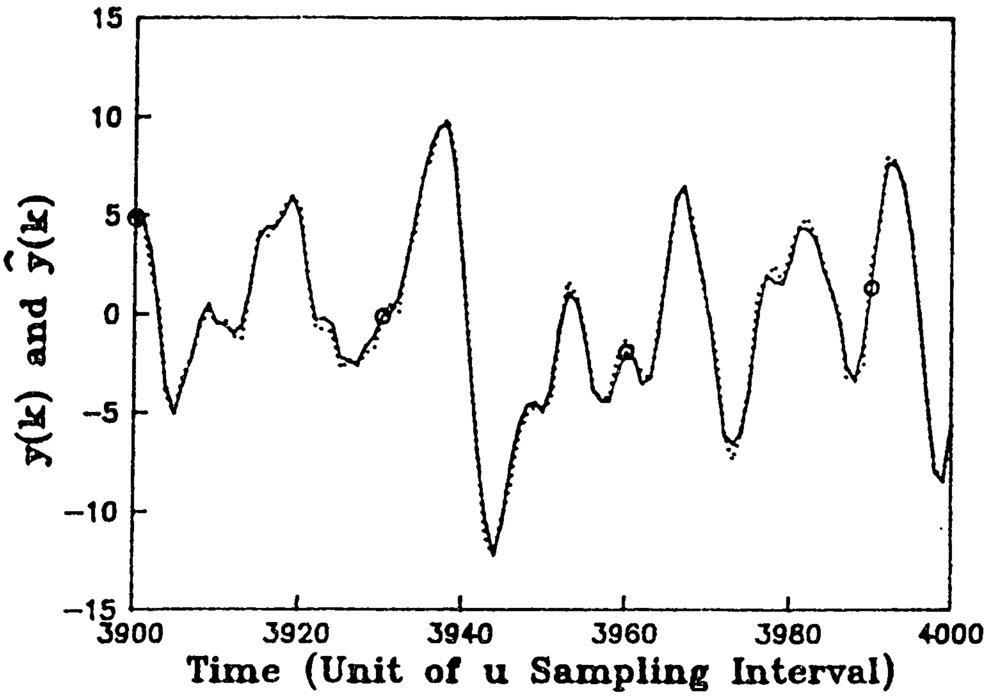

Figure 3.1: Output estimation with J=30 (solid line=y(k); dotted line=$\hat{y}(k)$; circles=measured y(kJ)).

3.4 References

Crochiere R. E. and Rabina L. R., Multirate Digital Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1983.

Elliott H., "Hybrid Adaptive Control of Continuous Time Systems", IEEE Trans. Automat. Contr., vol. AC-27, pp. 419-426, Apr. 1982.

Gawthrop P. J., "Hybrid Self-tuning Control", Proc. Inst. Elec. Eng., vol 127, part D, no. 5, pp.973-998, Sept. 1980.

Goodwin, G. C. and Sin K. S. Adaptive Filtering, Prediction and Control, Prentice Hall, New Jersey, 1984.

Landau, I. D., Adaptive Control: the Model Reference Approach, Marcel Dekker, Inc., New York, 1979.

Narendra K. S., Khalifa I. H. and Annaswamy A. M., "Error Models for Stable Hybrid Adaptive Systems", IEEE Tran. Automat. Contr., vol. AC-30. pp. 339-347, Apr.1985.

4. MULTIRATE CONSTRAINED ADAPTIVE CONTROL[3]

4.1 Introduction

In discrete computer control applications, it is usually desirable to use a sampling rate which is consistent with the controlled system dynamics. However, the choice of sampling rate may be limited by the availability of the output measurement. For example, in chemical processing applications, composition analyzers such as gas chromatographes have a cycle time of say 5 to 10 minutes compared to a desired control interval of say 0.5 to 1 minute. If the control interval is increased to match the availability of measurements then control performance deteriorates significantly. The deterioration of control performance with the increased control interval is due to the fact that between measurements the controller works in open loop mode and its action is kept constant. This means that the control system cannot respond faster than the open loop system between the slow measurements and only low frequency inputs to the system, eg. slow setpoint changes, can be compensated.

However, instead of being constant, or following some other fixed action such as using a higher order hold, the

----

[3]A version of this chapter has been accepted for publication: Lu W., Fisher D. G. and Shah S. L., Int. Journal of Control, 1989.

control action between the slow measurements can be based on a process model. A good model would permit the use of a faster control sampling rate, and hence the controlled system should be able to respond to higher frequency inputs. This simple but fundamental idea is applied in this chapter to improve the servo control of linear time invariant systems with unknown parameters.

An adaptive control law is derived which accommodates two sampling rates: a slower one (due to slow measurements) for the output and a faster one (due to desirable control interval) for the manipulated input. The control action is based on an estimated discrete model with the faster sampling rate. The estimated discrete model is updated whenever the output measurement is available. The control law aims at minimizing a quadratic function of the output tracking error with dynamic weighting on the input at each fast input sampling instant. When there are upper and lower bounds imposed on the control action, as is always the case in real applications, the minimization is performed in the constrained region without extra computation. The stability and convergence properties of the control system are stated and formally proven.

The block diagram of the proposed multirate constrained adaptive control system is shown in Figure 4.1. Note that since only servo control is considered the disturbances are not considered and assumed to be zero.

4.1.1 Previous Work

This work is a natural extension of the work presented in Chapters 2 and 3 on the identification and output estimation of multirate systems and is intended to show how these previous results can be applied to formulate and analyze the multirate counterparts of adaptive control algorithms developed for equal rate sampling systems.

Textbooks such as the one by Goodwin and Sin (1984) include a comprehensive treatment of adaptive control for the case of equal rate sampling. However, very little work has been published on the adaptive control of multirate systems. Minimum variance control (or d+1 step ahead control for deterministic systems) of a first order system was studied by Soderstrom (1980), who also used a slower output sampling rate. The first control law considered by Soderstrom coincides with the usual minimum variance controller only at the (slower) output sampling instants. The second control law he considered coincides with the usual minimum variance controller only in an approximate sense, since the missing output measurements are replaced by interpolated and extrapolated values. Convergence properties were not formally given or proven. Zhang and Tomizuka (1988) proposed a multirate adaptive control algorithm, which is also restricted to first order systems and does not guarantee an asymptotically zero output estimation error or parameter convergence. Recently, a self-tuning controller for multirate systems was presented by Scattolini (1988). In his approach the performance index is defined <u>only</u> at the (slow) output sampling instants. To minimize the performance index function the control action is determined subject to the minimal change of control action at the (fast) input sampling instants. Stability and convergence properties were given and proven only for the reduced case of equal rate systems.

4.2 ADAPTIVE PREDICTIVE CONTROL SYSTEM

The multirate, constrained adaptive control system is derived in this section.

4.2.1 Plant Model

Consider the continuous, SISO, linear time-invariant system

$$\dot{x}(t)=A_c x(t)+B_c u(t) \qquad (4.1)$$

$$y(t)=C_c x(t-dT) \qquad (4.2)$$

The dimension of matrices A_c, B_c and C_c are nxn, nx1, and 1xn respectively, and the system is assumed to be a minimal representation of the plant. Assume that the delay d is an integer and T is selected as the input sampling period, i.e.

$$u(t)=u(kT), \qquad kT\le t<kT+T,\ k=0,1,\ldots \qquad (4.3)$$

For simplicity and without loss of generality let T=1. Then the continuous model (4.1)-(4.2) can be represented by the equivalent discrete model

$$A_1(q^{-1})y(k)=B_1(q^{-1})u(k-d), \qquad (4.4)$$

where

$$A_1(q^{-1})=1+ a_{11}q^{-1}+ a_{12}q^{-2}+\ldots+ a_{1n}q^{-n}, \qquad (4.5)$$

$$B_1(q^{-1})=b_{11}q^{-1}+\ldots+ b_{1n}q^{-n} \qquad (4.6)$$

and $A_1(q^{-1})$, $B_1(q^{-1})$ are coprime since the system of (4.1) and (4.2) is minimal.

If the output sampling period is greater than that of the input, e.g. only y(kJT) is available from the measurement data, where J is a positive integer, then multiplying both sides of (4.4) by some polynomial $C_1(q^{-1})$ provides an equivalent form of (4.4) (cf. Chapter 3)

$$A(q^{-J})y(k)=B(q^{-1})u(k-d), \tag{4.7}$$

where

$$A(q^{-J})=1+ a_1q^{-J}+ a_2q^{-2J}+\ldots+ a_nq^{-nJ}, \tag{4.8}$$

$$B(q^{-1})=b_1q^{-1}+b_2q^{-2}+ \ . \ . \ . \ +b_mq^{-m} \tag{4.9}$$

and $\quad m=Jxn. \qquad (4.10)$

The equivalent model (4.7) is convenient for parameter identification and output estimation in multirate applications. Output estimates can be generated not only at the output sampling instants, 0, JT, 2JT,..., but also at the input sampling instants, 0, T, 2T, ... and it can be proven (cf. Chapters 2 and 3) that the estimated output values converge to the real output values. This helps to improve the performance at the J-1 intersampling instants of the output y(kJ).

4.2.2 Identification and Output Estimation

Assume that the plant is represented by model (4.7), i.e.

$$y(k)=\phi(k-1)^T\theta_0 \tag{4.11}$$

where

$$\phi(k-1)^T=[-y(k-J),\ -y(k-2J),\ \ldots,\ -y(k-nJ),$$
$$u(k-d-1),\ u(k-d-2),\ \ldots,\ u(k-d-m)], \tag{4.12}$$

$$\theta_0^T=[a_1, \ldots, a_n, b_1, \ldots, b_m] \qquad \text{and } m=Jxn. \qquad (4.13)$$

If the estimates of θ_0 are given by

$$\hat{\theta}(k)^T=[\hat{a}_1(k), \ldots, \hat{a}_n(k), \hat{b}_1(k), \ldots, \hat{b}_m(k)]$$

$$k=0, 1, 2, \ldots \qquad (4.14)$$

then the a posteriori output estimate is generated by

$$\bar{y}(k)=\bar{\phi}(k-1)^T\hat{\theta}(k) \qquad (4.15)$$

where $\bar{\phi}(k-1)^T=[-\bar{y}(k-J), -\bar{y}(k-2J), \ldots, -\bar{y}(k-nJ),$

$$u(k-d-1), u(k-d-2), \ldots, u(k-d-m)] \qquad (4.16)$$

with the initial values

$\bar{\phi}(-1)=\phi(-1)$ (available from the measurement data)

and (4.17)

$$\bar{\phi}(i-1)=\text{arbitrary} \qquad \text{for } i=1, 2, \ldots, J-1. \qquad (4.18)$$

The a priori output estimate is given by

$$\hat{y}(k)=\bar{\phi}(k-1)^T\hat{\theta}(k-1). \qquad (4.19)$$

Define:

the a posteriori model output error

$$\eta(k)=y(k)-\bar{y}(k), \qquad (4.20)$$

the a priori model output error

$$e(k)=y(k)-\hat{y}(k), \qquad (4.21)$$

the generalized a posteriori output error

$$\bar{\eta}=D(q^{-J})\eta(k) \qquad (4.22)$$

and the generalized a priori output error

$$\bar{v}(k)=e(k)+[D(q^{-J})-1]\eta(k), \qquad (4.23)$$

where

$$D(q^{-J})=1+d_1q^{-J}+d_2q^{-2J}+\ldots+d_lq^{-lJ} \qquad (4.24)$$

is a fixed moving average filter.

The parameter adaptation law is given by

$$\hat{\theta}(kJ)=\hat{\theta}(kJ-J)+$$

$$P(kJ-2)\bar{\phi}(kJ-1)\bar{v}(kJ)/[1+\bar{\phi}(kJ-1)^{T}P(kJ-2)\bar{\phi}(kJ-1)] \quad (4.25)$$

$$\hat{\theta}(kJ+i)=\hat{\theta}(kJ), \qquad i=1,2,\ldots,J-1 \quad (4.26)$$

$$P[(k+1)J-2]=P(kJ-2)-P(kJ-2)\bar{\phi}(kJ-1)\bar{\phi}(kJ-1)^{T}P(kJ-2)$$

$$/[1+\bar{\phi}(kJ-1)^{T}P(kJ-2)\bar{\phi}(kJ-1)], \quad (4.27)$$

$$P(-2)>0 \quad (4.28)$$

$$\hat{\theta}(0)=\text{arbitrary}. \quad (4.29)$$

Note that the parameter estimator (4.25-4.29) is an output error method (Goodwin and Sin 1984). An equation error method (Goodwin and Sin 1984) for parameter estimation can be derived using a similar approach. However the two methods result in different intersampling output estimation error behaviour and it is show. in Chapter 3 that the output error method can have a smaller bound on the error when the input, $u(k)$, is not rich in excitation and slowly changing.

4.2.3 Adaptive Predictive Control Law

At time k, the predicted value of $y(k+d+1)$ is given by

$$y_e(k+d+1|k)=\phi_e^{T}(k+d|k)\hat{\theta}(k) \quad (4.30a)$$

where

$$y_e(\tau|k)=\phi_e^{T}(\tau|k)\hat{\theta}(k), \quad (4.30b)$$

$$\phi_e^{T}(\tau|k)=[-\bar{y}_e(\tau-J+1|k),-\bar{y}_e(\tau-2J+1|k),\ldots,$$

$$-\bar{y}_e(\tau-nJ+1|k),u(k),\ u(k-1),\ \ldots,\ u(k-m+1)]. \quad (4.31)$$

$$\bar{y}_e(\tau|k)=\bar{y}(\tau) \qquad \text{if } \tau\le k$$

and $$\bar{y}_e(\tau|k)=y_e(\tau|k) \qquad \text{if } \tau>k. \quad (4.32)$$

At time k the controller output $u(t)$ is calculated such that the performance index

$$I_e(k+d+1|k)\equiv[y_e(k+d+1|k)-y^*(k+d+1)]^2+[Q(q^{-1})u(k)]^2 \tag{4.33}$$

is minimized subject to the constraint

$$u_{min}\le u(k)\le u_{max}, \tag{4.34}$$

where

$$Q(q^{-1})=Q_N(q^{-1})/Q_D(q^{-1}), \tag{4.35}$$

$$Q_N(q^{-1})=q_{N0}+q_{N1}q^{-1}+\ldots+q_{Np}q^{-p}, \tag{4.36}$$

$$Q_D(q^{-1})q_{D0}+q_{D1}q^{-1}+\ldots+q_{Dp}q^{-p}, \tag{4.37}$$

$y^*(k+d+1)$ is the assumed known set point value. Since $I_e(k+d+1|k)$ is quadratic in $u(k)$, the minimization is such that one of the following three cases is satisfied

(i) $\partial I_e(k+d+1|k)/\partial u(k)=0$ and $u_{min}\le u(k)\le u_{max}$ (4.38)

(ii) $u(k)=u_{min}$ (4.39)

(iii) $u(k)=u_{max}$. (4.40)

This leads to the following control law

$$u'(k)=\{\hat{b}_1(k)[y^*(k+d+1)-(y_e(k+d+1|k)-\hat{b}_1(k)u(k))]\}/\{\hat{b}_1(k)^2+Q(q^{-1})q_{N0}\}, \tag{4.41}$$

$$\begin{aligned} u(k)&=u_{min} && \text{if } u'(k)<u_{min} \\ &=u'(k) && \text{if } u_{min}\le u'(k)\le u_{max} \\ &=u_{max} && \text{if } u_{max}<u'(k) \end{aligned} \tag{4.42}$$

Note that in (4.41) $y_e(k+d+1|k)-\hat{b}_1(k)u(k)$ contains only past data $y(k),y(k-1),\ldots,u(k-1),u(k-2),\ldots$. The control law (4.41-4.42) is an (output error method variant) extension of that given in Remark 6.3.3. of Goodwin and Sin (1984). The extension is in the sense that multirate sampling and input weighting are included. To improve the information content and robustness of the overall control system, a small amplitude external excitation signal may be added to the calculated control signal, i.e.

$$u(k)=u_{min}+v(k) \qquad \text{if } u'(k)<u_{min}$$
$$=u'(k)+v(k) \qquad \text{if } u_{min}\le u'(k)\le u_{max}$$
$$=u_{max}+v(k) \qquad \text{if } u_{max}<u'(k) \qquad (4.43)$$

4.3 CONVERGENCE ANALYSIS

The convergence properties of the multirate adaptive control system derived in the previous section are developed and proven below.

4.3.1 Parameter and Output Estimation

Here a result presented in Chapter 3 is first stated in Theorem 4.1 for completeness, then parameter convergence is proven in Theorem 4.2 under a persistent excitation condition. (Note that the identified model (4.7) is a nonminimal model, i.e. $A(q^{-J})$ and $B(q^{-1})$ contain the common factor polynomial $C(q^{-1})$, and that conditions for parameter convergence are not as obvious as for minimal model cases.) These two theorems provide convergence properties for the parameter and output estimation algorithm described in section 2.2.

Theorem 4.1:

Consider the output estimation and parameter identification algorithm (4.15,4.19) and (4.25-4.29) applied to the plant model (4.7). Then, provided that the system $H(q^{-J})\equiv[D(q^{-J})/A(q^{-J})-1/2]$ is very strictly passive:

(i) $$\lim_{N\to\infty}\sum_{k=1}^{N}\eta(kJ)^2<\infty, \qquad (4.44)$$

which implies

$$\lim_{N\to\infty} \sum_{k=1}^{N} \bar{\eta}(kJ)^2<\infty \quad \text{and} \quad \lim_{k\to\infty}|y(kJ)-\bar{y}(kJ)|=0. \tag{4.45}$$

(ii)
$$\lim_{N\to\infty} \sum_{k=1}^{N} \bar{\phi}(kJ-1)^T P(kJ-2)\bar{\phi}(kJ-1)\bar{\eta}(kJ)^2<\infty \tag{4.46}$$

which implies

$$\lim_{N\to\infty} \sum_{k=1}^{N} \|\hat{\theta}(kJ)-\hat{\theta}[(k-s)J]\|^2<\infty \tag{4.47a}$$

or equivalently (cf. (4.26))

$$\lim_{N\to\infty} \sum_{k=1}^{N} \|\hat{\theta}(k)-\hat{\theta}(k-s)]\|^2<\infty \tag{4.47b}$$

where s is any finite integer.

(iii)
$$\|\hat{\theta}(kJ)-\theta_0\|^2\le\kappa_1\|\hat{\theta}(0)-\theta_0\|^2 \qquad \forall k\ge 0 \tag{4.48a}$$

or equivalently (cf.(4.26))

$$\|\hat{\theta}(k)-\theta_0\|^2\le\kappa_1\|\hat{\theta}(0)-\theta_0\|^2 \qquad \forall k\ge 0 \tag{4.48b}$$

where $0<\kappa_1<\infty$.

(iv)
$$\tilde{\theta}(kJ)^T P(kJ-2)^{-1}\tilde{\theta}(kJ)<c<\infty \qquad \forall k\ge 0, \tag{4.49a}$$

or equivalently

$$\tilde{\theta}(kJ+i)^T P(kJ-2)^{-1}\tilde{\theta}(kJ+i)<c<\infty$$
$$\forall k\ge 0 \text{ and } i=0,1,\ldots J-1, \tag{4.49b}$$

where c is a constant and $\tilde{\theta}(k)\equiv\hat{\theta}(k)-\theta_0$.

If $\{u(k)\}$is bounded then:

(v)
$$\lim_{k\to\infty}|y(kJ)-\hat{y}(kJ)|=0 \tag{4.50}$$

and

(vi) there exists a positive number $\epsilon_7<\infty$ such that $\|\hat{\theta}(0)-\theta_0\|^2<\epsilon_7$ implies

$$|y(k)-\bar{y}(k)|$$
$$\le\delta[\limsup_{k\to\infty}\|\hat{\theta}(kJ)-\theta_0\|][\limsup_{k\to\infty}|u(k)-u(k-1)|]+\Delta(k)$$
$$\forall k\ge 0 \tag{4.51}$$

where $\Delta(k)$ is a sequence satisfying $\lim_{k\to\infty}\Delta(k)=0$, δ and the two limit superiors are positive finite numbers.

Proof: See Chapters 3 and 5. ■

Theorem 4.2:

Subject to the same conditions as Theorem 4.1 the implication (i)⇒(ii)⇒(iii)⇒(iv) holds for the following statements:

(i) $$c_{1max}I > \sum_{i=k}^{k+r} U(i)U(i)^T > c_{1min}I \qquad \forall k\geq 0, \qquad (4.52)$$

where $r<\infty$ is an integer, $\infty>c_{1max}>c_{1min}>0$ and

$$U(k-1)=[u(k-1),u(k-2),\ldots,u(k-2m)]^T. \qquad (4.53)$$

(ii) $$c_{2max}I > \sum_{i=k}^{k+r} \phi(iJ-1)\phi(iJ-1)^T > c_{2max}I$$

$$\forall k \text{ sufficiently large,} \qquad (4.54)$$

and

$$c_{2max}I > \sum_{i=k}^{k+r} \bar{\phi}(iJ-1)\bar{\phi}(iJ-1)^T > c_{2max}I$$

$$\forall k \text{ sufficiently large,} \qquad (4.55)$$

where $\infty>c_{2max}>c_{2min}>0$.

(iii) $$\lim_{k\to\infty} \lambda_{min}(P(kJ-2)^{-1})=\infty \qquad (4.56)$$

where $\lambda_{min}(\cdot)$ means the minimum eigenvalue.

(iv) $$\lim_{k\to\infty} \hat{\theta}(k)=\theta_0. \qquad (4.57)$$

Proof:

(i)⇒(ii): It is easy to check that the equivalent model (4.7) is uniquely determined by model (4.4) and the integer J, i.e. the common factor polynomial, $C_1(q^{-1})$, is uniquely fixed by the model structure. This implies that a state space realization of $\phi(\cdot)$ versus $u(\cdot)$ as the input is completely reachable (Heymman 1988). Therefore (4.52) implies (4.54) (Anderson and Johnson 1982), which yields (4.55) since $\lim_{k\to\infty}\|\bar{\phi}(kJ-1)-\phi(kJ-1)\|=0$ by (i) of Theorem 4.1.

(ii)⇒(iii): This is obvious since

$$P(kJ-2)^{-1}=P(-2)^{-1}+\sum_{i=0}^{k-1} \bar{\phi}(iJ-1)\bar{\phi}(iJ-1)^T.$$

(iii)⇒(iv): Combining (4.56) and (4.49b) gives this implication. ∎

Note that for the more general cases, where A_1 and B_1 are not coprime the result of Chapters 6 and 7 can be used to determine the limiting set of $\hat{\theta}(k)$. Theorem 4.2 given above would then be a special case of this more general result. However, the above proof is more concise and provides a direct link with other publications.

4.3.2 Adaptive Control System

Theorems 4.1 and 4.2 only give convergence properties for the parameter and output estimation algorithm. When the values of parameter and output estimates are incorporated into in the adaptive control law (4.41,4.43), it is also desirable to prove the stability and convergence properties of the overall control system. Note that with $u_{min}=-\infty$ and $u_{max}=\infty$ a stability and convergence proof is difficult to carry out. The main difficulty is that the Key Technical Lemma (Lemma 6.1 of Goodwin and Sin, 1984) is not applicable in this multirate case because the adaptive control law (4.41,4.43) allows the control, u(t), to change between two successive output measurements. Since a constraint with $u_{min}>-\infty$ and $u_{max}<\infty$ is more typical of practical control applications a convergence proof without the constraint is not discussed here.

The following Theorem 4.3 shows that under given conditions the adaptive control system guarantees boundedness for all its internal signals and asymptotically behaves like a reference control system, defined by the corresponding control system with known plant parameters and known output values at the fast (input) sampling instants.

The following assumptions will be used in Theorem 4.3:

Assumptions:

(1) The time delay d+1 is known.

(2) $d+1 \le J$.

(3) The order of the plant is known.

(4) $H(q^{-J})$ is very strictly passive.

(5) There exists $0<M<\infty$ such that $\forall k$, the 'frozen' transfer function $1/[\hat{b}_1(k)^2+Q(q^{-1})q_{N0}]$ is asymptotically stable and has a steady state gain bounded by -M from below and M from above.

(6) $\{y^*(k)\}$ is bounded and $y^*(k+d+1)$ is known at time instant k.

(7) $-\infty<u_{min}<u_{max}<\infty$ and $\{v(k)\}$ is bounded.

Remarks:

(1)Note that to use the adaptive control algorithm (4.41,4.43) Assumptions (1-3) are not necessary. Upper bounds of the plant order and the delay are all that is required. For simplicity these assumptions are used in Theorem 4.3.

(2) The very strictly passive condition in Assumption (4) is not due to the multirate sampling. It is inherent in the output error identification method (Landau 1979). This condition is not required if an equation error method is

used. Discussion of the guidelines for selecting the D filter of (4.23) and the relationship between the very strictly passive condition and the original plant is included in Goodwin and Sin (1984) and Landau (1979).

(3) From (4.47b) $[\hat{b}_1(k)^2+Q(q^{-1})q_{N0}]^{-1}$ is slowly time varing. Therefore Assumption (5) guarantees that the linear, slowly time varing system $[\hat{b}_1(k)^2+Q(q^{-1})q_{N0}]^{-1}$ is BIBO (bounded input produces bounded output). Note that for a corresponding adaptive control system with known output values at the fast (input) sampling instants and $u_{min}=-\infty$ and $u_{max}=\infty$, i.e. an adaptive equal (fast) rate control system without input constraints, the offline choice of Q has to make $B_1(q^{-1})+A_1(q^{-1})Q(q^{-1})q_{n0}/b_{11}$ stable (Tsiligiannis and Svoronos 1986), which requires some a priori knowledge about the plant parameters. Since Assumption (5) only involves one parameter, $\hat{b}_1(k)$, and $\hat{b}_1(k)^2>0$, $\forall k$, the offline choice of Q to meet the assumption is much easier. For example $Q=\lambda$ always satisfies the assumption for any constant scalar constant $\lambda\neq 0$. To satisfy Assumption (5), similar simple rules can be deduced for a first or second order Q, etc.

<u>Theorem 4.3:</u>

Subject to Assumptions (1-7) the control law (4.41,4.43) when applied to the system (4.11) yields:

(i) $\{u(k)\}$ and $\{y(k)\}$ are bounded sequences.

(ii) If $\|\hat{\theta}(0)-\theta_0\|^2<\epsilon_7$, where ϵ_7 is as defined in Theorem 4.1, then $\{\bar{y}_e(\tau|k),\tau\leq k\}$ and $\{y_e(k+d+1|k)\}$ and $\{u'(k)\}$ are all bounded and the following reference performance index

$$I(k+d+1|k) \equiv [(y(k+d+1)-y^*(k+d+1))]^2+[Q(q^{-1})u(k)]^2 \quad (4.58)$$

is approximated by $I_e(k+d+1|k)$ with the following error

$$|I_e(k+d+1|k)-I(k+d+1|k)| \le \alpha \lim_{k\to\infty} \sup \|\hat{\theta}(kJ)-\theta_0\| + \Delta_1(k) \quad (4.59)$$

where $\Delta_1(k)$ is a sequence satisfying $\lim_{k\to\infty} \Delta_1(k)=0$, α and $\lim_{k\to\infty} \sup\|\hat{\theta}(kJ)-\theta_0\|$ are two finite numbers.

(iii) If $\{v(k)\}$ is selected such that (4.52), the excitation condition for $u(k)$ holds, then

$$\lim_{k\to\infty} \hat{\theta}(kJ)=\theta_0 \quad (4.60)$$

which implies from (4.59) that

$$\lim_{k\to\infty} |I_e(k+d+1|k)-I(k+d+1|k)|=0 \quad (4.61)$$

<u>Proof:</u>

(i) Conclusion (i) is obvious since $u(t)$ is constrained and the system is asymptotically stable by the very strictly passive assumption.

(ii) Since $d+1 \le J$

$$\bar{y}_e(\tau|k)=\bar{y}(\tau) \qquad \forall\ \tau \le k \text{ and } k=1,2,... \quad (4.62)$$

and $$\phi_e(k+d|k)=\bar{\phi}(k+d). \quad (4.63)$$

Therefore by (vi) of Theorem 4.1 and (i) of this theorem $\{\bar{y}_e(\tau|k), \tau \le k\}$ and

$$\|\phi_e(k+d|k)\|=\|\bar{\phi}(k+d)\| \le n \sup_k |\bar{y}(k)| + m \sup_k |u(k)| \quad (4.64)$$

are bounded, which implies that $\{y_e(k+d+1|k)\}$ (cf. (4.30)) is also bounded since $\hat{\theta}(k)$ is from (4.47b) or (4.48b). From (4.41) the boundedness of $\{u'(k)\}$ is assured by Assumption (5), and the boundedness of $y^*(k)$, $y_e(k+d+1|k)$, $u(k)$ and $\hat{\theta}(k)$. To prove (4.59) note that from (4.33) and (4.58)

$$|I_e(k+d+1|k)-I(k+d+1|k)|$$

$$=|y_e(k+d+1|k)^2-y(k+d+1)^2-2y^*(k+d+1)(y_e(k+d+1|k)-y(k+d+1))|$$

$$=|[y_e(k+d+1|k)+y(k+d+1)-2y^*(k+d+1)](y_e(k+d+1|k)-y(k+d+1))| \quad (4.65)$$

and

$$|y_e(k+d+1|k)-y(k+d+1)|$$
$$=|\bar{\phi}(k+d)^T\hat{\theta}(k)-\phi(k+d)^T\theta_0|$$
$$\le\|\bar{\phi}(k+d)-\phi(k+d)\|\|\hat{\theta}(k)\|+\|\phi(k+d)\|\|\hat{\theta}(k)-\theta_0\|$$
$$\le m_1 \limsup_{k\to\infty} \|\hat{\theta}(k)-\theta_0\|+\Delta_e(k)=m_1 \limsup_{k\to\infty} \|\hat{\theta}(k)-\theta_0\|+\Delta_e(k) \quad (4.66)$$

where $0<m_1<\infty$, $\lim_{k\to\infty} \Delta_e(k)=0$, and $\|\bar{\phi}(k+d)-\phi(k+d)\|\le n \sup_k|\bar{y}(k)-y(k)|$ and (4.51) are used in (4.66). Substituting (4.66) into (4.65) and noting that $y_e(k+d+1|k)$, $y(k)$, $y^*(k)$ are all bounded yields (4.59).

(iii) Conclusion (iii) is obvious. Note that conditions for v(k) under which excitation conditions for u(k) are satisfied have been discussed by Moore (1987). ∎

It is important to realize that conclusion (ii) of Theorem 4.3 is a local stability and convergence result because of the condition $\|\hat{\theta}(0)-\theta_0\|^2<\epsilon_7$. Since in practical applications an adaptive control system should not be switched on with poor initial parameters this condition does not appear to be especially critical and can be achieved (cf. Theorem 4.2) in practice by running the parameter estimation algorithm for a finite time with sufficient excitation before the adaptive control law (4.41, 4.43) is switched on.

4.4 A SIMULATION EXAMPLE

This section presents a simulation example which shows that the proposed multirate, constrained, adaptive control system behaves asymptotically like the reference control system, which minimizes, at each control sampling instant k, the reference performance index $I(k+d+1|k)$ instead of $I_e(k+d+1|k)$.

The simulated continuous plant system (cf. model (4.1-4.2)) is represented by

$$\dot{x}(t)=\begin{bmatrix} 0.7504 & -0.8668 \\ 1.2383 & -1.1071 \end{bmatrix}x(t)+\begin{bmatrix} 0.5821 \\ -0.6562 \end{bmatrix}u(t) \tag{4.67}$$

$$y(t)=[\ 1 \quad 0.5\ ]x(t-\tau_d) \tag{4.68}$$

With $\tau_d=0$, the open loop response of the plant is shown in Figure 4.2 for a unit step input. Common design guidelines suggest that the discrete sampling interval for the system should be approximately T=1 if the controlled system is required to have faster dynamics than the open loop system. For T=1, the corresponding discrete model (cf. model (4.4)) is

$$(1-1.5q^{-1}+0.7q^{-2})y(k)=(q^{-1}+0.5q^{-2})u(k-d) \tag{4.69}$$

where $d=\tau_d/T=\tau_d$.

Assume that the output, y, can be measured only every 10 control intervals. It is obvious from Figure 4.2 that a control interval of JT=10 would not be expected to give good control using conventional equal (slow) rate algorithms. However, applying the adaptive control law (4.41,4.43) with

$J=10$, $v(k)\equiv 0$, $Q(q^{-1})=0$, $u_{min}=-1$ and $u_{max}=1$, i.e. d+1 step ahead constrained adaptive control, to system (4.67-4.68) yields the responses shown in Figure 4.3 (a-d) for d=0 and Figure 4.4 (a-d) for d=12. All figures are plotted using continuous time t, instead of discrete time k, to show the intersampling behaviour. The identification and estimation algorithm is on from time t=0 and the parameter estimates start from zero initial values. A conventional PI controller with a control interval of 10 is used from t=0 to t=300 and a small amount of noise is added to the control action at every interval (T=1) to yield additional excitation during this initial period, after which this noise is removed. The adaptive control law is turned on at t=300. The set point is given by

$$y^*(t+d+1)=5 \qquad 100i\le t<100i+100 \quad i=0,2,4$$
$$=-5 \qquad 100i\le t<100i+100 \quad i=1,3,5 \qquad (4.70)$$

Figure 4.5 (a-d) shows the response of the corresponding reference control system with known plant parameters and known output values at every interval, T=1.

Remarks:

(1) For $0\le t<300$, with a control interval of 10 and a PI controller it is difficult to achieve a satisfactory response (cf. Figure 4.2a where tuning was done by trial and error). When the system has a delay of 12 the response is even worse (cf. Figure 4.4a). Since fast output measurements are assumed for the reference control system in Figure 4.5a the PI controller gives a much better performance than in

Figures 4.3a and 4.4a. Note that in Figures 4.3b and 4.4b the small ripple in u(t) between the control intervals for $0 \le t < 300$ is due to the small excitation or probing signal.

(2) For $t \ge 300$ there is essentially no difference between the multirate adaptive control system and the reference control system (compare Figure 4.3 versus Figures 4.5). This indicates that the parameter estimates (not plotted) were sufficiently close to the true plant parameters even though only 30 measured values of the output were available during the initial identification period $0 \le t < 300$. After t=500 the adaptive system shown in Figure 4.3behaves essentially the same as the 'ideal' (fast and equal rate sampling, known parameters) reference system shown in Figure 4.5.

(3) When the system has a delay of 12 the predictive structure of the proposed multirate adaptive control still ensures excellent performance (cf. Figure 4.4). This simulation result also shows that the control algorithm (4.41,4.43) is applicable with $d > J$ even though $d \le J$ is assumed in the proof of Theorem 4.3.

(4) The response of the reference system with a single, slow sampling rate of 10 is not shown. However, as may be expected, since between the sampling instants the control u(t) is constant and the response is open loop (cf. the open loop response shown in Figure 4.2), the output response is definitely not as good as Figure 4.3 or Figure 4.5 (Note that in Figures 4.3 and 4.5 the output tracking errors, for $300 \le t$, are essentially settled to zero only 5 or 6 steps

after every setpoint change.).

(5) Control in the presence of external disturbances is not discussed. Obviously information about the disturbances enters the system only via the slowly measured output variable. Therefore even though the control sampling rate is faster than the output sampling rate the proposed controller still cannot compensate disturbance components beyond the frequency range determined by the output sampling rate.

(6) Note that the output estimation y_e represented in (30) relies explicitly only on u(k) and $\hat{\theta}(k)$, i.e. in Figure 4.1 the line from y(kJT) to the output estimation model block is not connnected. Therefore the output feedback is only indirectly via the parameter estimate $\theta(k)$. Since the simulation example is linear and allows an exact model match, this 'weak' form output feedback does not cause problems. However, in most control applications an exact model match is impossible and it should be more practical to connnect the line from y(kJT) to the output estimation model block by some adhoc method to strengthen the output feedback and/or to improve the output estimation and prediction.

4.5 CONCLUSIONS

A multirate, constrained adaptive servo controller is derived for applications where the output sampling rate is restricted to a rate J times slower than the desired control (input sampling) rate.

Local stability and convergence properties of the adaptive control system are given and formally proven. It is shown that the the multirate, constrained adaptive control system asymptotically approaches the performance of the analogous fast and equal sampling rate system with known, constant system parameters.

Performance is excellent, as shown by the simulated results. Since almost all practical applications have input constraints $u_{min} \leq u(k) \leq u_{max}$ and many applications have restrictions on the output sampling rate, the proposed control algorithm should be of significant practical interest.

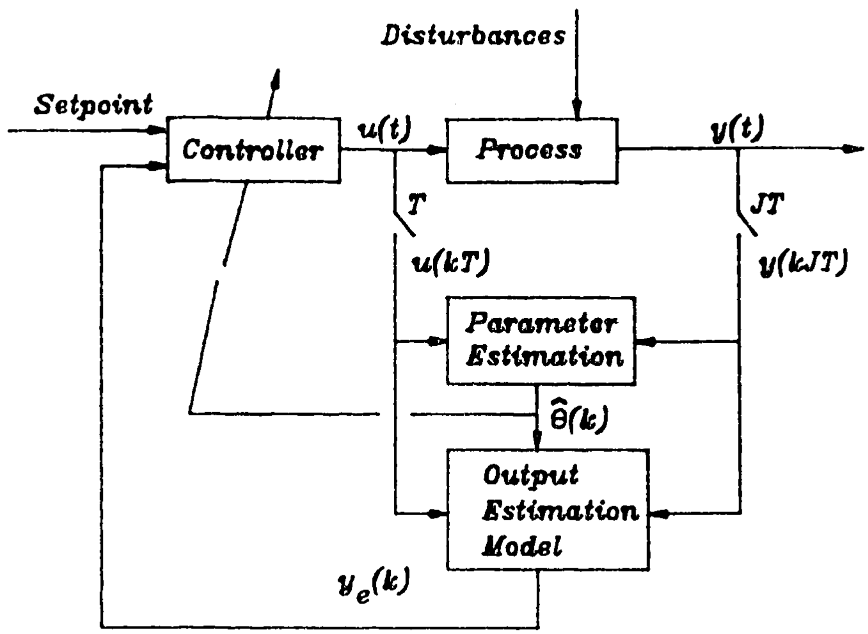

Figure 4.1: Multirate constrained adaptive control system.

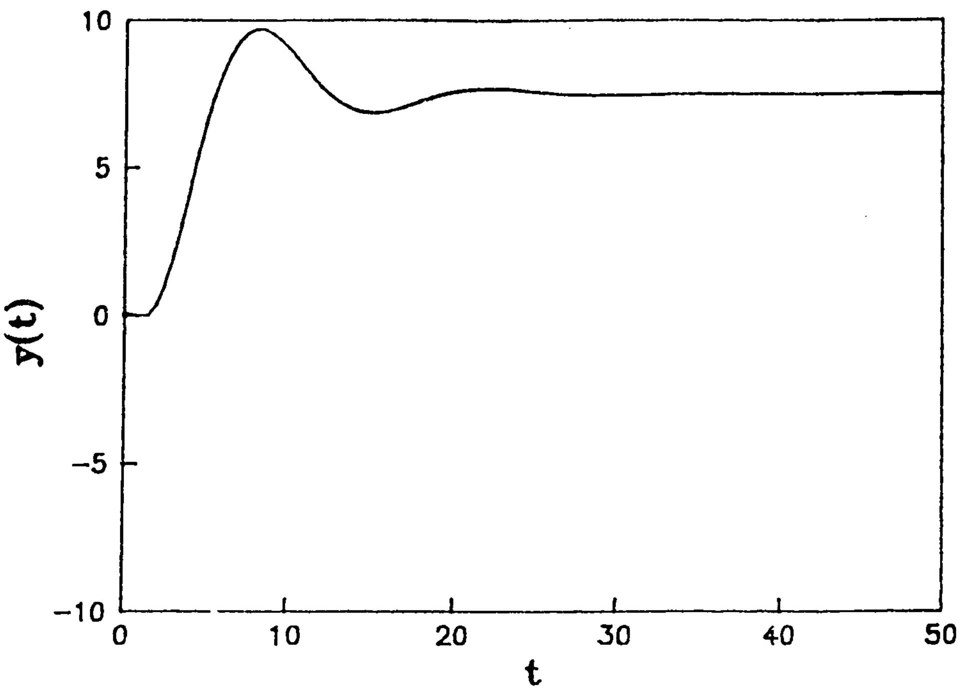

Figure 4.2: Open loop output response of continuous model to a unit step input.

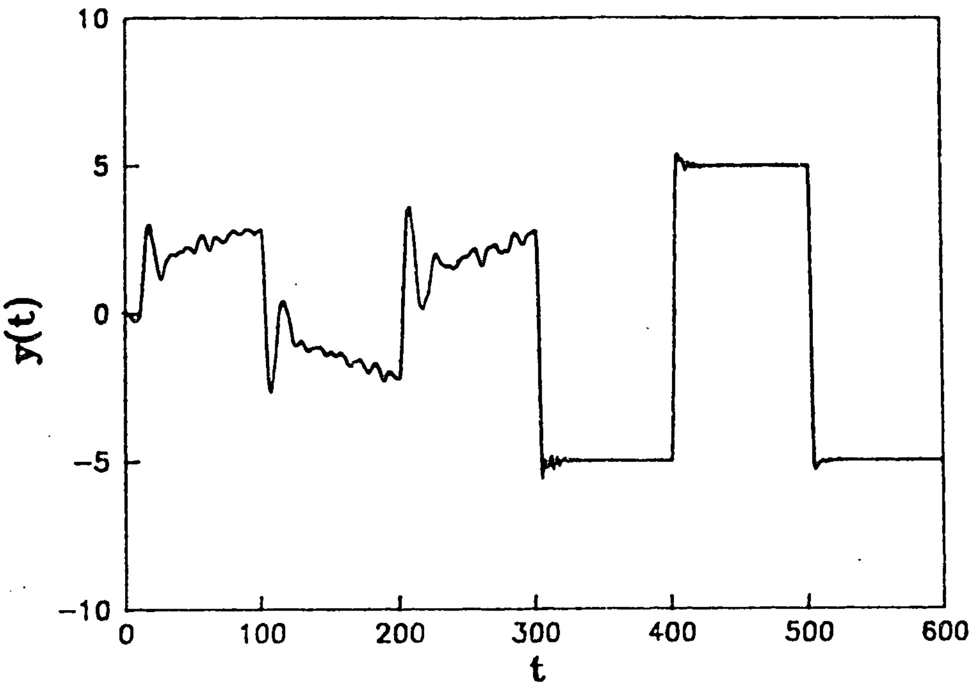

Figure 4.3a: Output response using multirate, constrained adaptive d+1 step ahead control (for t>300). System delay d=0, control interval=1, output measurement interval=10.

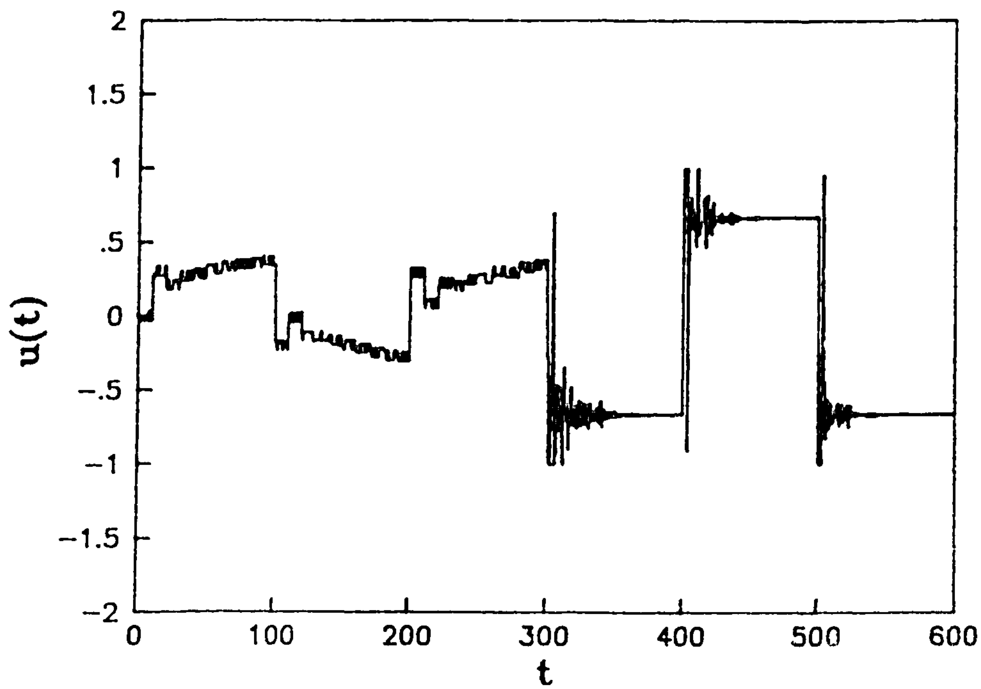

Figure 4.3b: Manipulated variable with constraints $-1 \le u(t) \le 1$.

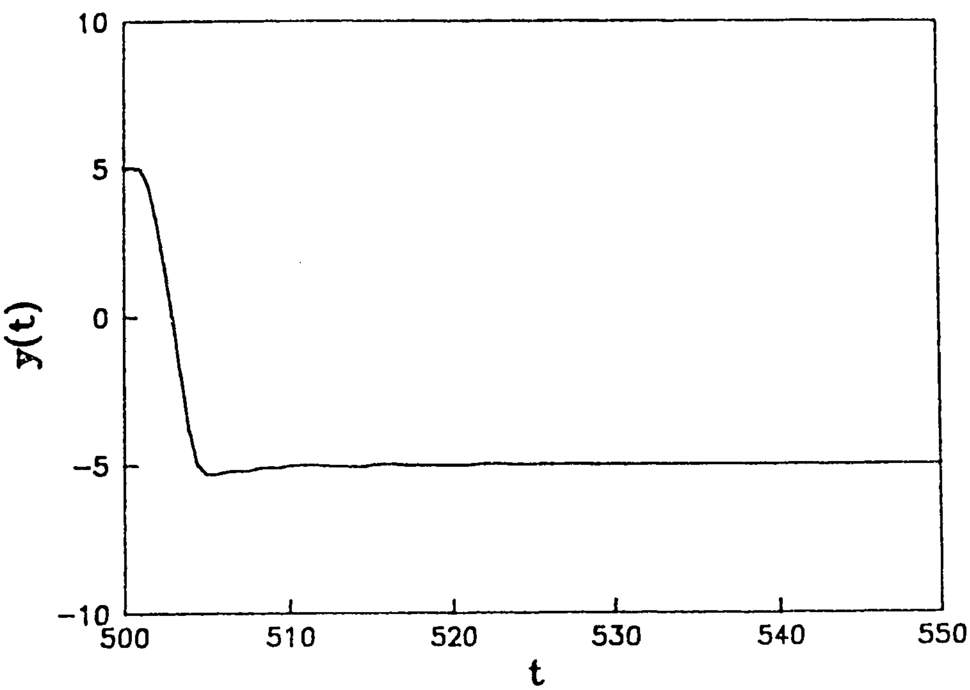

Figure 4.3c: Output response with an expanded time scale.

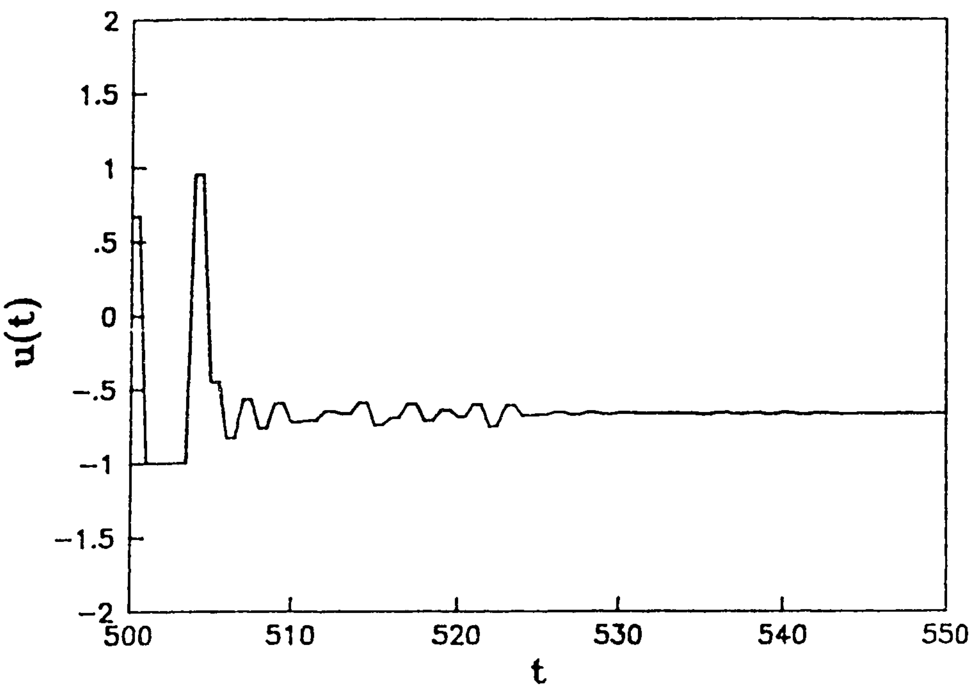

Figure 4.3d: Manipulated variable with an expanded time scale (to show control action every one time interval despite output sampling every 10 time intervals).

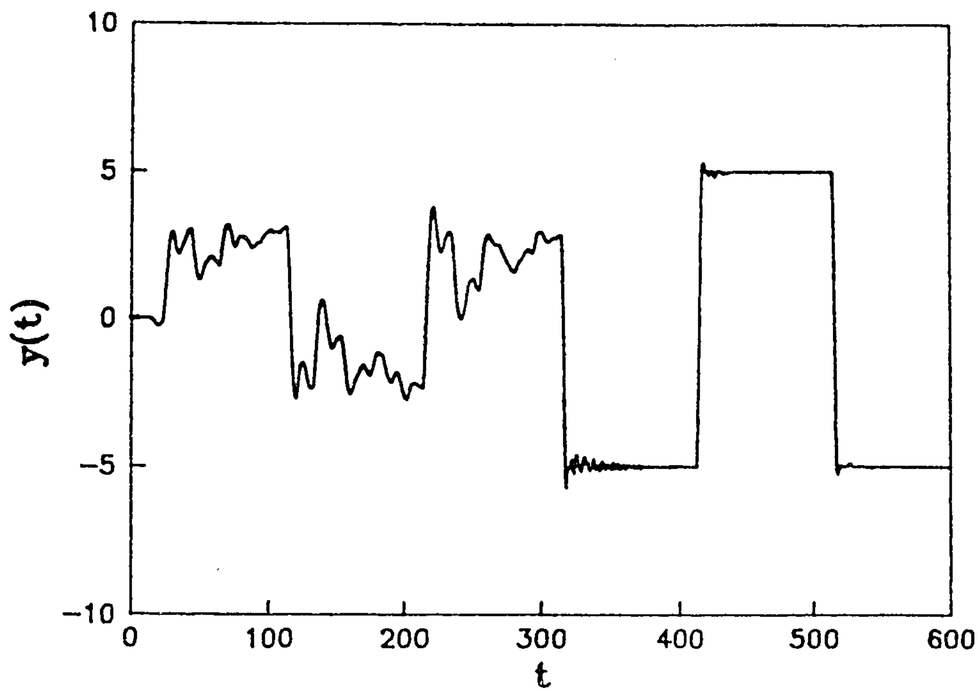

Figure 4.4a: Output response comparable to Figure 4.3 but with system delay d=12.

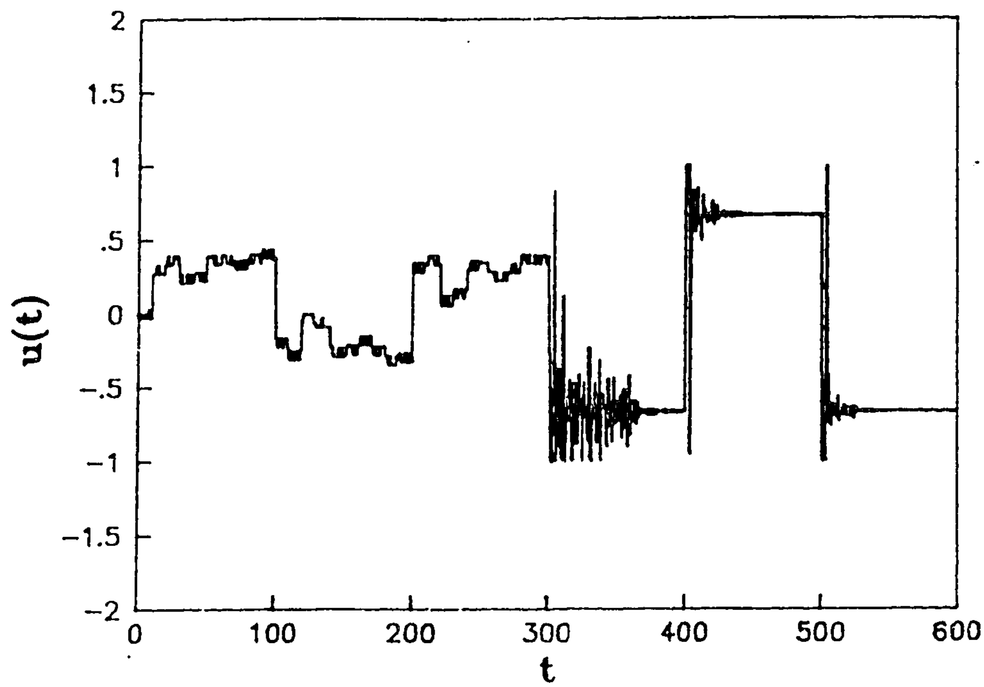

Figure 4.4b: Manipulated variable.

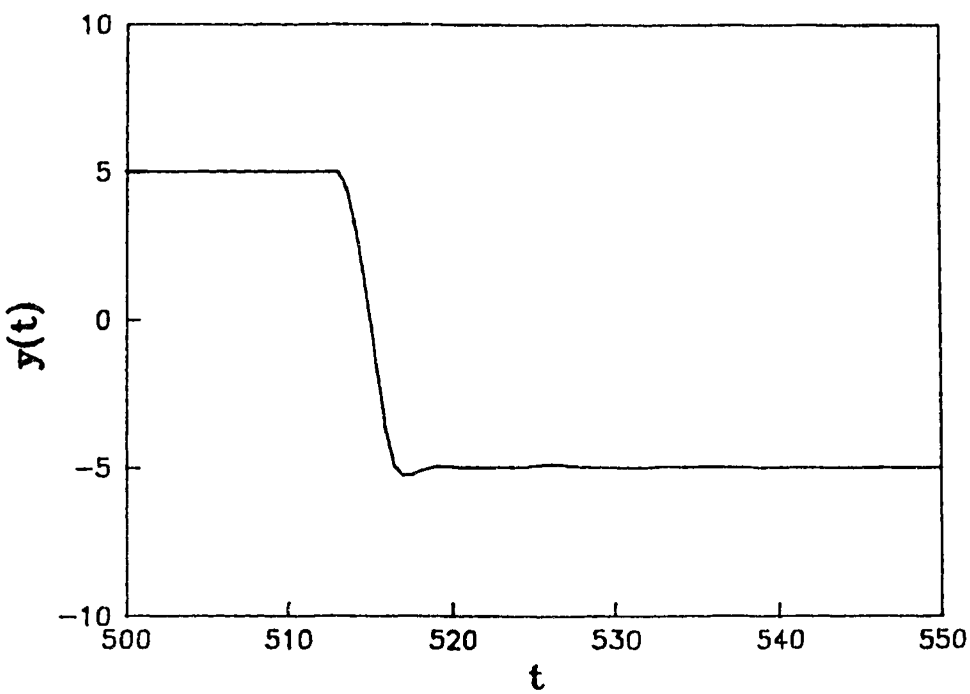

Figure 4.4c: Output response with an expanded time scale.

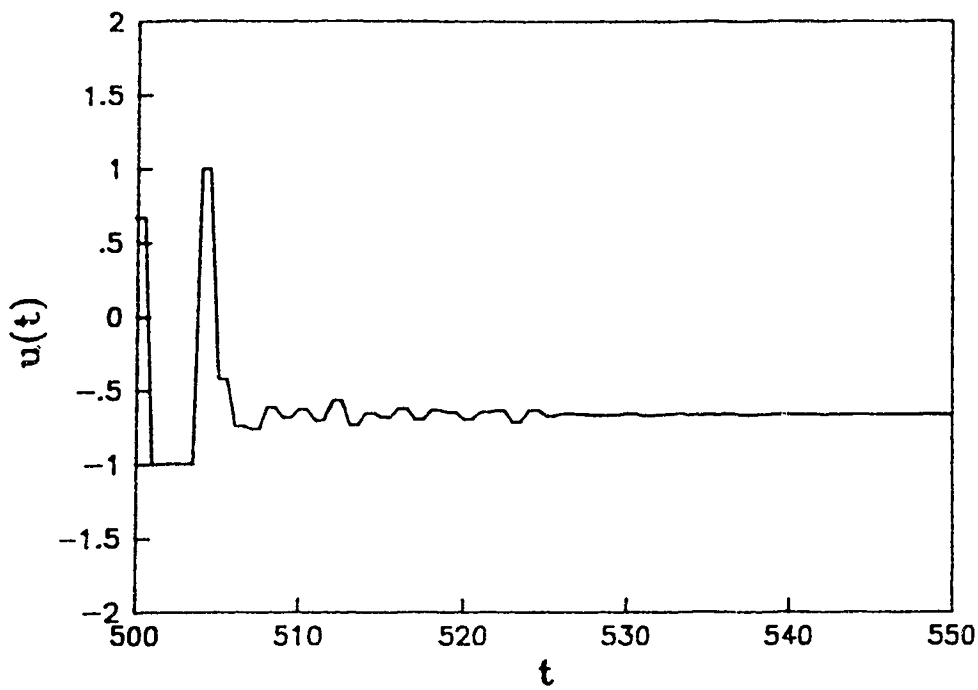

Figure 4.4d: Manipulated variable with an expanded time scale.

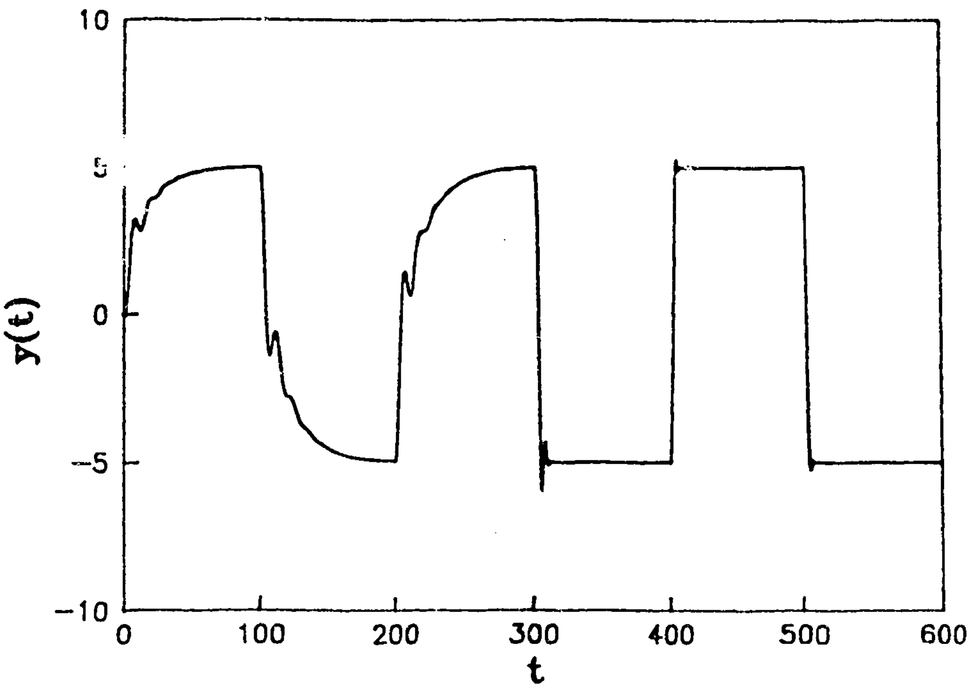

Figure 4.5a: Output response of the reference control system representing the best control performance achievable, i.e. output measurement interval=1 and plant parameters are known and used in the constrained d+1 step ahead control law. System delay d=0.

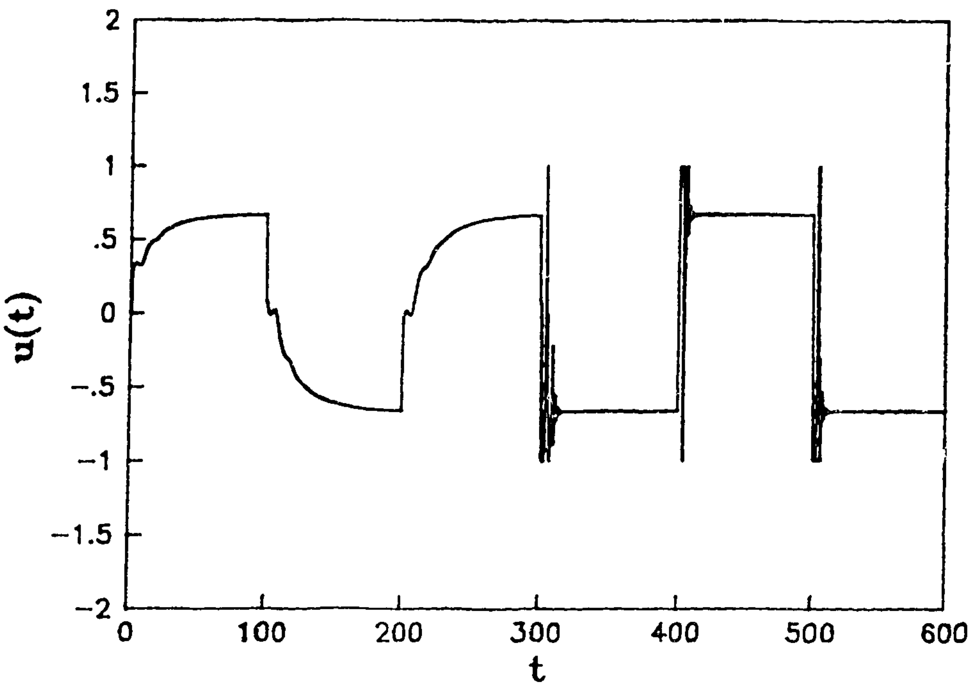

Figure 4.5b: Manipulated variable.

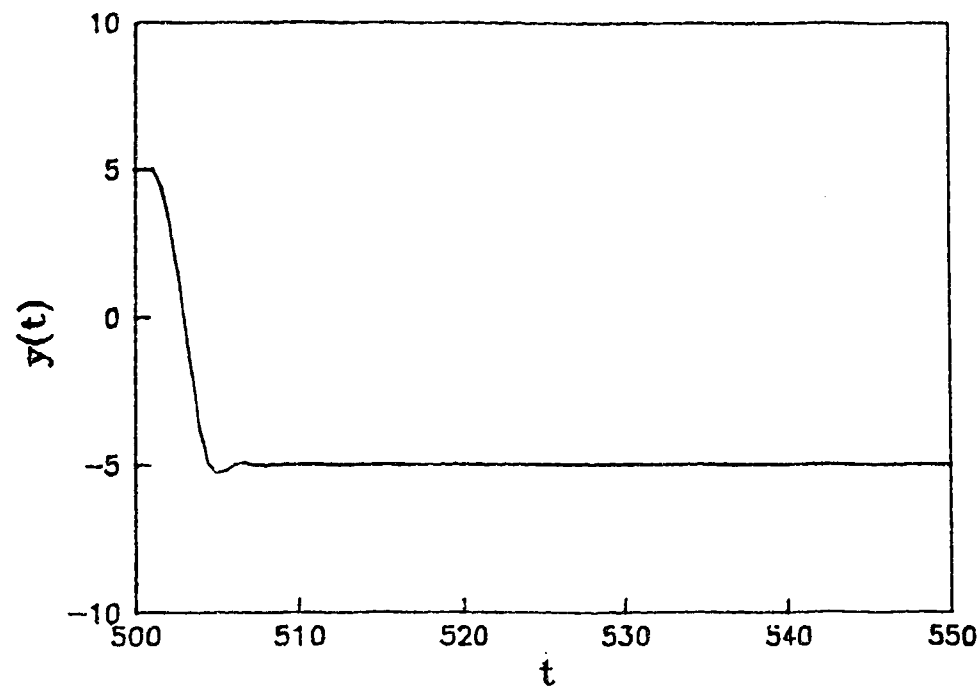

Figure 4.5c: Output response with an expanded time scale.

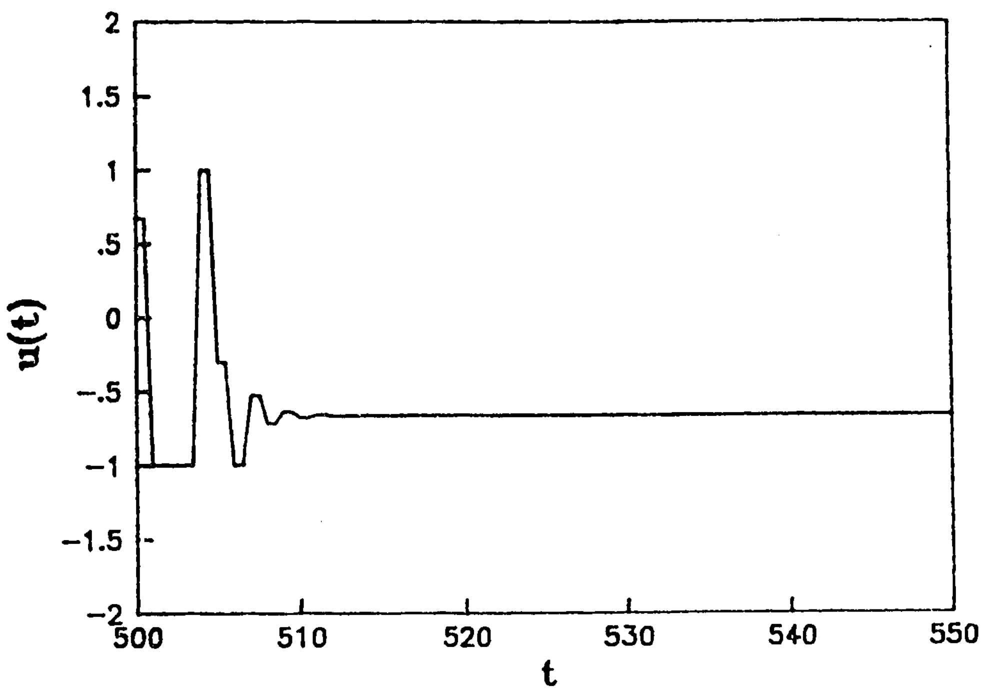

Figure 4.5d: Manipulated variable with an expanded time scale.

4.6 REFERENCES

Anderson B. D. O. and Johnson, Jr. C. R., "Exponential Convergence of Adaptive Identification and Control Algorithms," Automatica, vol. 18, pp. 1-13, Jan. 1982.

Goodwin G. C. and Sin K. S., Adaptive Filtering Prediction and Control, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

Heymann M., "Persistency of Excitation Results for Structured Nonminimal Models," IEEE Trans. Automat. Contr., vol. AC-33, pp. 112-116, 1988.

Landau, I. D., Adaptive Control: the Model Reference Approach, Marcel Dekker, Inc., New York, 1979.

Moore J. B., "A Universality Advantage of Stochastic Excitation Signals for Adaptive Control," Systems & Control Letters 9, pp. 55-58, 1987.

Scattolini R., "Self-tuning Control of Systems with Infrequent and Delayed Output Sampling," IEE Proc. D, Control Theory & Appl., 135, Pt. D, No. 4 pp. 213-221 July 1988.

Soderstrom T., Methods and Applications in Adaptive Control, Springer Verlag, 1980.

Tsiligiannis C. A. and Svoronos S. A., "Deterministic Convergence of a Clark-Gawthrop Self-tuning Controller," Automatica, vol. 22, No. 2, pp. 193-197, 1986.

Zhang Q. and Tomizuka M., "Multirate Sampling Adaptive Control and Its Application to Thermal Mixing Systems," Int. J. Control, vol. 47 No. 3, pp. 135-144, 1988.

5. MULTIRATE ADAPTIVE INFERENTIAL ESTIMATION AND CONTROL[*]

5.1 Introduction

This chapter combines some of the concepts used in multirate output estimation presented in Chapters 2 and 3 and inferential control (Parrish and Brosilow 1985). The resulting multirate adaptive inferential estimation and control system is shown in Figure 5.1. The primary process output, y(t), is sampled with a period JT while the input u(t) and secondary output v(t) are sampled at the desired control interval, T. These measurements are sent in parallel to a parameter identification algorithm and an output estimation algorithm. The output estimates, $y_e(t)$, are produced every sampling interval, T, and can be used, as shown in Figure 5.1, as the basis for control. If y_e is a good estimate of y then control should improve because it can be implemented with period T rather than JT.

The multirate adaptive inferential estimation and control system shown in Figure 5.1 was first proposed by Guilandoust et al. (1986, 1987a, 1987b, 1988). Their work includes two approaches: state space and input-output. It is original and practical but is incomplete in its theoretical foundation. For example, the state-space approach requires

----

[*]A version of this chapter has been submitted for publication: Lu W. and Fisher D. G., IEE Proc. D, Control Theory & Appl., 1989.

that the process be completely observable from the secondary output v(t). This requirement is severely restrictive since in most cases the dynamic modes of the primary output y(t) are not all included in the dynamic modes of v(t). The input-output approach does not require this observability assumption. It directly assumes that the process has two input-output models, one for y(t) versus u(t), another for v(t) versus u(t). However the formulation requires that the same white noise term be present in each of the two models to relate y(t) with v(t) to obtain the working equation of the algorithm. This approach does not adequately reflect the link between y and u, e.g. the link is only via the external white noise disturbance and if this disturbance vanished then there would be no theoretical basis for the working equation. Furthermore, it is difficult to interpret the physical meaning of the working equation, e.g. the relationship between the order of the polynomials in the working equation and the characteristics of the actual process is not clearly defined.

This chapter formulates the working equation based on the framework of linear models (Wolovich 1974). The working equation reflects fully and more fundamentally the inferential relation from v to y via the internal system structure. The proposed approach does not require the limiting assumptions made by Guilandoust et al. (1986, 1987a, 1987b, 1988); quantitatively defines the relationship between the working equation and the original process model plus the external disturbances; formally proves the output convergence properties; and provides a solid theoretical background for extending the result to multi-input multi-output cases.

When J, i.e. the output sampling interval, is increased then the number of parameters to be identified increases proportionally. For cases when the number of estimated parameters must be reduced, a simplified algorithm is proposed which works well in simulations but lacks a formal proof of convergence.

5.2 Models for Multirate Inferential Estimation

In the following discussion, models are derived first for multirate systems without external disturbances and then for systems with external deterministic and/or stochastic disturbances. It is assumed that the process shown in Figure 5.1 is completely observable from v(t) plus y(t), which is much less restrictive than assuming that it is completely observable from v(t) alone (Guilandoust et al. 1986, 1987a, 1987b, 1988). The process is of order n with an observability index nv from v(t). To simplify the notation it is assumed that the input sampling interval T=1 and t is also used to indicate discrete time.

5.2.1 Case 1: No Disturbances

By the observability assumption, the system can be represented as (Wolovich 1974):

$$x(t+1)=\begin{bmatrix} 0 & \dots & 0 & -a_1 & 0 & \dots & 0 & -\bar{a}_1 \\ & I_{nv-1} & & \vdots & \vdots & & \vdots & \vdots \\ & & & -a_{nv} & 0 & \dots & 0 & -\bar{a}_{nv} \\ 0 & \dots & 0 & -a_{nv+1} & 0 & \dots & 0 & -\bar{a}_{nv+1} \\ \vdots & & \vdots & \vdots & & I_{ny-1} & & \vdots \\ 0 & \dots & 0 & -a_n & & & & -\bar{a}_n \end{bmatrix} x(t)+\begin{bmatrix} b_1 \\ \vdots \\ b_{nv} \\ b_{nv+1} \\ \vdots \\ b_n \end{bmatrix} u(t) \tag{5.1}$$

$$v(t)=x_{nv}(t) \tag{5.2}$$

$$y(t)=hx_{nv}(t)+x_n(t) \tag{5.3}$$

where ny=n-nv. The corresponding input-output relation from u(t) and v(t) to y(t) can be expressed as follows

$$A(q^{-1})[y(t)-hv(t)]=B(q^{-1})u(t)+\bar{C}(q^{-1})v(t) \tag{5.4}$$

where $A(q^{-1})=1+\bar{a}_nq^{-1}+\bar{a}_{n-1}q^{-2}+\dots+\bar{a}_{nv+1}q^{-ny}=\prod_{i=1}^{ny}[1-(\lambda_iq)^{-1}]$. (5.5)

The λ_i's are the roots of $A(q^{-1})$,

$$B(q^{-1})=b_nq^{-1}+b_{n-1}q^{-2}+\dots+b_{nv+1}q^{-ny} \tag{5.6}$$

and $\bar{C}(q^{-1})=-a_nq^{-1}-a_{n-1}q^{-2}-\dots-a_{nv+1}q^{-ny}$. (5.7)

Let $C(q^{-1})=A(q^{-1})h+\bar{C}(q^{-1})$

$$=h+(\bar{a}_nh-a_n)q^{-1}+\dots+(\bar{a}_{nv+1}h-a_{nv+1})q^{-ny}. \tag{5.8}$$

Then

$$A(q^{-1})y(t)=B(q^{-1})u(t)+C(q^{-1})v(t). \tag{5.9}$$

Multiplying both sides of (5.9) by

$$\prod_{i=1}^{ny}[1+(\lambda_iq)^{-1}+\dots+(\lambda_iq)^{2-J}+(\lambda_iq)^{1-J}]$$

results in the equivalent form

$$A_J(q^{-J})y(t)=B_J(q^{-1})u(t)+C_J(q^{-1})v(t) \tag{5.10}$$

where $A_J(q^{-J})=1+a_{J1}q^{-J}+a_{J2}q^{-2J}+\ldots+a_{Jny}q^{-nyJ}$, (5.11)

$B_J(q^{-1})=b_{J1}q^{-1}+b_{J2}q^{-2}+\ldots+b_{Jm}q^{-m}$, (5.12)

$C_J(q^{-1})=C_{J0}+C_{J1}q^{-1}+C_{J2}q^{-2}+\ldots+C_{Jm}q^{-m}$, (5.13)

and $m=J\times ny$.

<u>Remarks:</u>

(1) The working equation (5.10) can also be derived directly from the original continuous model of the process with a discretized input (cf. Chapter 2).

(2) The stability property of $A_J(q^{-J})$ follows that of $A(q^{-1})$. If the parameters of $A(q^{-1})$ are real, so are those of $A_J(q^{-J})$.

5.2.2 Case 2: Deterministic and Stochastic Disturbances

If the dynamic modes of the deterministic disturbances do not result in pole-zero cancellation with the dynamic modes of the process, the composite system, i.e. the process system plus deterministic disturbances, can be represented by

$$x(t+1)=\begin{bmatrix} 0 & \cdots & 0 & -a_1 & 0 & \cdots & 0 & -\bar{a}_1 \\ & I_{nv-1} & & \vdots & \vdots & & \vdots & \vdots \\ & & & -a_{nv} & 0 & \cdots & 0 & -\bar{a}_{nv} \\ 0 & \cdots & 0 & -a_{nv+1} & 0 & \cdots & 0 & -\bar{a}_{nv+1} \\ \vdots & & \vdots & \vdots & & I_{ny-1} & & \vdots \\ 0 & \cdots & 0 & -a_n & & & & -\bar{a}_n \end{bmatrix} x(t)+\begin{bmatrix} b_1 \\ \vdots \\ b_{nv} \\ b_{nv+1} \\ \vdots \\ b_n \end{bmatrix} u(t)+$$

$$[r_1 \ \ldots \ r_{nv} \ r_{nv+1} \ \ldots \ r_n]^T\bar{w}(t) \quad (5.14)$$

$$v(t)=x_{nv}(t)+\eta_v(t), \quad (5.15)$$

$$y(t)=hx_{nv}(t)+x_n(t)+\eta_y(t), \quad (5.16)$$

where the augmented state variable x includes the dynamics of deterministic disturbances. The order of the composite system (5.14) is greater than or equal to the order of the process but is still represented by n. The stochastic disturbances, $\bar{w}(t)$, $\eta_v(t)$ and $\eta_y(t)$ are assumed to be white, Gaussian sequences with finite variances.

As in the zero disturbance case, the input-output relationship betwwen u(t), $\bar{w}(t)$, v(t) and y(t) can be obtained as follows:

$$A(q^{-1})[y(t)-hv(t)+h\eta_v(t)-\eta_y(t)]$$
$$=B(q^{-1})u(t)+\bar{C}(q^{-1})v(t)+R(q^{-1})\bar{w}(t) \quad (5.17)$$

where A, B and $\bar{C}$ have the form of (5.5-5.7) and

$$R(q^{-1})=r_nq^{-1}+r_{n-1}q^{-2}+\ldots+r_{nv+1}q^{-ny}. \quad (5.18)$$

Similarly, like (5.9)

$$A(q^{-1})y(t)=B(q^{-1})u(t)+C(q^{-1})v(t)+D(q^{-1})z(t) \quad (5.19)$$

where z(t) is white and has a finite variance. In (5.19) $R(q^{-1})\bar{w}(t)-A(q^{-1})h\eta_v(t)+A(q^{-1})\eta_y(t)$ has been replaced by $D(q^{-1})z(t)$ using the representation and spectral factorization principle (Astrom 1970). The polynomial $D(q^{-1})$ is defined by

$$D(q^{-1})=d_0+d_1q^{-1}+\ldots+d_{ny}q^{-ny}. \quad (5.20)$$

Like (5.10), an equivalent form of (5.19) is

$$A_J(q^{-J})y(t)=B_J(q^{-1})u(t)+C_J(q^{-1})v(t)+D_J(q^{-1})z(t) \quad (5.21)$$

where

$$D_J(q^{-1})=d_{J0}+d_{J1}q^{-1}+d_{J2}q^{-2}+\ldots+d_{Jm}q^{-m}. \quad (5.22)$$

Remarks:

(1) If the deterministic disturbances result in pole-zero cancellation with the process, the observability property of

the composite system is lost. This situation, called input zero blocking, is not discussed in here.

(2) The derivation of the working equation (5.10) or (5.21) is easily extended to MIMO cases. For example if there are several secondary measured variables rather than a single v(t) the same approach can be used to derive an appropriate working equation. On the other hand if v(t) is not available, i.e. if nv=0, the working equation naturally reduces to the SISO case treated in Chapters 2 and 3. Note that it is almost impossible to make either the extension to the MIMO case or the reduction to the SISO case when using the approach of Guilandust et al. (1986, 1987a, 1987b, 1988).

5.3 Estimation Algorithm

5.3.1 Definitions

The notation used to describe the process in (5.10) and (5.21) can be simplified by dropping the J subscript. The inferential model of the process then becomes

$$A(q^{-J})y(t)=B(q^{-1})u(t)+C(q^{-1})v(t)+D(q^{-1})z(t) \qquad (5.23)$$

where

$$A(q^{-J})=1+a_1q^{-J}+a_2q^{-2J}+...+a_nq^{-nJ}, \qquad (5.24)$$

$$B(q^{-1})=b_1q^{-1}+b_2q^{-2}+...+b_mq^{-m}, \qquad (5.25)$$

$$C(q^{-1})=c_0+c_1q^{-1}+c_2q^{-2}+ ...+c_mq^{-m}, \qquad (5.26)$$

$$D(q^{-1})=d_0+d_1q^{-1}+d_2q^{-2} +...+d_mq^{-m}, \qquad (5.27)$$

$m=Jxn$.

Note that here n plays the role of ny in the previous section. It is further assumed that the original state space representation (5.14-5.16) has all its eigenvalues inside the stable region.

Let

$$\phi(t-1)^T=[-y(t-J),-y(t-2J),\ldots -y(t-nJ),u(t-1),u(t-2),\ldots,u(t-m),v(t),v(t-1),\ldots,v(t-m)] \qquad (5.28)$$

and

$$\theta_0{}^T=[a_1,\ldots,a_n,b_1,\ldots,b_m,c_0,c_1,\ldots,c_m]. \qquad (5.29)$$

Then

$$y(t)=\phi(t-1)^T\theta_0+D(q^{-1})z(t). \qquad (5.30)$$

Next, Define the following:

<u>A posteriori model output</u>

$$\bar{y}(t)=\bar{\phi}(t-1)^T\hat{\theta}(t) \qquad (5.31)$$

where

$$\bar{\phi}(t-1)^T=[-\bar{y}(t-J),-\bar{y}(t-2J),\ldots -\bar{y}(t-nJ),u(t-1),u(t-2),\ldots,u(t-m),v(t),v(t-1),\ldots,v(t-m)] \qquad (5.32)$$

with the initial values

$\bar{\phi}(-1)=\phi(-1)$ (available from the measurement data),

and $\bar{\phi}(-1+i)$=arbitrary and

$$\hat{\theta}(t)^T=[\hat{a}_1(t),\ldots,\hat{a}_n(t),\ \hat{b}_1(t),\ldots,\hat{b}_m(t),\hat{c}_0(t),\ldots,\hat{c}_m(t)] \qquad (5.33)$$

i.e. $\hat{\theta}(t)$ is the estimate of θ_0 at time t.

<u>A posteriori model output error</u>

$$\eta(t)=y(t)-\bar{y}(t) \qquad (5.34)$$

<u>A priori model output</u>

$$\hat{y}(t)=\bar{\phi}(t-1)\hat{\theta}(t-1) \qquad (5.35)$$

<u>A priori model output error</u>

$$e(t)=y(t)-\hat{y}(t) \tag{5.36}$$

<u>Generalized a posterior output error</u>

$$\bar{\eta}(t)=L(q^{-J})\eta(t) \tag{5.37}$$

where

$$L(q^{-J})=1+l_1q^{-J}+...+l_{1'}q^{-1'J} \tag{5.38}$$

is a fixed, moving-average filter.

<u>Generalized a priori output error</u>

$$\bar{v}(t)=e(t)+[L(q^{-J})-1]\eta(t) \tag{5.39}$$

5.3.2 Estimation Algorithm

The parameter estimation algorithm is given by

$$\hat{\theta}(Jt)=\hat{\theta}(Jt-J)+$$

$$P(Jt-2)\bar{\phi}(Jt-1)\bar{v}(Jt)/[1+\bar{\phi}(Jt-1)^TP(Jt-2)\bar{\phi}(Jt-1)], \tag{5.40}$$

$$\hat{\theta}(Jt+i)=\hat{\theta}(Jt) \qquad (i=1,2,...,J-1), \tag{5.41}$$

$$P[J(t+1)-2]=P(Jt-2)-P(Jt-2)\bar{\phi}(Jt-1)\bar{\phi}(Jt-1)^TP(Jt-2)$$

$$/[1+\bar{\phi}(Jt-1)^TP(Jt-2)\bar{\phi}(Jt-1)], \tag{5.32}$$

$$\hat{\theta}(0)=\text{arbitrary}, \tag{5.43}$$

$$P(-2)>0. \tag{5.44}$$

The regressor $\bar{\phi}(t)$ has been defined in (5.32). At each unity time step, the estimated outputs $\bar{y}(t)$ and $\hat{y}(t+1)$ can be calculated by using (5.31) and (5.35) although the output is measured only every J sampling intervals.

5.3.3 Convergence at Output Sampling Instants (for $z(t)\equiv 0$)

The first step is to define the convergence properties of the output estimates at the output (slow) sampling interval, JT.

<u>Theorem 5.1:</u>

Consider the algorithm (5.40-5.44) applied to inferential model (5.23) with $z(t)\equiv 0$; then, provided that the system $H(q^{-J})=[L(q^{-J})/A(q^{-J})-1/2]$ is very strictly passive:

(i)
$$\lim_{N\to\infty}\sum_{t=1}^{N} \eta(Jt)^2<\infty, \tag{5.45}$$

which implies

$$\lim_{N\to\infty}\sum_{t=1}^{N} \bar{\eta}(Jt)^2<\infty \tag{5.46}$$

and
$$\lim_{t\to\infty}|y(Jt)-\bar{y}(Jt)|=0; \tag{5.47}$$

(ii)
$$\lim_{N\to\infty}\sum_{t=1}^{N} \bar{\phi}(Jt-1)^T P(Jt-2)\bar{\phi}(Jt-1)\bar{\eta}(Jt)^2<\infty, \tag{5.48}$$

which implies

$$\lim_{N\to\infty}\sum_{t=1}^{N} \|\hat{\theta}(Jt)-\hat{\theta}[J(t-s)]\|^2<\infty \tag{5.49}$$

where s is any finite integer;

(iii) If $\{u(t)\}$ is bounded, then

$$\lim_{t\to\infty} \bar{v}(Jt)=0 \tag{5.50}$$

and
$$\lim_{t\to\infty} |y(Jt)-\hat{y}(Jt)|=0. \tag{5.51}$$

<u>Proof:</u>

Define
$$b(t)=-\bar{\phi}(t-1)^T\tilde{\theta}(t) \tag{5.52}$$

where
$$\tilde{\theta}(t)=\hat{\theta}(t)-\theta_0. \tag{5.53}$$

Combining (5.30) and (5.31) (and noting $z(t)\equiv 0$) gives

$$A(q^{-J})\eta(t)=b(t), \tag{5.54}$$

or particularly
$$A(q^{-J})\eta(Jt)=b(Jt). \tag{5.55}$$

(5.37) and (5.55) give

$$A(q^{-J})\bar{\eta}(Jt)=D(q^{-J})b(Jt). \tag{5.56}$$

Multiplying (5.40) by $\bar{\phi}(Jt-1)^T$ and then subtracting from y(Jt) gives

$$\eta(Jt)=e(Jt)$$
$$-\bar{\phi}(Jt-1)^T P(Jt-2)\bar{\phi}(Jt-1)\bar{v}(Jt)$$
$$/[1+\bar{\phi}(Jt-1)^T P(Jt-2)\bar{\phi}(Jt-1)]. \quad (5.57)$$

Combining (5.57) with (5.37) and (5.39), yields

$$\bar{\eta}(Jt)=\bar{v}(Jt)/[1+\bar{\phi}(Jt-1)^T P(Jt-2)\bar{\phi}(Jt-1)] \quad (5.58)$$

Substituting (5.58) into (5.40) gives

$$\hat{\theta}(Jt)=\hat{\theta}(Jt-J)+P(Jt-2)\bar{\phi}(Jt-1)\bar{\eta}(Jt). \quad (5.59)$$

Substracting θ_0 from both sides yields

$$\tilde{\theta}(Jt)-P(Jt-2)\bar{\phi}(Jt-1)\bar{\eta}(Jt)=\tilde{\theta}(Jt-J). \quad (5.60)$$

Let

$$V(Jt)=\tilde{\theta}(Jt)^T P[J(t+1)-2]^{-1}\tilde{\theta}(Jt) \quad (5.61)$$

and then from (5.60) and (5.61)

$$[\tilde{\theta}(Jt)-P(Jt-2)\bar{\phi}(Jt-1)\bar{\eta}(Jt)]^T P(Jt-2)^{-1}$$
$$[\tilde{\theta}(Jt)-P(Jt-2)\bar{\phi}(Jt-1)\bar{\eta}(Jt)]^T=V(Jt-J) \quad (5.62)$$

or $$\tilde{\theta}(Jt)^T P(Jt-2)^{-1}\tilde{\theta}(Jt)-2\tilde{\theta}(Jt)^T\bar{\phi}(Jt-1)\bar{\eta}(Jt)$$
$$+\bar{\phi}(Jt-1)^T P(Jt-2)\bar{\phi}(Jt-1)\bar{\eta}^2(Jt) =V(Jt-J) \quad (5.63)$$

Using the Inversion Lemma (Lemma 3.3.4, Goodwin and Sin 1984)

$$\tilde{\theta}(Jt)^T P(Jt-2)^{-1}\tilde{\theta}(Jt)=$$
$$\tilde{\theta}(Jt)^T\{P[J(t+1)-2]^{-1}-\bar{\phi}(Jt-1)\bar{\phi}(Jt-1)^T\}\tilde{\theta}(Jt)^T$$
$$=V(Jt)-\tilde{\theta}(Jt)^T\bar{\phi}(Jt-1)\bar{\phi}(Jt-1)^T\tilde{\theta}(Jt). \quad (5.64)$$

Combining (5.52), (5.62) and (5.63)

$$V(Jt)=V(Jt-J)-2[\bar{\eta}(Jt)-b(Jt)/2]b(Jt)$$
$$-\bar{\phi}(Jt-1)^T P(Jt-2)\bar{\phi}(Jt-1)\bar{\eta}^2(Jt). \quad (5.65)$$

The remainder of the proof is the same as that of Theorem

3.5.1 in (Goodwin and Sin 1984), except that the very strictly passive condition is used for the relation between $[\bar{\eta}(Jt)-b(Jt)/2]$ and $b(Jt)$, where the input sequence is $b(0), b(J), b(2J), \ldots$ and the output sequence is $[\bar{\eta}(0)-b(0)/2]$, $[\bar{\eta}(J)-b(J)/2]$, $[\bar{\eta}(2J)-b(2J)/2], \ldots$. ■

5.3.4 Convergence at the Input Sampling Instants (for $z(t)\equiv 0$)

Using the results from Theorem 5.1 it is now possible to define the convergence properties of the output estimate at each input sampling interval, T, i.e. at the output intersampling points.

Theorem 5.2:

Under the same conditions as Theorem 5.1:

(i) $$\|\hat{\theta}(Jt)-\theta_0\|^2 \le \kappa_1 \|\hat{\theta}(0)-\theta_0\|^2 \qquad \forall t>0 \tag{5.66}$$

where $\kappa_1=\lambda_{max}[P(-2)^{-1}+\phi(-1)\phi(-1)^T]/\lambda_{min}[P(-2)^{-1}]$, and $\lambda_{max}(\cdot)$ and $\lambda_{min}(\cdot)$ represent the maximum and minimum eigenvalues respectively.

(ii) There exists a positive number ϵ such that if $\{u(t)\}$ is bounded then $\|\hat{\theta}(0)-\theta_0\|^2<\epsilon$ implies

(a) $|y(t)-\hat{y}(t)| \le$

$$\delta[\limsup_{t\to\infty}\|\hat{\theta}(Jt)-\theta_0\|][\limsup_{t\to\infty}|u(t)-u(t-1)|]+\Delta(t) \qquad \forall t>0 \tag{5.67}$$

where $\Delta(t)$ is a sequence satisfying $\lim_{t\to\infty}\Delta(t)=0$, δ and the two limit superiors are finite numbers.

(b) $$\lim_{t\to\infty}|y(t)-\hat{y}(t)|=0 \tag{5.68}$$

provided that $\lim_{t\to\infty}\hat{\theta}(Jt)=\theta_0$ or $\lim_{t\to\infty}[u(t)-u(t-1)]=0$.

Proof:

(i) Using (5.65),

$$V(Jt)=V(0)-2\sum_{i=1}^{t} b(Ji)[\bar{\eta}(Ji)-b(Ji)/2] -\sum_{i=1}^{t} \bar{\phi}(Ji-1)^T P(Ji-2)\bar{\phi}(Ji-1)\bar{\eta}^2(Ji). \quad (5.69)$$

From (5.56)

$$[\bar{\eta}(Jt)-b(Jt)/2]=[D(q^{-J})/A(q^{-J})-1/2]b(Jt) =H(q^{-J})b(Jt). \quad (5.70)$$

By the very strictly passive assumption of $H(q^{-J})$

$$\sum_{i=1}^{t} b(Ji)[\bar{\eta}(Ji)-b(Ji)/2] \geq -b(0)[\bar{\eta}(0)-b(0)/2]. \quad (5.71)$$

Considering the initial conditions of $\bar{\phi}$,

$$\bar{\eta}(0)=D(q^{-J})\eta(t)|_{t=0}=\eta(0)=y(0)-\bar{y}(0) =\phi(-1)^T\theta_0-\bar{\phi}(-1)^T\hat{\theta}(0)=\phi(-1)^T[\theta_0-\hat{\theta}(0)] =-\tilde{\theta}(0)^T\phi(-1). \quad (5.72)$$

But

$$b(0)=-\tilde{\theta}^T(0)\bar{\phi}(-1)=-\tilde{\theta}^T(0)\phi(-1) \quad (5.73)$$

and therefore

$$\sum_{i=1}^{t} b(Ji)[\bar{\eta}(Ji)-b(Ji)/2] \geq -b^2(0)/2=-\tilde{\theta}^T(0)\phi(-1)\phi(-1)^T\tilde{\theta}(0)/2. \quad (5.74)$$

Substituting (5.74) into (5.69) and using the definition of V give

$$\tilde{\theta}(Jt)^T P[J(t+1)-2]^{-1}\tilde{\theta}(Jt) \leq \tilde{\theta}^T(0)[P(-2)+\phi(-1)\phi(-1)^T]^{-1}\tilde{\theta}(0) \quad (5.75)$$

where the positive definite property of P is used.

By the Inversion Lemma (Lemma 3.3.4, Goodwin and Sin 1984) it is easy to verify that

$$\lambda_{min}[P[J(t+1)-2]^{-1}] \geq \lambda_{min}[P(-2)^{-1}] \qquad \forall t\geq 0. \quad (5.76)$$

Then (5.75) and (5.76) immediately yield (5.66).

(ii) From (5.54)

$$A(q^{-J})\eta(t)=b(t) \tag{5.77}$$

or equivalently, using (41),

$$A(q^{-J})\eta(Jt+i)=-\bar{\phi}(Jt+i-1)\tilde{\theta}(Jt) \quad i=0,1,2,...,J-1. \tag{5.78}$$

From (5.78)

$$A(q^{-J})[\eta(Jt+i)-\eta(Jt)]=-[\bar{\phi}(Jt+i-1)-\bar{\phi}(Jt-1)]\tilde{\theta}(Jt)$$

$$i=1,2,..,J-1 \tag{5.79}$$

But

$$\bar{\phi}(Jt+i-1)=$$

$$\{[F_1(Jt-J,q^{-1})q^{-J},F_1(Jt-2J,q^{-1})q^{-2J},...,F_1(Jt-nJ,q^{-1})q^{-nJ}]u(Jt+i)$$
$$+[F_2(Jt-J,q^{-1})q^{-J},F_2(Jt-2J,q^{-1})q^{-2J},...,F_2(Jt-nJ,q^{-1})q^{-nJ}]v(Jt+i),$$
$$[q^{-1},q^{-2},...,q^{-m}]u(Jt+i),$$
$$[1,q^{-1},q^{-2},...,q^{-m}]v(Jt+i)\} \quad i=1,2,...,J-1 \tag{5.80}$$

where

$$F_1(t,q^{-1})=\hat{B}(t,q^{-1})/\hat{A}(t,q^{-J}),\ F_2(t,q^{-1})=\hat{C}(t,q^{-1})/\hat{A}(t,q^{-J}),$$

$$\hat{A}(t,q^{-J})=1+\hat{a}_1(t)q^{-J}+\hat{a}_2(t)q^{-2J}+...+\hat{a}_n(t)q^{-nJ},$$

$$\hat{B}(t,q^{-1})=\hat{b}_1(t)q^{-1}+\hat{b}_2(t)q^{-2}+...+\hat{b}_m(t)q^{-m}$$

and $\hat{C}(t,q^{-J})=\hat{c}_0(t)+\hat{c}_1(t)q^{-J}+\hat{c}_2(t)q^{-2J}+...+\hat{c}_n(t)q^{-nJ}$.

By the very strictly passive assumption, $A(q^{-J})$ is asymptotically stable. Thus there exists $\epsilon_1>0$ such that: if $\|\hat{\theta}(0)-\theta_0\|^2<\epsilon_1$, $\hat{A}(0,q^{-J})$ is also asymptotically stable. From (i) it is concluded that $\hat{A}(Jt,q^{-J})$ $\forall t\geq 0$ has uniformly all its eigenvalues strictly inside the unit circle for $\epsilon=\epsilon_1/\kappa_1$. Also from (5.49), (5.80) is slowly time varying and it is not difficult to show that ϵ can be such that the systems $F_1(t,q^{-1})$ and $F_2(t,q^{-1})$ are slowly time varing and global uniform exponentially stable uniformly in their parameters.

Therefore bounded $\{u(t)\}$ and $\{v(t)\}$ imply bounded $\{\bar{\phi}(t)\}$ and that there exist $0<M_1,M_2,M<\infty$ such that

$$\begin{aligned}\|\bar{\phi}(Jt+i-1)-\bar{\phi}(Jt-1)\| \le\ & M_1|u(Jt+i)-u(Jt)|+M_2|v(Jt+i)-v(Jt)| \\ & \le M|u(Jt+i)-u(Jt)| \\ & i=1,2,\ldots,J-1 \end{aligned} \tag{5.81}$$

since $\bar{\phi}(kJ+i-1)-\bar{\phi}(kJ-1)$ versus $u(kJ+i)-u(kJ)$ and $v(kJ+i)-v(kJ)$ $i=1,\ldots,J-1$ also satisfies (5.80) and the relationship between $v(t)$ and $u(t)$ is linear time invariant and asymptotically stable by the assumption on the eigenvalues of the original process.

Considering that $\lim_{t\to\infty} \eta(Jt)=0$ (from (5.47)) and $A(q^{-J})$ is asymptotically stable, (5.79) and (5.81) yield

$$|\eta(Jt+i)|\le\delta_1[\limsup_{t\to\infty}\|\tilde{\theta}(Jt)\|][\limsup_{t\to\infty}|u(Jt+i)-u(Jt)|]+\Delta(Jt+i) \qquad i=1,2,\ldots,J-1 \tag{5.82}$$

where $\Delta(Jt+i)$ $i=1,2,\ldots,J-1$ are some sequences satisfying $\lim_{t\to\infty} \Delta(Jt+i)=0$, δ_1 and the two limit supperios are finite positive numbers. Note that for $i=1,2,\ldots,J-1$

$\limsup_{t\to\infty}|u(Jt+i)-u(Jt)| \le (J-1)\limsup_{t\to\infty}|u(t)-u(t-1)|$.

Part (a) of (ii) is obtained by letting $|y(Jt)-\hat{y}(Jt)|=\Delta(Jt)$, $\delta=(J-1)\delta_1$ and noting that

$$\begin{aligned}\eta(Jt+i)&=y(Jt+i)-\bar{y}(Jt+i)=y(Jt+i)-\bar{\phi}(Jt+i-1)\tilde{\theta}(Jt+i) \\ &=y(Jt+i)-\bar{\phi}(Jt+i-1)^T\tilde{\theta}(Jt) \\ &=y(Jt+i)-\hat{y}(Jt+i) \qquad i=1,2,\ldots,J-1\end{aligned} \tag{5.83}$$

and using (5.51) and (5.82). Part (b) of (ii) is obvious. ■

5.3.5 Remarks on Convergence Theorems

(1) Theorems 5.1 and 5.2 are extensions of the results

presented in Chapters 2 and 3. Here the secondary output $v(t)$ is included in the algorithm and the proofs are presented in greater detail.

(2) The convergence results of Theorems 5.1 and 5.2 are applicable to cases with $z(t)\equiv 0$ and can be applied to the convergence analysis of any adaptive servo control using the multirate, inferential estimation algorithm.

(3) The estimation algorithm (5.40-5.44) will handle processes operating in noisy, stochastic disturbance environments. In general, if external stochastic disturbances are present then there is model mismatch since the parameters of the external stochastic disturbance model, $D(q^{-1})$, are not identified. This makes the convergence analysis quite difficult.

(4) Note that only output convergence is proven and no conclusion is made about parameter convergence to the true parameter vector. It is not difficult to observe from the derivation in the previous section that the parmeterization of the inferential equation (eg. (5.23) with $z(t)\equiv 0$) is, in general, not unique, i.e. θ_0 can be anything belonging to an equivalence class set in the parameter vector space. It would be desirable, in cases where parameter convergence is important, to use some improved algorithm with structurally constrained, inferential working equations so that only a unique convergence point in the parameter vector space exists for identification.

5.3.6 Simplified Algorithm

The algorithm (5.40-5.44) has one practical disadvantage: the number of the parameters to be estimated increases linearly with J. Since an exact model match can not be achieved by the algorithm (5.40-5.44) when $D(q^{-1})z(t)\neq 0$, it may be desirable to simplify the algorithm for such cases. Decreasing the number of parameters in the algorithm would, in general, increase the model mismatch and result in poorer performance. However, in some applications reducing the number of parameters to be estimated improves the numerical conditioning of the estimation algorithm and reduces the variance of the output estimate, thus resulting in better overall performance. Based on this observation (cf. parsimony principle, Ljung 1988) it is proposed that the algorithm defined by (5.40-5.44) be simplified by reducing the number of $\hat{b}$ and $\hat{c}$ parameters. One extreme case, considered here as an demonstration example, is to set $\hat{b}_{i+1}(t)=\hat{c}_i(t)\equiv 0$ for $i\neq 0,J,2J,...,(n-1)J$. The number of parameters to be estimated in the proposed simplified algorithm is 3n and is therefore independent of J. The algorithm (5.40-5.44) could be simplified or modified in other ways, which should depend on the application. The simplified algorithm considered here is similar in form to the original algorithm and can achieve an exact model match (if $\hat{c}_{nJ}$ is not set to zero and included in the parameter estimate vector) at the output sampling intervals if $D(q^{-1})z(t)=0$ and the input u is kept constant within each primary output sampling interval, i.e. over J intervals.

5.3.7 Predictive Estimation

A one step ahead prediction of y(t) can be calculated from the a priori model (5.13):

$$\hat{y}(t+1)=\bar{\phi}(t)^T\hat{\theta}(t). \tag{5.84}$$

In general, to predict y(t) k steps ahead:

$$y_e(t+k)=\phi^T{}_e(t-1+k)\hat{\theta}(t) \tag{5.85}$$

where $\phi^T{}_e(t-1+k)=$

$$[-y_e(t-J+k),-y_e(t-2J+k),\ldots,-y_e(t-nJ+k),\ldots] \tag{5.86}$$

and $y_e(\tau)=\bar{y}(\tau)$ if $\tau\le t$

$$y_e(\tau)=y_e(\tau) \quad \text{if } t<\tau. \tag{5.87}$$

5.4 Simulation Results

The simulated process is given by

$$y=[\frac{-5.9}{7.8s+1}+\frac{a}{3.6s+1}]e^{-ks}u+\frac{-6.34}{19.2s+1}e^{-ks}w \tag{5.88}$$

$$v=\frac{-16.877}{3.6s+1}u+\frac{-18.0}{9s+1}w \tag{5.89}$$

This model is a modified version of the linearized distillation column model described by Patke et al. (1982), where y is the composition of one component of the overhead, u is the reflux rate, v is a temperature measured at an appropriate stage in the column and w is a disturbance in the feed. The process output, y(t), was scaled to a comparable numerical magnitude with respect to that of u and v, and for convenience the model was considered to be dimensionless. The term $\frac{a}{3.6s+1}e^{-ks}u$ in (5.88) was added to the original model of Patke et al. (1982) to illustrate different modes of coupling between v(t) and y(t). (In the following simulation examples a=0.0 or a=0.5.)

5.4.1 Open-loop Output Estimation without Disturbances (J=10)

For the process defined by (88-89), the time delay k=0, the disturbance w=0 and the input u is a PRBS sequence passed through a zero order hold. The sampling interval for u and v is one time unit but the sampling interval for the output, y, is 10 units, i.e. J=10. The appropriate working equation, as defined in section 2 is

$$(1+a_1q^{-10})y(t)=\sum_{i=1}^{10} b_iq^{-i}u(t)+\sum_{i=0}^{10} c_iq^{-i}v(t). \qquad (5.90)$$

For the simplified algorithm, the working equation is

$$(1+a_1q^{-10})y(t)=b_1q^{-1}u(t)+c_0v(t) \qquad (5.91)$$

Remarks:

(1) Since a zero-order hold is used, the algorithm (5.40-5.44) with the working equation (5.90) (full algorithm) allows an exact model match to the simulated plant model.

(2) If the simplified algorithm is used then there is model mismatch because of the insufficient number of the estimated parameters.

(3) When using the full algorithm, it is not necessary to use v(t) even if $\alpha \neq 0$. However, if v(t) is not used then the assumptions about the state representation of the process change accordingly, e.g. the observability index with respect to v(t) becomes zero. Therefore the structure of the working equation has to be reformulated (cf. Chapters 2 and 3), i.e. the working equation cannot be obtained by simply dropping the v(t) terms in (5.90). The output estimate will

still converge to the real output and the number of parameters to be estimated does not change (cf. Chapters 2 and 3). The advantage of using v(t) is that, each common mode shared by y and v will, in general, reduce the length of the data window by J and improve the numerical conditioning of the algorithm.

(4) When the simplified algorithm is used, the secondary measurement v(t) is necessary even if $\alpha=0$ because it partially compensates for the information lost by omitting J-1 values of u(t) in the regressor.

(5) Note that with the approach of the Guilandanst et al. (1986, 1987a, 1987b, 1988) their working equation can not be formulated for this case with w(t)=0, since the existence of a sustained external stochastic disturbance is assumed by their approach.

The performance of the full algorithm is shown in Figure 5.2 ($\alpha=0$) and Figure 5.3 ($\alpha=0.5$) and shows excellent convergence of the output estimates to the real output. The output estimates $\tilde{y}(t)$ in Figures 5.4 and 5.5 produced by the simplified algorithm are not as good as the corresponding estimates in Figures 5.2 and 5.3, but are still a good approximation of y(t). The sacrifice in output estimation accuracy may be worthwhile since the number of estimated parameters is significantly reduced (from 22 to 3).

5.4.2 PI Feedback Control with Disturbances (J=10)

The disturbance w(t) is defined by the following series of step changes

$$w(t)=+0.7 \qquad 100(i-1)\le t<100i \qquad i=1,2,3,\ldots$$

$$w(t)=-0.7 \qquad 100i\le t\le 100(i+1) \qquad i=1,2,3,\ldots \qquad (5.92)$$

The control objective is to maintain the output at the desired value y=0 using u as the control variable. A conventional PI feedback controller is chosen for simplicity so that

$$u(It)=k_c y_b(It)+(k_c I/T_i)\sum_{i=0}^{t} y_b(Ii) \qquad t=1,2,\ldots \qquad (5.93)$$

When (93) is used with the adaptive inferential estimation algorithms a small perturbation signal is added to u to improve the excitation. y_b and I (I=1 or J) are selected for each specific case as described below.

Ideal case: The best control should be obtained when all measurements are available at the desired control interval, i.e. I=1 and $y_b(It)=y(t)$ in (5.93). Figure 5.6 shows the closed loop response with and without the process time delay. The controller constants $K_c=0.2$ and $T_i=10$ were obtained by trial and error tuning.

Practical case: With the conventional PI control scheme if J=10 then the control interval must be increased from one time unit to 10, i.e. I=J=10 and $y_b(It)=y(It)$ in (93). As expected, the control performance (solid line Figure 5.7) becomes oscillatory and control detuning is required (dotted line, T_i increased to 60). If a time delay k=5 is included the performance degradation is even more severe (results not shown).

Multirate inferential control with delay compensation: The output is sampled every J=10 intervals but estimates of the output, y_e, are produced at every control interval, i.e. I=1 and $y(It)=y_e(t+k)$ in (5.93). As shown in Figure 5.8 (k=0) and 9 (k=5) the estimated output values are very close to the true values and control performance is very close to the ideal case plotted in Figure 5.6. (However the case with time delay (Figure 5.9) is not as good as the corresponding case with no delay.)

Simplified multirate inferential control: The open-loop output estimation results in Figures 5.2-5.5 showed that the output estimation error increased when the simplified algorithm was used. However, Figure 5.10 shows that under closed loop conditions the simplified algorithm produced output estimates and control performance equal to, or better than the full algorithm. When a process delay is included (Figure 5.11) the output estimates and the control performance are again better than the full algorithm (Figure 5.9).

The fact that the simplified algorithm can result in better closed loop performance can be demonstrated by simulation examples and can be supported by intuitive arguments, e.g. the parsimony principle in parameter estimation (Ljung, 1987). However, at the present time there is no formal proof of convergence on improved performance for the simplified algorithm. Neverthless for a given application it is obviously worthwhile to evaluate both the full and simplified algorithms by simulation or experiment as shown in this section.

5.5 Conclusions

(1) A multirate, inferential estimation algorithm based on {u(t),v(t),y(Jt),t=0,1,2...}, is derived. The output convergence properties are formally proven for the case without unmeasured external stochastic disturbances.

(2) Multirate, inferential control using the output estimates, $y_e(t)$, rather than the measured values y(Jt) is significantly better than the comparable conventional, single-rate control scheme using y(Jt) and approaches that of the ideal case where the output is measured every sampling interval, i.e. the output values y(t) are used.

(3) A simplified inferential algorithm is presented which actually outperforms the full algorithm in the closed loop simulation example. However, the convergence properties are not formally proven.

(4) The algorithm has direct application in the process industries (e.g. distillation columns) where the output measurements, y(Jt) (e.g. composition), are available only at intervals J times slower than the desired control interval, and a secondary measurement v(t) (e.g. temperature) is available. It can also be used as an alternative to many conventional cascade control loops in which the outer loop operates with a sampling interval JT and the inner loop with T.

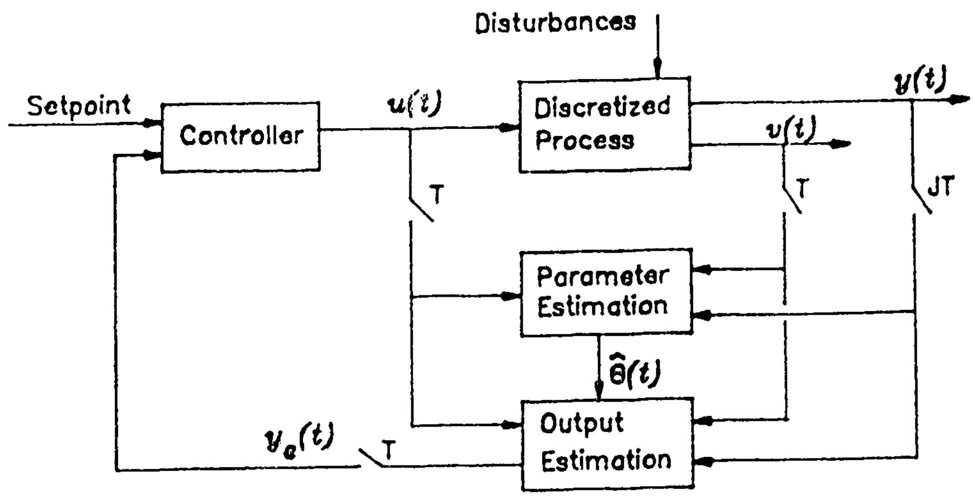

Figure 5.1: Multirate inferential estimation and control system.

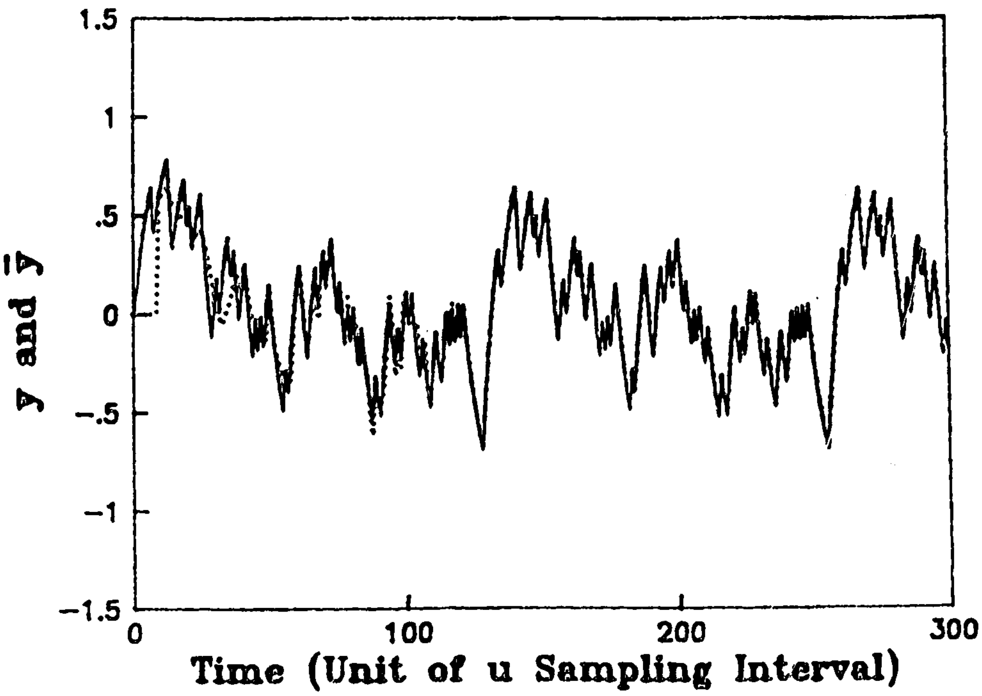

Figure 5.2: Output estimation with the full algorithm, u=PRBS, a=0 (solid line=y; dotted line=$\bar{y}$).

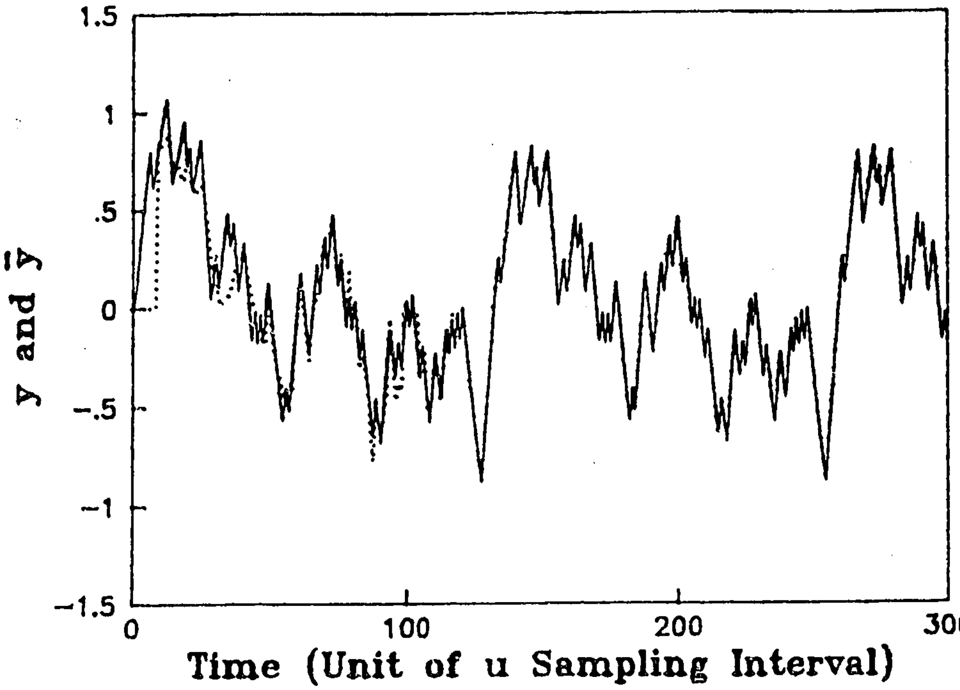

Figure 5.3: Output estimation with the full algorithm, u=PRBS, a=0.5 (solid line=y; dotted line=$\bar{y}$).

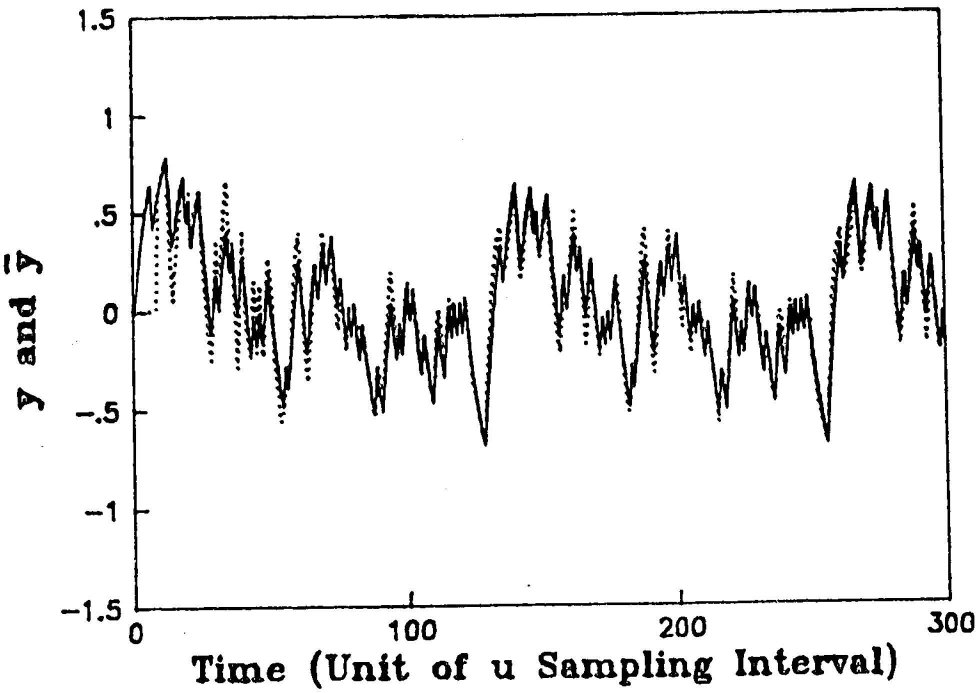

Figure 5.4: Output estimation with the simplified algorithm, u=PRBS, a=0 (solid line=y; dotted line=$\bar{y}$).

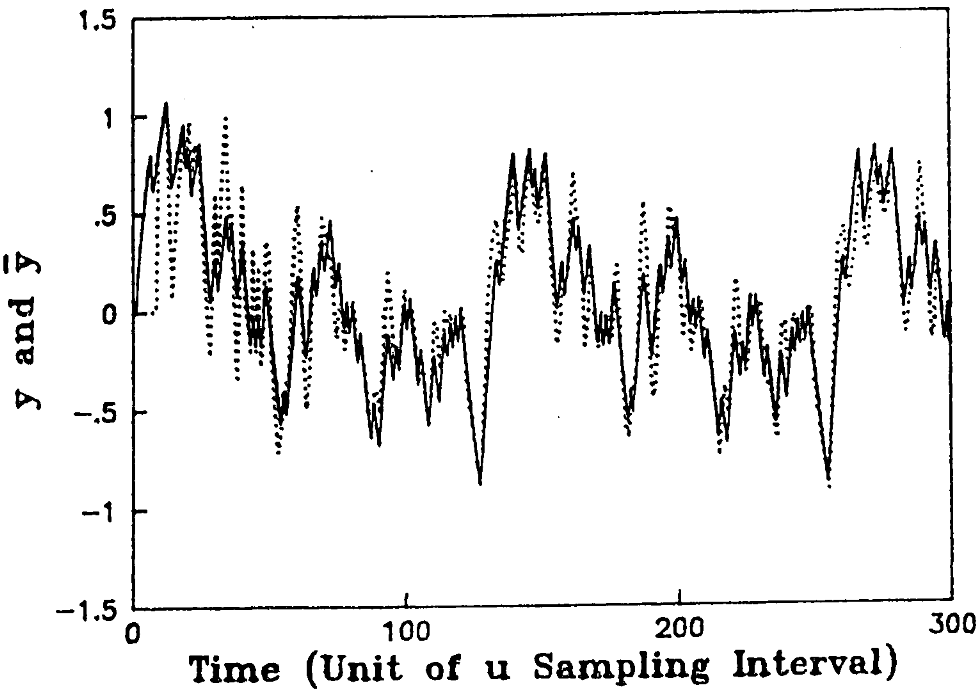

Figure 5.5: Output estimation with the simplified algorithm, u=PRBS, a=0.5 (solid line=y; dotted line=$\bar{y}$).

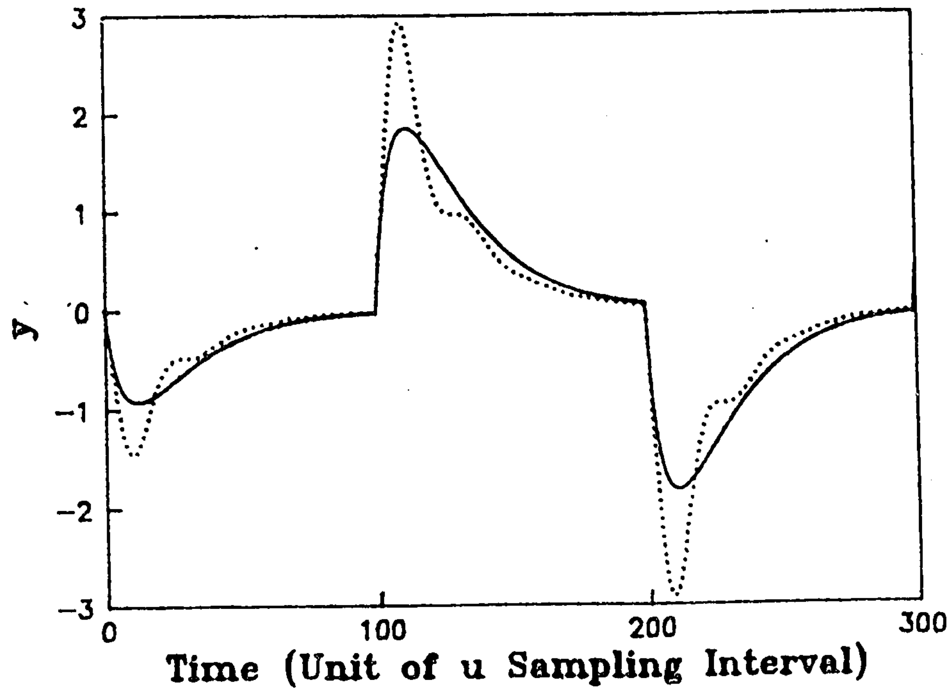

Figure 5.6: Closed loop PI control, ideal case: (using the process output y(t) at the fast control sampling intervals) with delay (dotted line) or no delay (solid line).

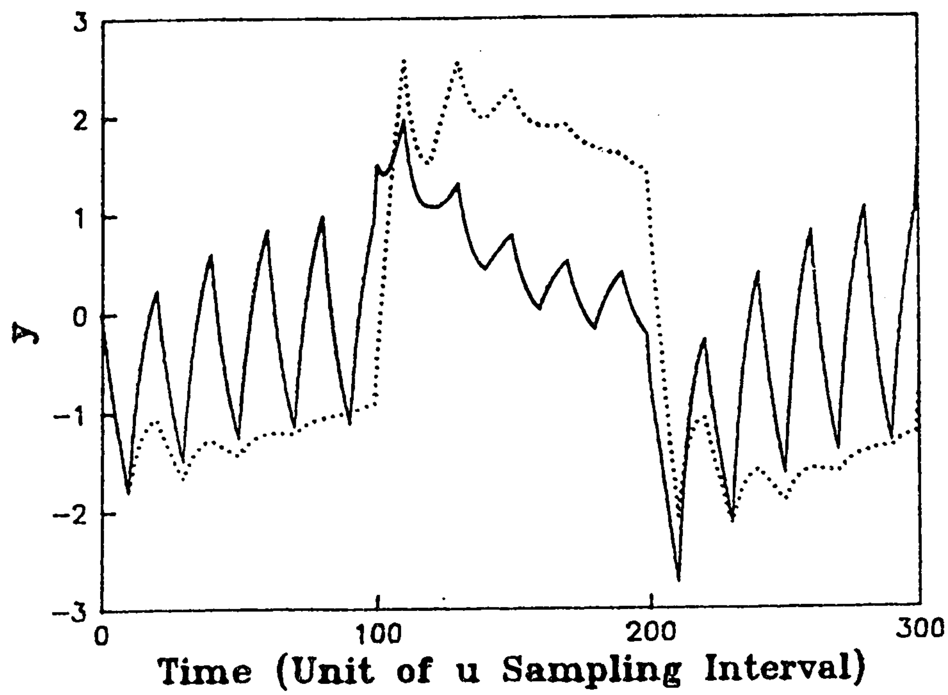

Figure 5.7: Closed loop PI control, Practical case: (using the process output y(t) at the slow sampling intervals, i.e. every JT intervals) with no delay, K_c=0.2, T_i=10 (solid line) and T_i=60 (dotted line).

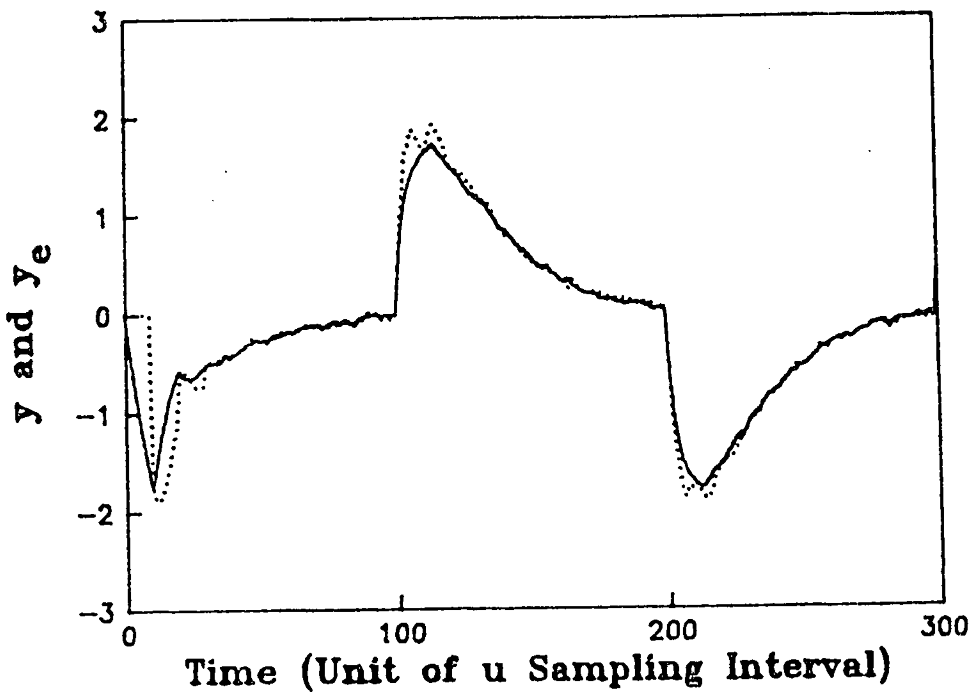

Figure 5.8: Multirate inferential control, with no delay (solid line=y; dotted line=y_e).

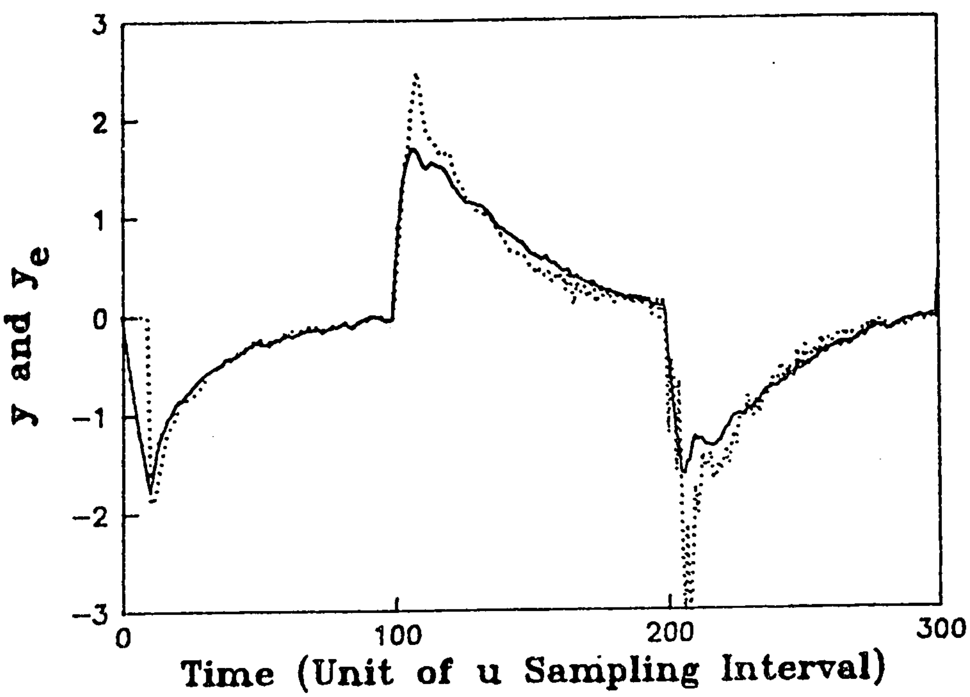

Figure 5.9: Multirate inferential control, with delay k=5 (solid line=y; dotted line=y_e).

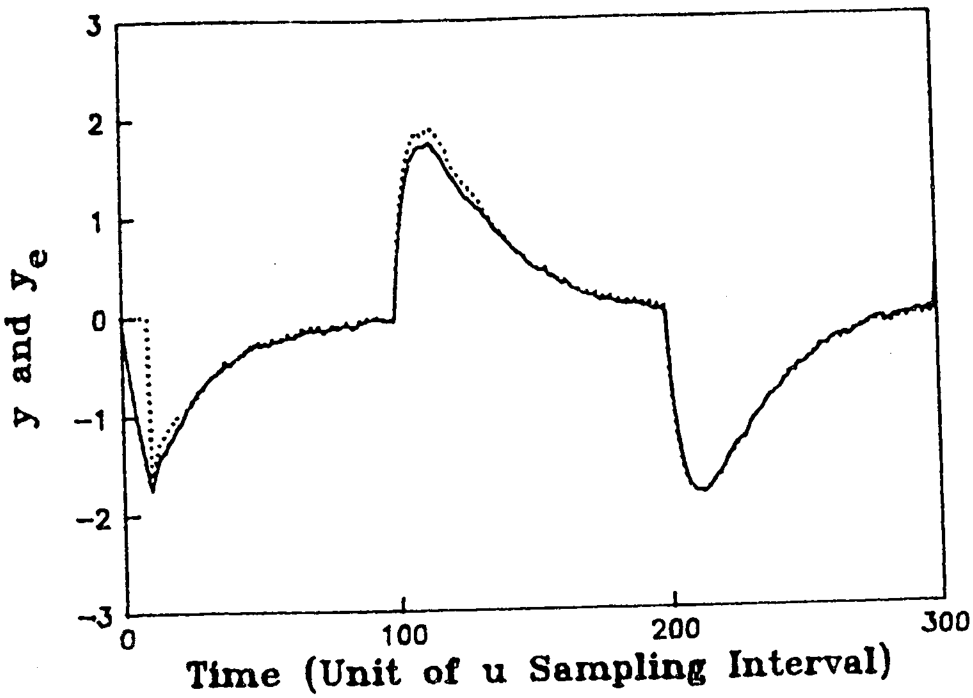

Figure 5.10: Multirate simplified inferential control, with no delay (solid line=y; dotted line=y_e).

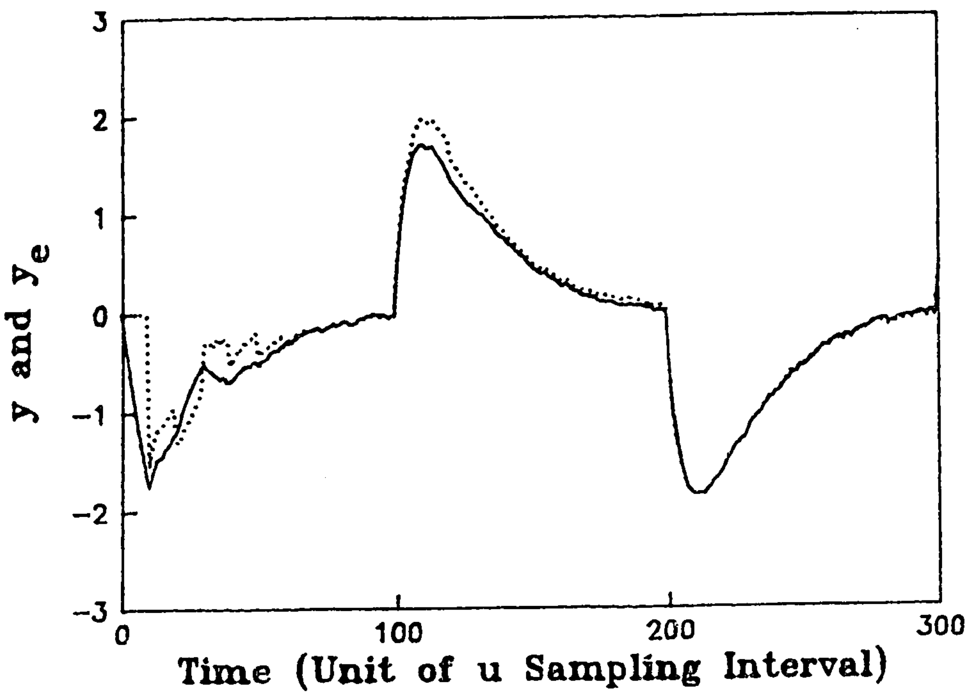

Figure 5.11: Multirate simplified inferential control, with delay (solid line=y; dotted line=y_e).

5.6 References

Astrom K. J. Introduction to Stochastic Control Theory. Academic Press, 1970.

Goodwin G. C. and Sin K. S., Adaptive Filtering Prediction and Control, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

Guilandoust M. T., Morris A. J. and Tham M. T., "Estimation and Control of Distillation Product Composition Using Tray Temperature Measurments," IFAC Symposium on Dynamics and Control of Chemical Reactors and Distillation Columns, pp. 203-208, Bournemouth, UK, 1986.

Guilandoust M. T., Morris A. J. and Tham M. T., "Adaptive Estimation in Inferential Process Control," Proc. American Control Conf., pp1743-1748, 1987a.

Guilandoust M. T., Morris A. J. and Tham M. T., "Adaptive Inferential Control," IEE Proc. vol. 134, Pt. D, NO. 3, pp. 171-179, May 1987b.

Guilandoust M. T., Morris A. J. and Tham M. T., "An Adaptive Estimation Algorithm for Inferential Control," Ind. Eng. Chem. Process Des Dev., 27, pp. 1658-1664, 1988.

Ljung L., System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

Parrish, J. R. and Brosilow, C. B., "Inferential Control Applications," Automatica, 21, (5), pp. 527-538, 1985.

Patke, N. G., Deshpande, P. B. and Chou, A. C., "Evaluation of Inferential and Parallel Cascade Schemes for

Distillation Control," Ind. Eng. Chem. Process Des Dev., 21, pp. 266-272, 1982

Soderstrom T., Methods and Applications in Adaptive Control, Springer Verlag, 1980.

Wolovich, W. A., Linear Multivariable System, Springer Verlag, New York, 1974.

6. RLS PARAMETER CONVERGENCE WITH OVERPARAMETERIZED MODELS[5]

6.1 Introduction

Overparameterization is frequently used in parameter identification applications. For example, the identified model is often overparameterized to avoid model mismatch when the delay or the order of the plant model is not exactly known, or when more design freedom is required in an adaptive control strategy. However, most of the theoretical analyses of parameter convergence to date require that the exact model order be known.

One significant study which does apply to overparameterized systems is that by Ljung (1987). He showed that for prediction error type algorithms the criterion minimizing arguments converge, if the system signals are **informative**, to a set, each point of which results in the same input-output relations as that of the plant. (To use this result it is necessary to show that the estimated parameters for a particular prediction type algorithm asymptotically attain the values of the criterion minimizing arguments. This can be done for RLS by applying the ordinary differential equation (ODE) technique (Ljung 1977, Ljung and Soderstrom 1983) subject to some regularity conditions.) However, Ljung's results are based on the very restrictive

----

[5] A version of this chapter has been published: Lu W. and Fisher D. G., System and Control Letters, vol. 12, No. 2, pp. 133-138, Feb. 1989.

assumption of quasi-stationary data. (A signal u(t) is quasi-stationary only if
$\lim_{T\to\infty} \sum_{t=1}^{T} u^2(t)/T$ **exists** and is **finite** (Ljung 1987).) Since many practical applications and theoretically significant systems do not satisfy this quasi-stationary data assumption, several authors (Solo 1979, Anderson and Johnson 1982, Moore 1983, Lai and Wei 1986a) have used the alternative Lyapunov type function approach (also referred to as a Martingale approach for stochastic systems). The Lyapunov type function approach provides greater insight into the algorithms. For example it provides a basis for studying convergence rates (Anderson and Johnson 1982) and a basis for obtaining parameter convergence conditions in the presence of possibly unbounded signals (Goodwin and Teoh 1985) or in terms of "much weaker types of excitation conditions than persistent excitation" (Lai and Wei 1986b), etc. Unfortunately, in order to prove parameter convergence, all of these authors using the Lyapunov type function approach had to assume that the system order was not overestimated.

This chapter shows how the Lyapunov type function approach can be extended to prove parameter convergence for overparameterized systems in such a way that all the advantages of this approach for minimum systems are preserved for overparameterized systems. For example, by using this approach the properties of the common factor polynomials defined by the limiting set of parameter estimates can be addressed and the excitation requirements can be relaxed. The approach used in this chapter is new and has application beyond the RLS case treated in this chapter.

Recently there have been a number of studies dealing with parameter convergence for overparameterized systems (Goodwin et al. 1985, Janecki 1987, Xia and Moore 1987, Heymann 1988, Xia and Moore 1988). Goodwin et al. showed a persistency of excitation result for nonminimal models of systems having purely deterministic disturbances. The same result was obtained by Janecki (1987) using a more direct approach. Xia et al. (1987,1988) showed that when the RLS or ELS algorithm is applied to overparameterized models it is possible, by modifying the data regressor in the algorithm, to insure the persistent excitation condition for the modified data regressor and thus guarantee parameter convergence to a unique point. Heymann (1988) showed that the parameters in structured nonminimal models can be uniquely determined if and only if a certain design identity has a unique solution. To insure parameter convergence, all of these results implicitly require that the degree of overparameterization be known or the common factor polynomial be uniquely fixed by the estimated model structure. Note that if the system signals are quasi-stationary then the parameter convergence results of Goodwin et al. (1985), Janecki (1987) and Heymann (1988) are essentially covered by those of Ljung (1987). Thus one of the main contributions of these papers (Goodwin et al. 1985, Janecki 1987, Heymann 1988) is to avoid the quasi-stationary data assumption.

This chapter proves parameter convergence for RLS applied to overparameterized systems without assuming quasi-stationary data, without requiring any priori knowledge of the degree of overparameterization and without modifying the RLS algorithm.

6.2 Main convergence Results

The following set notation is used. If W is a vector space over the real scalar field R, $G \subset W$, $F \subset W$, $w \in W$ and $r \in R$, let

$$w+G \equiv \{w+g:\ g \in G\},$$

$$w-G \equiv \{w-g:\ g \in G\},$$

$$G+F \equiv \{g+f:\ g \in G,\ f \in F\},$$

$$rG \equiv \{rg:\ g \in G\}$$

etc. .

Assume that the plant is represented by the n th order irreducible DARMA model

$$(1+\sum_{i=1}^{n} a_i q^{-i})y(t)=(\sum_{i=1}^{n} b_i q^{-i})u(t) \qquad (6.1)$$

and the RLS, (3.3.46)-(3.3.47) of (Goodwin and Sin 1984), is used to identify the parameters of (6.1):

$$\hat{\theta}(t)=\hat{\theta}(t-1)+ P(t-2)\phi(t-1)[y(t)-\phi(t-1)^T\hat{\theta}(t-1)]$$
$$/[1+\phi(t-1)^T P(t-2)\phi(t-1)] \qquad t\geq 1 \qquad (6.2)$$

$$P(t-1)=P(t-2)-P(t-2)\phi(t-1)\phi(t-1)^T P(t-2)$$
$$/[1+\phi(t-1)^T P(t-2)\phi(t-1)] \qquad t\geq 1 \qquad (6.3)$$

with $\hat{\theta}(0)$ and $P(-1)>0$ given, and

$$\hat{\theta}(t)=[\hat{a}_1(t),\ldots,\hat{a}_{n+m}(t),\hat{b}_1(t),\ldots,\hat{b}_{n+m}(t)]^T, \tag{6.4}$$

$$\phi(t-1)=[-y(t-1),\ldots,-y(t-(n+m)),u(t-1),\ldots,u(t-(n+m))]^T, \tag{6.5}$$

where $1\le m<\infty$. In other words, RLS is applied to identify a model from the following model set

$$\{(1+\sum_{i=1}^{n+m}\bar{a}_i q^{-i})y(t)=(\sum_{i=1}^{n+m}\bar{b}_i q^{-i})u(t): [\bar{a}_1,\ldots,\bar{a}_{n+m},\bar{b}_1,\ldots,\bar{b}_{n+m}]^T\in R^{2(n+m)}\}. \tag{6.6}$$

An important subset of (6.6) is

$$\{(1+\sum_{i=1}^{n}a_i q^{-i})(1+\sum_{j=1}^{m}c_j q^{-j})y(t)=(\sum_{i=1}^{n}b_i q^{-i})(1+\sum_{j=1}^{m}c_j q^{-j})u(t): c\equiv[c_1,\ldots,c_m]^T\in R^m\}. \tag{6.7}$$

Any element of (6.7) contains the plant model (6.1) and a monic common factor polynomial of order $\le$ m. Fixing $c\in R^m$, let

$$\theta_c\equiv[a_{1c},\ldots,a_{n+mc},b_{1c},\ldots,b_{n+mc}]^T \tag{6.8}$$

be the parameter vector of model (6.7). By expressing the θ_c versus c relation in an explicit matrix form, it easily follows that

$$H\equiv(\theta_c:c\in R^m) \tag{6.9}$$

is an m dimensional linear hypersurface of $R^{2(n+m)}$, or more precisely, for any fixed $h_0\in H$, $H-h_0$ is an m dimensional subspace of $R^{2(n+m)}$.

<u>Theorem 6.1:</u>

For the RLS algorithm applied to the DARMA model (6.1) with overparameterization of degree m, if

$$\lim_{t\to\infty}\lambda_{2n+m}(t)=\infty \tag{6.10}$$

where $\lambda_i(t)$, i=1,...,2(n+m), is defined to be i th largest eigenvalue of $P(t)^{-1}$,

then $\lim_{t\to\infty} \inf_{h\in H\cap\theta+B_e} \|\hat{\theta}(t)-h\|=0$ (6.11)

where $B_r\equiv(z\in R^{2(n+m)}: \|z\|\le r)$ for $r>0$, (6.12)

$e\equiv\sqrt{\kappa}\,\|\hat{\theta}(0)-\theta\|$, (6.13)

$\kappa\equiv$the condition number of $P(-1)^{-1}$, (6.14)

θ is the point on H which is closest to $\hat{\theta}(0)$, i.e. $\|\theta-\hat{\theta}(0)\|=\inf_{h\in H}\|h-\hat{\theta}(0)\|$. Such a point θ exists and is unique since H is closed and convex.

Theorem 6.2:

Under the conditions of Theorem 6.1, if the condition number, κ, of $P(-1)^{-1}$, is unity, then

$\lim_{t\to\infty} \hat{\theta}(t)=\theta$ if $m\ge1$ (6.15)

and $\hat{\theta}(t)\in S$ $\forall t$ if $m\ge2$ (6.16)

where $S\equiv(s\in R^{2(n+m)}: s=r(\hat{\theta}(0)-\theta)+\theta,\ |r|\le1)$, (6.17)

and θ is the unique point as defined in Theorem 6.1.

Remarks on the Theorems:

(1) Theorem 6.1 implies that the limiting set of parameter estimates always belongs to H, which represents the equivalence class model set (6.7), any element of which is a true model of the plant. This conclusion implied by Theorem 6.1 can be obtained by Ljung's approach if the quasi-stationary data and $\lim_{t\to\infty} \lambda_{2n+m}(t)/t>0$ are assumed. Theorem 6.1 also points out that the initial conditions of the algorithm determine the bounded region, $H\cap\theta+B_e$, of the

limiting set on H or equivalently the common factor polynomials represented by this bounded region. Note that parameter convergence to H is global since $H \cap \theta + B_e \subset H$ always holds irrespective of the initial estimate which determines the position of θ on H.

(2) For parameter convergence to the set H, the condition ($\lim_{t\to\infty} \lambda_{2n+m}(P(t)^{-1})=\infty$) is the weakest possible and compatible with the "much weaker types of excitation conditions than persistent excitation" discussed by Lai and Wei (1986b) as well as the "persistency of excitation in the presence of possibly unbounded signals" discussed by Goodwin and Teoh (1985), while Ljung's condition (Ljung, 1987) (which implies that $P(t)^{-1}/t$ **converges** to a semipositive definite matrix with a rank greater or equal than 2n+m) is significantly stronger and is not applicable to more general signals.

(3) The geometric meaning of Theorem 6.2 is very informative. S is a straight line section which is normal to the linear hypersurface H with $\hat{\theta}(0)$ at one end and $\theta \in H$ as the symmetric point. This symmetric point θ is uniquely defined by model (6.1), the degree of overparameterization m, and the initial value of the parameter estimate $\hat{\theta}(0)$.

(4) If $\hat{\theta}(0) \in H$ then $\hat{\theta}(t)=\hat{\theta}(0)$, $\forall t \geq 0$ and Theorem 6.1 and 6.2 are trivially verified.

6.3 Proof of Theorem 6.1 and 6.2

Several lemmas are presented first to facilitate the proof of Theorems 6.1 and 6.2.

Lemma 6.1: $\forall t \geq 0$ there exists an orthogonal matrix Q(t) such that

$$P(t)^{-1} = Q^T(t)\,\mathrm{diag}(\lambda_1(t), \ldots, \lambda_{2(n+m)}(t))\,Q(t) \qquad (6.18)$$

where $\lambda_i(t) \geq 0$, $i=1,2,\ldots,2(n+m)$ is the i^{th} largest eigenvalue of $P(t)^{-1}$.

Proof:

The existence of Q(t) is obvious since $P(t)^{-1}$ is real symmetric and positive. ∎

$$\text{Let } \tilde{\theta}_c(t) \equiv \hat{\theta}(t) - \theta_c \qquad (6.19)$$

and

$$V_c(t) \equiv \tilde{\theta}_c(t)^T P(t-1)^{-1} \tilde{\theta}_c(t) \qquad (6.20)$$

Lemma 6.2: $V_c \geq 0$ is nonincreasing $\forall c \in R^m$. Thus

$$\lim_{t\to\infty} V_c(t) \equiv V_c(\infty) < \infty \qquad (6.21)$$

exists. Note that $V_c(\infty) < \infty$ is c dependent and is not uniformly bounded.

Proof:

Note that any θ_c, $c \in R^m$, can be used as the true parameter 'θ_0' in Lemma 3.3.6. of (Goodwin and Sin 1984), since each model in the model set (6.7) is a true plant model as well as model (6.1). Then this lemma is proven by simply applying Lemma 3.3.6. of (Goodwin and Sin, 1984). ∎

Lemma 6.3:

$$\text{Let } x_c(t) \equiv Q(t-1)\tilde{\theta}_c(t) \qquad (6.22)$$

$$\text{Then } x_c(t)^T x_c(t) = \tilde{\theta}_c(t)^T \tilde{\theta}_c(t) \qquad (6.23)$$

and $V_c(t)=\sum_{i=1}^{2(n+m)} \lambda_i(t-1)x_{ci}^2(t)$ (6.24)

where $x_{ci}(t)$, $i=1,\ldots,2(n+m)$ is the i th component of $x_c(t)$.

Proof:

Combining (6.18), (6.20) and (6.22) yields the proof since Q(t) is orthogonal. ∎

Lemma 6.4:

Let the conditions of Theorem 6.1 be satisfied. Then

$$\forall c\in R^m \quad \lim_{t\to\infty} \inf_{x\in X}\|x_c(t)-x\|=0 \qquad (6.25)$$

where $X\equiv\{x\in R^{2(n+m)}:\ x_i=0,\ 1\le i\le 2n+m\}$ (6.26)

Proof:

Since $V_c(\infty)<\infty$ $\forall c$, $\lambda_i(t)\ge 0$, $\forall 1\le i\le 2(n+m)$ and $t\ge 0$ and $\lim_{t\to\infty}\lambda_i(t)=\infty$ $1\le i\le 2n+m$, we have $\lim_{t\to\infty} x_{ci}(t)=0$, $1\le i\le 2n+m$ by (6.24) of Lemma 6.3. This immediately gives (6.25). ∎

Proof of Theorem 6.1:

From Lemma 6.4 we have:

$\forall c\in R^m$ and $\epsilon 1>0$ there exists $0<T_c<\infty$ such that

$$x_c(t)=Q(t-1)\tilde{\theta}_c(t)\in X+B_{\epsilon 1} \qquad \forall t>T_c \qquad (6.27)$$

where $B_{\epsilon 1}$ is as defined in (6.12). Since Q(t) is orthogonal, there exists $0<d<\infty$ such that $\|Q(t)^T\|\le d$ $\forall t$. Here d may depend on the matrix norm used. Letting $\epsilon\equiv d\cdot\epsilon 1$ we have $Q(t-1)^{-1}B_{\epsilon 1}=Q(t-1)^TB_{\epsilon 1}\subset B_\epsilon$ $\forall t$. Thus by (6.27)

$$\hat{\theta}(t)\in\theta_c+Q(t-1)^TX+B_\epsilon \qquad \forall t>T_c \qquad (6.28)$$

Since $\theta\in H$, by using a similar argument to that of Lemma 6.2, Lemma 3.3.6. of (Goodwin and Sin, 1984) shows that

$$\|\hat{\theta}(t)-\theta\|\le e \qquad \forall t \tag{6.29}$$

Therefore $\hat{\theta}(t)\in(\theta_c+Q(t-1)^T X+B_\epsilon)\cap(\theta+B_e) \qquad \forall t>T_c$ (6.30)

Since

$$Q(t)^T\in\Omega_d\equiv\{\|U\|\le d,\ U\in\text{set of } [2(n+m)]^2 \text{ matrices}\} \qquad \forall t\ge 0 \tag{6.31}$$

and Ω_d is compact, there exists a subsequence $\{Q(t_k-1)^T\}^{\infty}_{k=1}$ which converges in the matrix norm to some $\bar{Q}^T\in\Omega_d$. It is easy to see that $\bar{Q}^T$ is also orthogonal. Since $\lim_{k\to\infty} Q(t_k-1)^T=\bar{Q}^T$ and $(\theta_o+Q(t_k-1)^T X+B_\epsilon)\cap(\theta+B_e)$ is **bounded**, there exists $\bar{T}_c$, $T_c\le\bar{T}_c<\infty$, such that

$$\hat{\theta}(t_k)\in(\theta_c+Q(t_k-1)^T X+B_\epsilon)\cap(\vartheta+B_e)$$
$$\subset(\theta_c+\bar{Q}^T X+B_{2\epsilon})\cap(\theta+B_e) \qquad \forall t_k>\bar{T}_c \tag{6.32}$$

Indeed the linear hypersurface $\theta_c+\bar{Q}^T X$ which contains θ_c, does not depend on $\theta_c\in H$, and thus contains H. To see this, note that for h_1, $h_2\in H$, $h_1+\bar{Q}^T X$ and $h_2+\bar{Q}^T X$ are two parallel linear hypersurfaces (i.e. the distance from any point on one hypersurface to the other is constant) since X is an m dimensional subspace and so is $\bar{Q}^T X$, and that $h_1+\bar{Q}^T X$ and $h_2+\bar{Q}^T X$ either coincide with each other or are separated by a distance greater than zero. The latter case is impossible since from (6.32) $h_1+\bar{Q}^T+B_{2\epsilon}\cap h_2+\bar{Q}^T+B_{2\epsilon}$ is not empty $\forall\epsilon>0$. Since the dimension of H=m=the dimension of $\theta_c+\bar{Q}^T X$ and $H\subset\theta_c+\bar{Q}^T X$ it follows that

$$H=\theta_c+\bar{Q}^T X \qquad \forall\theta_c\in H, \tag{6.33}$$

which means that $\theta_c+\bar{Q}^T X$ $\forall\theta_c\in H$ and H are the same element in the quotient space $R^{2(n+m)}$ modulo $\bar{Q}^T X$ (Rudin, 1973). Since $2\epsilon>0$ is arbitrary, (6.32) is equivalent to

$$\liminf_{k\to\infty} \inf_{h\in H\cap\theta+B_e} \|\hat{\theta}(t_k)-h\|=0 \qquad (6.34)$$

Similar arguments can be applied to show that every subsequence of $\{\hat{\theta}(t)\}^{\infty}_{t=1}$ has a subsequence having the property of (6.32) like $\{\hat{\theta}(t_k)\}^{\infty}_{k=1}$. This just implies (6.11). ■

<u>Proof of Theorem 6.2:</u>

Again from Lemma 3.3.6. of (Goodwin and Sin, 1984)

$$\|\hat{\theta}(t)-h\|^2\le\kappa\|\hat{\theta}(0)-h\|^2= \|\hat{\theta}(0)-h\|^2 \qquad \forall h\in H \qquad (6.35)$$

Suppose $m\ge2$, $\hat{\theta}(0)\notin H$ (if $\hat{\theta}(0)\in H$, then $\hat{\theta}(t)=\hat{\theta}(0)\in S$, $\forall t\ge0$) and $\hat{\theta}(t)\notin S$ for some $t>0$. Let

$$h_r\equiv r(\theta-h(t))+h(t) \qquad r\ge1 \qquad (6.36)$$

where θ is as defined in Theorem 6.1 and $h(t)$ is the unique point on H closest to $\hat{\theta}(t)$. Since dimension of $H=m\ge2$ H contains at least a two dimensional plane in $R^{2(n+m)}$. Therefore, if r is sufficiently large, it is easy to show that $\|\hat{\theta}(t)-h_r\|^2>\|\hat{\theta}(0)-h_r\|^2$, which is a contradiction to (6.35). To see this, let $(x_0,y_0)\in R^2$ and fix $y_0>0$. Let $y_0\ge\epsilon_x>0$, $y_0\ge\epsilon_y\ge0$. Then

$$(x_0+\epsilon_x)^2+(\pm(y_0-\epsilon_y))^2-(x_0{}^2+y_0{}^2)>0$$

if $x_0>0$ is sufficiently large. In our case, the plane determined by $\hat{\theta}(0)$, $\hat{\theta}(t)$ and θ corresponds to R^2, $\hat{\theta}(0)$ corresponds to $(0,y_0)$, $\hat{\theta}(t)$ corresponds to $(-\epsilon_x,\pm(y_0-\epsilon_y))$, θ corresponds to $(0,0)$ and h_r corresponds to $(x_0,0)$. Thus (6.16) is proven. If $m=1$, H is a line in $R^{2(n+m)}$ and the same argument shows that $\hat{\theta}(t)$ $\forall t$ falls in a plane which contains S and has H as its normal line. Since θ is the only point in the intersection of this plane and H, $\lim_{t\to\infty} \hat{\theta}(t)=\theta$ by Theorem 6.1. If $m\geq 2$ we also have

$\lim_{t\to\infty} \hat{\theta}(t)=\theta$ by Theorem 6.1 since $S\cap H=\theta$. Thus (6.15) is proven. ■

Remarks on the Proofs:

1) The above two proofs provide clear geometric pictures. For example, when $\kappa=1$ and $m\geq 2$, the parameter estimates follow the shortest straight path from the initial value to the linear hypersurface H.

2) Both $P(t)^{-1}/t$ and $Q(t)^T$ may not converge to anything at all in general for a wide class of signals. This is the most difficult part in achieving the result of Theorem 6.1. The argument of compactness and quotient space overcomes this difficulty and indeed shows that every converging subsequence of $Q(t)^T$ must yield the same m dimensional subspace $\bar{Q}^T X$ regardless of whether $\bar{Q}^T$ depends on the individual subsequence.

3) To keep the condition of Theorem 6.1 in its weakest possible form we use (6.10) instead of a more specific condition imposed on u(t). When the plant model (6.1) is stable a sufficient condition for (6.10) to hold is that u(t) be weakly persistently exciting of order 2n+m (Definition 3.4.B. of (Goodwin and Sin, 1984)). The basic

idea in this argument is that a state space realization of

$[-y(t-1),\ldots,-y(t-n)),u(t-1),\ldots,u(t-(n+m))]^T$ versus

u(t) as the input is completely reachable even though that of the full regressor

$[-y(t-1),\ldots,-y(t-(n+m)),u(t-1),\ldots,u(t-(n+m))]^T$ versus

u(t) is not. More generally, condition (6.10) can be satisfied by a very wide class signals, e.g. those discussed by Lai and Wei (1986a) and Goodwin and Teoh (1985).

6.4 Conclusions

In this chapter it is proven that for RLS applied to an overparameterized DARMA model the limiting set of parameter estimates belongs to a set, H, each element of which contains the irreducible plant model plus a monic common factor polynomial. Furthermore the common factor polynomials defined by the limiting set of parameter estimates are shown to be quantitatively related to the initial conditions of the algorithms and the conditions on the required excitation are the weakest possible.

By introducing the concept of quotient space the approach taken in this chapter preserves all the merits of the Lyapunov type function approach and makes it generally applicable to both minimal and non-minimal systems.

6.5 References

Anderson B. D. O. and Johnson Jr. C. R., "Exponential Convergence of Adaptive Identification and Control Algorithms," Automatica vol. 18, pp. 1-13, Jan. 1982.

Goodwin G. C., Norton J. P., and Viswanathan M. N., "Persistency of excitation for nonminimal models of systems having purely deterministic disturbances," IEEE Trans. Automat. Contr., vol. AC-30, pp. 589-590, June 1985.

Goodwin G. C. and Sin K. S., Adaptive Filtering Prediction and Control, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

Goodwin G. C. and Teoh E. K., "Pesistency of Excitation in the Presence of Possibly Unbounded Signals," IEEE Trans. Automat. Contr., vol. AC-30, pp. 595-597, 1985.

Heymann M., "Persistency of Excitation Results for Structured Nonminimal Models," IEEE Trans. Automat. Contr., vol. AC-33, pp. 112-116, 1988.

Janecki D., "Remarks on Persistency of Excitation for Nonminimal Models of Systems Having Purely Deterministic Disturbances," IEEE Trans. Automat. Contr., vol. AC-32, pp. 1123-1125, 1987.

Lai T. L. and Wei C. Z., "Extended Least Squares and Their Applications to Adaptive Control and Prediction in Linear Systems," IEEE Trans. Automat. Contr., vol. AC-31, pp. 898-906, 1986.

Lai T. L. and Wei C. Z., "On the Concept of Excitation in

Least Squares Identification and Adaptive Control," Stochastics, vol. 16, pp. 227-254, 1986.

Ljung L., "Analysis of Recursive Stochastic Algorithms," IEEE Trans. Automat. Contr., vol. AC-22, pp. 551-575, 1977.

Ljung L., System Identification: Theory for the User, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

Ljung L. and Soderstrom T., Theory and Practice of Recursive Identification. Cambridge, MA: M.I.T. Press, 1983.

Moore J. B., "Persistence of Excitation in Extended Least Squares," IEEE Trans. Automat. Contr., vol. AC-28, pp. 60-68, Jan. 1983.

Rudin W., Functional Analysis. McGraw-Hill, New York, 1973.

Solo V., "The Convergence of AML," IEEE Trans. Automat. Contr., vol. AC-24, pp. 958-962, 1979.

Xia L., Moore J. B. and Gevers M., "On adaptive estimation and pole assignment of overparameterized systems", Int. J. of Adaptive Contr. and Signal Processing, vol. 1, p143-160, 1987.

Xia L., Moore J. B., "Recursive Identification of Overparameterized Systems," Int. Rept., Dept. of System Engineering, Australian National University, Canberra, Australia, 1988.

7. A GENERAL APPROACH TO PARAMETER CONVERGENCE WITH OVERPARAMETERIZED MODELS'

7.1 Introduction

Nonminimal or overparameterized models occur in many parameter identification applications. For example, the identified model is often overparameterized when the delay or the order of a plant model is not exactly known, or when some specific structure is required due to restrictions on the observed data as illustrated by the multirate systems described in the previous chapters. Other typical examples can be seen in direct adaptive control applications where the identified models are specially structurally overparameterized (Heymann 1988) or when internal models of the external disturbances are included. The latter can result in overparameterized models either because the exact disturbance order is unknown or because the disturbance model changes order, e.g. when the disturbance temporarily goes to zero. However, almost all the theoretical analyses of parameter convergence to date (Ljung 1977, Solo 1979, Anderson and Johnson 1982, Goodwin and Sin 1984, Moore 1983, Lai and Wei 1986a) require that the exact model order be known.

Recently parameter convergence with overparameterized

'A version of this chapter has been submitted for publication: Lu W. and Fisher D. G., IEEE Trans. Automat. Contr., 1988.

models has received a great deal of attention in the literature (Goodwin et al. 1985, Janecki 1987, Xia et al. 1987, Heymann 1988). Goodwin et al. (1985) showed a persistency of excitation result for nonminimal models of systems having purely deterministic disturbances. The same result was obtained by Janecki (1987) using a more direct approach. Xia et al. (1987) showed that when the RLS algorithm is applied to overparameterized models it is possible, by modifying the data regressor in the algorithm, to insure the persistent excitation condition for the modified data regressor and thus guarantee parameter convergence to a unique point. Heymann (1988) showed that the parameters in structured nonminimal models can be uniquely determined if and only if a certain design identity has a unique solution. This was shown to be related to the output-reachability of an associated-signal system which in turn can be used to develop persistency of excitation results. To insure parameter convergence, all of these results implicitly require that the degree of overparameterization be known.

In a context of minimum variance adaptive control Becker et al. (1985) used the stochastic approximation algorithm to show that if the control system does not possess reduced order controllers (i.e. if the order of the control system is overspecified), then the limit of the parameter estimates is a hypersphere of dimension strictly less than that of the parameter vector space, and is contained in a set, each member of which yields a minimum variance control law. The geometric properties they showed are properties of the overall control system and not those of the plant. No conclusion regarding the convergence of the parameter estimates to the true values is made.

Ljung (1987) presented some general results that can be applied to overparameterized systems. He showed that for prediction error type algorithms the criterion minimizing arguments converge, under certain existence conditions, to a set, each point of which results in the same input-output relation as that of the plant. To show that the parameter estimates asymptotically attain the values of the criterion minimizing arguments the ordinary differential equation (ODE) technique (Ljung 1977) can be applied if some regularity conditions are satisfied. The key restriction in Ljung's approach is the quasi-stationary data assumption (Ljung 1987) which does not apply to a wide class of signals.

Using a Martingale approach several authors (Ljung 1977, Solo 1979, Anderson and Johnson 1982, Goodwin and Sin 1984, Moore 1983, Lai and Wei 1986a) showed that the parameter estimation error is related to a deterministic or stochastic Lyapunov type function. Parameter convergence properties for minimal models were determined from the properties of this Lyapunov function without requiring the quasi-stationary data assumption. Note that some authors (Ljung 1976, Solo 1979, Goodwin and Sin 1984) still made the quasi-stationary data assumption, i.e. the existence of some limits, although it is not strictly necessary in their derivation.

This chapter presents two general convergence results (Theorems 7.1 and 7.4) closely related to the Lyapunov type function approach described above. These results can be applied to obtain parameter convergence properties without the quasi-stationary data assumption when nonminimal models with an unknown degree of overparameterization are involved. Theorem 7.1 shows that under conditions similar to those required for parameter convergence in the minimal model cases, if the input and the noise of the plant are sufficiently rich in an overparameterized sense, e.g. u(t) is weakly persistently exciting of order 2n+r (Definition 3.4.B of Goodwin and Sin 1984) instead of order 2n etc., where n is the plant order and r is the degree of overparameterization, then the parameter estimates at least converge to a set, say H, each element of which gives a model containing the irreducible plant model plus a common factor polynomial. The application of Theorem 7.1 is demonstrated by RLS and ELS algorithms (Theorem 7.2 and 7.3). Theorem 7.4 extends the convergence results to overparameterized systems with structured constraints (i.e. the constrained model structure allows exact model matching with the plant model) and is applied to RLS with overparameterized DARMA models subject to parameter constraints (Theorem 7.5) or a constraint on the initial covariance matrix (Theorem 7.6). The use of Theorem 7.4 is also illustrated by considering conditions for parameter convergence for the nonminimal systems having purely deterministic disturbances (Theorem 7.7) previously considered by Goodwin et al. (1985) and Janecki (1987). Remarks are also made to show that Theorem 7.4 can be used to obtain: (i) the parameter convergence result for the structured nonminimal models presented by Heymann (1988), (ii) an alternative and simpler approach to the problem, presented by Xia et al. (1987), of ensuring a unique parameter convergence point when an overparameterized model is used in an adaptive control algorithm, and (iii) the parameter convergence result of the multirate identification algorithm of Chapter 3.

7.2 Parameter Convergence with Minimal Models

This section presents a brief review of the literature on parameter convergence with minimal models, which provides a basis for the nonminimal model results developed in the later sections.

Consider the linear system

$$A(q^{-1})y(t)=B(q^{-1})u(t)+C(q^{-1})w(t) \quad (7.1)$$

where $A(q^{-1})=1+a_1q^{-1}+...+a_pq^{-p}$, (7.2)

$$B(q^{-1})=b_0+b_1q^{-1}+...+b_lq^{-l}, \quad (7.3)$$

$$C(q^{-1})=1+c_1q^{-1}+...+c_mq^{-m}, \quad (7.4)$$

the sequence $\{w(t)\}$ is a real stochastic process defined on a probability space $(\Omega,\mathbf{F},\mathbf{P})$ adapted to the sequence of increasing sub-sigma algebras ($\mathbf{F}_t$, t=0,1,...), where $\mathbf{F}_t$ is generated by the observations up to time t, and such that {w(t)} satisfies

$$E\{(w(t)|\mathbf{F}_{t-1}\}=0 \qquad \text{a.s. (almost surely)} \qquad (7.5)$$

$$E\{(w(t)^2|\mathbf{F}_{t-1}\}\leq\sigma^2<\infty \qquad \text{a.s.} \qquad (7.6)$$

$$\sup_N \sum_{t=1}^{N} w(t)^2/N<\infty \qquad \text{a.s.} \qquad (7.7)$$

We assume that $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ do not contain a common factor if $\sigma^2\neq0$ and that $A(q^{-1})$ and $B(q^{-1})$ are coprime if $\sigma^2=0$.

For RLS and ELS type algorithms, when applied to system (7.1), (if the plant model is contained in the identified model set and satisfies a certain positive real condition) a Lyapunov type function can be found, which usually has the form (Ljung 1977, Solo 1979, Goodwin and Sin 1984, Moore 1983, Lai and Wei 1986a)

$$\tilde{\theta}(t)^T[P^{-1}(t-1)/s(t-1)]\tilde{\theta}(t)\to0 \qquad \text{as } t\to\infty \qquad \text{a.s.} \qquad (7.8)$$

where $\tilde{\theta}(t)$ is the parameter estimation error, P(t-1) is the covariance matrix and s(t-1)>0 $\forall$t. Obviously if

$$\liminf_{t\to\infty} \lambda_{min}[P(t-1)^{-1}/s(t-1)]>0 \qquad \text{a.s.} \qquad (7.9)$$

then

$$\lim_{t\to\infty} \tilde{\theta}(t)=0 \qquad \text{a.s.} \qquad (7.10)$$

where $\lambda_{min}(M)$ ($\lambda_{max}(M)$) denotes the minimal (maximal) eigenvalue of a symmetric matrix M.

Many workers (Ljung 1977, Solo 1979, Goodwin and Sin 1984, Moore 1983, Lai and Wei 1986a) have shown that s(t)=t satisfies (7.8). Lai and Wei (1986a) were able to show that {s(t)} can be any positive sequence such that

$$s(t)/\log\{e+\lambda_{max}(\sum_{i=1}^{t} \psi(i)\psi(i)^T)\}\to\infty \quad \text{as } t\to\infty \quad \text{a.s.,} \qquad (7.11)$$

which will lead to weaker persistent excitation type conditions imposed on the plant signals for parameter convergence. In (7.11)

$$\psi(t-1)=[-y(t-1),\ldots,-y(t-p-r1),u(t),\ldots,u(t-l-r2),\quad w(t-1),\ldots,w(t-m-r3)], \tag{7.12}$$

where r1, r2, r3≥0.

7.2.1 Conditions for Parameter Convergence

To guarantee parameter convergence it is necessary to impose conditions on the data, e.g. (7.9) defines conditions imposed on $P(t)^{-1}$.

Translation of conditions imposed on $P(t)^{-1}$ into conditions on the plant signals $\{\psi(t)\}$ was discussed by Solo (1979). He concluded that the ELS (AML) algorithm (subject to certain conditions, e.g. the positive real condition) yields

$$P(t)^{-1}/s(t)-\sum_{i=1}^{t}\psi(i)\psi(i)^{T}/s(t)\rightarrow 0 \quad \text{as } t\rightarrow\infty \quad \text{a.s.} \tag{7.13}$$

for s(t)=t. Similar results have also been presented by Goodwin and Sin (1984), Moore (1983) and Lai and Wei (1986a).

Translation of conditions imposed on the plant signals $\{\psi(t)\}$ into conditions on the plant input u(t) and the noise w(t) was done by Anderson and Johnson (1980) for deterministic systems (where w(t) is not present) and by Moore (1983) and Lai and Wei (1986b) for stochastic systems by a reachability approach. Essentially the conclusion is that if u(t) and w(t) are sufficiently rich in excitation $\sum_{i=1}^{t} \psi(i)\psi(i)^T/s(t)$ has "sufficiently" many eigenvalues greater than a positive number.

Up to this point virtually no minimal model assumption is imposed. For minimal models, since an associated state space realization of $\psi(t)$ with u(t) and w(t) as inputs is completely reachable

$$\liminf_{t\to\infty} \lambda_{min}\left(\sum_{i=1}^{t} \psi(i)\psi(i)^T/s(t)\right)>0. \qquad (7.14)$$

i.e. all the eigenvalues are nonzero, and parameter convergence follows immediately from this property. However, for nonminimal models since the associated state space realization is only partially reachable $\sum_{i=1}^{t} \psi(i)\psi(i)^T/s(t)$ may possess eigenvalues arbitrarily close to zero. In general $\lim_{t\to\infty} \left(\sum_{i=1}^{t} \psi(i)\psi(i)^T/s(t)\right)$ does not exist. These two factors make it difficult to reach any conclusions about the asymptotic properties of the parameter estimates. Therefore one of the main contributions of this chapter is to identify mathematical analysis tools which get around this problem and ensure that the limiting set of the parameter estimates can be determined under conditions such that $\sum_{i=1}^{t} \psi(i)\psi(i)^T/s(t)$ has "sufficiently" many eigenvalues greater than a positive number.

7.3 Parameter Convergence with Overparameterized Models

Theorem 7.1 in this section defines sufficient conditions for parameter convergence of a class of nonminimal systems. The results do not require that the degree of overparameterization be known; do not require the quasi-stationary data assumption used by Ljung; and are applicable to any parameter estimation algorithm and model for which a Lyapunov type function can be found. The general results of Theorem 7.1 are applied to RLS with DARMA models and ELS with ARMAX models in section 7.3.1 and 7.3.2 respectively.

The following set notation is used. If W is a vector space over the real scalar field R, $G \subset W$, $F \subset W$, $w \in W$ and $r \in R$, let

$$w+G \equiv \{w+g:\ g \in G\},$$

$$w-G \equiv \{w-g:\ g \in G\},$$

$$G+F \equiv \{g+f:\ g \in G,\ f \in F\},$$

$$rG \equiv \{rg:\ g \in G\}$$

etc. .

Let

$$\theta^i \in R^N \qquad i=0,1,\ldots,r \leq N \qquad (7.15)$$

be r+1 vectors such that they uniquely determine an r dimensional linear hypersurface H. Specifically if r=1 then H is a linear line, and if r=2 then H is a linear plane, etc.. By this definition it is easy to see that H-h is an r dimensional subspace of R^N for any $h \in H$. Let

$$\tilde{\theta}^i(t) \equiv \hat{\theta}(t) - \theta^i \qquad i=0,1,\ldots,r \qquad (7.16)$$

where $\{\hat{\theta}(t)\}$ is a vector valued time sequence.

<u>Theorem</u> <u>7.1:</u>

$$\liminf_{t\to\infty} \inf_{h\in H\cap\bar{C}} \|\hat{\theta}(t)-h\|=0 \tag{7.17}$$

provided that the following assumptions are verified:

(1) There exists a real symmetric NxN matrix sequence $\{S(t)\}$ such that

(i) $S(t)\geq 0 \qquad \forall t$ (7.18)

(ii) $V_i(t)\equiv\tilde{\theta}^i(t)^T S(t)\tilde{\theta}^i(t)\to 0$ as $t\to\infty$ $i=0,1,\ldots,r$ (7.19)

(iii) $\liminf_{t\to\infty} \lambda_{N-r}(S(t))>0$ (7.20)

where $\lambda_k(M)$ denotes the k th largest eigenvalue of the symmetric matrix M.

(2) $\hat{\theta}(t)\in\bar{C} \qquad \forall t$ (7.21)

where $\bar{C}$ is a compact subset of R^N and

$$\{\theta^0,\theta^1,\ldots,\theta^r\}\subset\bar{C} \tag{7.22}$$

Proof:

Theorem 7.1 is a generalized version of Theorem 6.1 presented in Chapter 6. Herein we give a more direct and concise proof.

Step 1:

Since S(t) is a real symmetric NxN matrix, there exists an orthogonal matrix Q(t) such that

$$S(t)=Q^T(t)\mathrm{diag}(\lambda_1(S(t)),\ldots,\lambda_N(S(t)))Q(t), \tag{7.23}$$

where $\lambda_i(S(t))\geq 0 \ \forall 1\leq i\leq N$ by (7.18) and $\liminf_{t\to\infty} \lambda_{N-r}(S(t))>0$ by (7.20).

Step 2:

Let $x^i(t)\equiv Q(t)\tilde{\theta}^i(t) \qquad i=0,1,\ldots,r.$ (7.24)

By the orthogonality of $Q(t)$, (7.19), (7.23) and (7.24) yield

$$V_i(t)=\sum_{j=1}^{N} \lambda_j(S(t))x^i{}_j(t)^2\to 0 \quad \text{as } t\to\infty \quad i=0,1,\ldots,r \tag{7.25}$$

which implies $\lim_{t\to\infty} \inf_{x\in X}\|x^i(t)-x\|=0 \qquad i=0,1,\ldots,r \qquad (7.26)$

where

$$X\equiv\{x\in R^N: x_j \text{ (the j th element of x) } =0 \ j=1,\ldots,N-r\},$$

since $\lim_{t\to\infty} \inf \lambda_{N-r}(S(t))>0$ and $\lambda_i(S(t))\geq 0$ $i=1,\ldots,N$.

Step 3:

Since $Q(t)$ $\forall t$ is orthogonal, $\{Q(t)\}^\infty{}_{t=1}$ is bounded in the NxN matrix norm space and thus belongs to some compact set. By also considering (7.21) ($\bar{C}$ is compact in R^N) this implies that there exists a subsequence $\{t_k\}^\infty{}_{k=1}$ such that

$$\lim_{k\to\infty} Q(t_k)=\bar{Q} \quad \text{exists,} \tag{7.27}$$

$$\text{and } \lim_{k\to\infty} \hat{\theta}(t_k)=\bar{\theta} \quad \text{exists.} \tag{7.28}$$

Therefore (7.26) implies that

$$\begin{aligned}\lim_{k\to\infty} x^i(t_k)&=\lim_{k\to\infty} Q(t_k)(\hat{\theta}(t_k)-\theta^i)\\ &=\bar{Q}(\bar{\theta}-\theta^i)\in X \qquad i=0,1,\ldots,r\end{aligned} \tag{7.29}$$

since X is a closed subspace, or

$$\lim_{k\to\infty} \hat{\theta}(t_k)=\bar{\theta}\in\theta^i+\bar{Q}^T X \qquad i=0,1,\ldots,r. \tag{7.30}$$

$\bar{Q}^T X$ is an r dimensional subspace since X is and $\bar{Q}$ is orthogonal (and therefore nonsingular). (7.30) implies that the cosets $\theta^i+\bar{Q}^T X$ $i=0,1,\ldots,r$ are the same element in the quotient space R^N modulo $\bar{Q}^T X$ (Rudin 1973). In other words they are the unique r dimensional linear hypersurface in R^N passing through $\theta^0,\theta^1,\ldots,\theta^r$, i.e. they are all identical to H according to the definition of H. So by considering

(7.21), (7.30) gives

$$\lim_{k\to\infty} \hat{\theta}(t_k)=\bar{\theta}\in H\cap\bar{C} \tag{7.31}$$

Step 4:

Similar arguments can be applied to show that every subsequence of $\{\hat{\theta}(t)\}_{t=1}^{\infty}$ has a subsequence converging to a point in $H\cap\bar{C}$. This implies (7.17). It is worth noting that $\bar{Q}^T X$ does not depend on the particular subsequences selected even though $\bar{Q}$ does. ▪

Corollary 7.1:

For a stochastic system, if the assumptions of Theorem 7.1 are all satisfied almost surely (a.s.), then

$$\lim_{t\to\infty} \inf_{h\in H\cap\bar{C}} \|\hat{\theta}(t)-h\|=0 \qquad \text{a.s.} \tag{7.32}$$

Proof:

Applying Theorem 7.1 pointwise to the probability space yields the proof. ▪

7.3.1 RLS Applied to Overparameterized DARMA Models

In this case $C(q^{-1})w(t)\equiv 0$, and the parameter adaptation is (Goodwin and Sin 1984)

$$\hat{\theta}(t)=\hat{\theta}(t-1)+ P(t-2)\phi(t-1)[y(t)-\phi(t-1)^T\hat{\theta}(t-1)] \\ /[1+\phi(t-1)^T P(t-2)\phi(t-1)] \qquad t\geq 1 \tag{7.33}$$

$$P(t-1)=P(t-2)-P(t-2)\phi(t-1)\phi(t-1)^T P(t-2) \\ /[1+\phi(t-1)^T P(t-2)\phi(t-1)] \qquad t\geq 1 \tag{7.34}$$

with $\hat{\theta}(0)$ and $P(-1)>0$ given, and

$$\hat{\theta}(t)=[\hat{a}_1(t),\ldots,\hat{a}_{p+r},\hat{b}_0(t),\ldots,\hat{b}_{1+r}(t)]^T \qquad (7.35)$$

$$\phi(t-1)=[-y(t-1),\ldots,-y(t-p-r),u(t),\ldots,u(t-1-r)]^T \qquad (7.36)$$

where $r\geq 0$ is the degree of overparameterization.

Theorem 7.2:

For the RLS algorithm (7.33)-(7.36) applied to ARMAX model (7.1) with $C(q^{-1})w(t)\equiv 0$, if

$$\lim_{t\to\infty} \lambda_{N-r}(P(t-1)^{-1})=\infty \qquad (7.37)$$

where $N=p+1+2r+1$, then

$$\lim_{t\to\infty} \inf_{h\in H_d\cap\theta+B_e} \|\hat{\theta}(t)-h\|=0, \qquad (7.38)$$

where $B_e\equiv(z\in R^N\text{: } \|z\|\leq e_1)$, (7.39)

$$e_1\equiv\sqrt{\kappa}\,\|\hat{\theta}(0)-\theta\|, \qquad (7.40)$$

$$\kappa\equiv\lambda_1(P(-1)^{-1})/\lambda_N(P(-1)^{-1}), \qquad (7.41)$$

$H_d\subset R^N$ is an r dimensional linear hypersurface such that each element of H_d, h, gives a model with the following form

$$A(q^{-1})E_h(q^{-1})y(t)=B(q^{-1})E_h(q^{-1})u(t) \qquad (7.42)$$

where $E_h(q^{-1})=1+e_{h1}q^{-1}+\ldots+e_{hr}q^{-r}$ (7.43)

and θ is the unique point on H_d which is closest to $\hat{\theta}(0)$.

Proof:

It is easy to verify that H_d is an r dimensional linear hypersurface of R^N and $\hat{\theta}(t)\in\theta+B_e$ $\forall t$ (Chapter 6). Let $H=H_d$, $S(t)=P(t-1)^{-1}/\lambda_{N-r}(P(t-1)^{-1})$, $\bar{C}=\theta+B_e$, $\theta^{\circ}=\theta$ and $\theta^1,\ldots,\theta^r$ be such that $\{\theta^1-\theta,\ldots,\theta^r-\theta\}\subset B_e$ forms a basis of $H_d-\theta$. Condition (7.19) is verified by applying Lemma 3.3.6. of

(Goodwin and Sin) considering that each θ^i $1\le i\le r$ can represent a true plant parameter vector. Conditions (7.18) and (7.20-7.22) are obviously verified. Then application of Theorem 7.1 yields the proof. ∎

7.3.2 ELS Applied to Overparameterized ARMAX Models

In this case $C(q^{-1})w(t)$ is present, and the parameter adaptation has the same form as (7.33) and (7.34) with $\hat{\theta}(0)$ and $P(-1)>0$ given, and the matrices and vectors have their corresponding dimensions, e.g.

$$\hat{\theta}(t)=[\hat{a}_1(t),\ldots,\hat{a}_{p+r},\hat{b}_0(t),\ldots,\hat{b}_{l+r}(t),\\ \hat{c}_1(t),\ldots,\hat{c}_{m+r}(t)]^T \tag{7.44}$$

$$\phi(t-1)=[-y(t-1),\ldots,-y(t-p-r),u(t),\ldots,u(t-l-r),\\ \eta(t-1),\ldots,\eta(t-m-r)]^T \tag{7.45}$$

$$\text{where } \eta(t)\equiv y(t)-\phi(t-1)^T\hat{\theta}(t) \tag{7.46}$$

<u>Theorem 7.3:</u>

For the ELS algorithm (7.33-7.34) and (7.44-7.46) applied to ARMAX model (7.1), ∀ positive k sufficiently large

$$\liminf_{t\to\infty}\ \inf_{h\in H_s\cap B_k} \|\hat{\theta}(t)-h\|=0 \tag{7.47}$$

a.s. on the event $\{\|\hat{\theta}(t)\|\le k,\ \forall t\}$, if there exists some positive sequence $\{s(t)\}$ such that

$$\lim_{t\to\infty} s(t)=\infty \qquad \text{a.s.,} \tag{7.48}$$

and

$$\tilde{\theta}^i(t)^T[P(t-1)^{-1}/s(t-1)]\tilde{\theta}^i(t)\to 0 \qquad \text{as } t\to\infty \qquad \text{a.s.}\\ i=0,1,\ldots,r \tag{7.49}$$

and $\liminf_{t\to\infty} [\lambda_{N-r}(P(t-1)^{-1})/s(t-1)]>0$ a.s., (7.50)

where $N=p+l+m+3r+1$, $B_k\subset R^N$ is as defined in (7.39), H_s is an r dimensional hypersurface in R^N such that each element of H_s, h, gives a model with the following form

$$A(q^{-1})E_h(q^{-1})y(t)=B(q^{-1})E_h(q^{-1})u(t)+C(q^{-1})E_h(q^{-1})w(t) \quad (7.51)$$

where $E_h(q^{-1})$ has the same form of (7.43),

$\{\theta^i,\ i=0,1,\ldots,r\}\subset H_s\cap B_k$ for k sufficiently large and

$\{\theta^i-\theta^0,\ i=1,\ldots,r\}$ forms a basis of $H_s-\theta^0$.

Proof:

Let $H=H_s$, $S(t)=P(t-1)^{-1}/s(t-1)$, $\bar{C}=B_k$. Now all the assumptions of Theorem 7.1 are verified a.s. on the event $\{\|\hat{\theta}(t)\|\le k,\ \forall t\}$. Thus the proof is done by applying Corollary 1. ∎

7.3.3 Remarks

(1) The concept of "quotient space" (Rudin 1973) used in the proof of Theorem 7.1 has a very meaningful geometric interpretation. When transformed from the parameter vector space R^N onto the quotient space the limiting set of parameter estimates, which represents a linear hypersurface in R^N, is condensed to a unique point in the quotient space. Theorems 7.2 and 7.3 both show that the model defined by any point on the limiting set of parameter estimates contains the plant model (7.1) and a monic common factor polynomial.

(2) Condition (7.49) is imposed implicitly on θ^i. For example, if $1/[C(q^{-1})E_{\theta i}(q^{-1})]-1/2$ is very strictly passive, then (7.49) is satisfied for s(t)=t (Solo 1979) or for {s(t)>0} such that (7.11) is satisfied (Lai and Wei 1986a). The existence of the r+1 θ^i's satisfying the very strictly passive condition may only require that $C(q^{-1})$ is stable (Shah and Franklin 1982), which is less demanding than the very strict passivity of $1/C(q^{-1})-1/2$.

(3) The rate at which s(t) approaches infinity in (7.48) affects the type of persistent excitation conditions required on the plant signals for the parameter estimates to converge (to a set in nonminimal model cases). The slower the rate, the weaker the type of persistent excitation conditions will be. This is the basis for Lai and Wei's claim that the type of excitation they used to guarantee parameter converge (to a point in minimal model cases) was weaker than those of previous results.

(4) Conditions (7.37) and (7.50) can be translated into persistent excitation type conditions on the plant input u(t) and noise w(t). For example, for open loop identification if $A(q^{-1})$ is stable a condition similar to the one given by equation (3.13) of (Moore, 1983), i.e. $\liminf_{t\to\infty} \sum_{i=1}^{t} E[V(i)V(i)^T|F_{t-1}]/t>0$ where $V(t)=[u(t),\ldots,u(t-2n-r),w(t-1),\ldots,w(t-2n-r)]^T$, $n=\max(p,l+1,m)$, is sufficient for

$\liminf_{t\to\infty} \lambda_{N-r}(\sum_{i=1}^{t} \psi(i)\psi(i)^T/s(t))>0$ to hold, which in turn implies (7.50). This kind of condition translation has essentially been done in several previous publications (Solo 1979, Anderson and Johnson 1980, Goodwin and Sin 1984, Moore 1983, Lai and Wei 1986b). The basic argument for applying these previous results to the overparameterized case is that, for example, if a state space realization of

$[-y(t-1),\ldots,-y(t-p),u(t),\ldots,u(t-l)]^T$ versus $u(t)$ as input is completely reachable so obviously is that of

$[-y(t-1),\ldots,-y(t-p),u(t),\ldots,u(t-l-r)]^T$ versus $u(t)$, even though that of the full regressor

$[-y(t-1),\ldots,-y(t- \quad r),u(t),\ldots,u(t-l-r)]^T$ versus $u(t)$ is not. A similar argument can be applied to the case where $w(t)$ is present.

In order to avoid unnecessarily lengthy descriptions of previous results (Solo 1979, Anderson and Johnson 1980, Goodwin and Sin 1984, Moore 1983, Lai and Wei 1986b) and to focus on the key contributions of this chapter, persistent excitation type conditions in this chapter are almost always imposed on $P(t)^{-1}$ in terms of equations like (7.37) and (7.50) rather than on $u(t)$ and $w(t)$.

(5) Conclusion (7.47) of Theorem 7.3 applies almost surely on the event $\{\|\hat{\theta}(t)\|\le k, \forall t\}$ for any sufficiently large positive integer k. It would be preferable to extend this result such that the limiting set of $\{\hat{\theta}(t)\}$ belongs to a bounded subset of H_s almost surely on the whole probability

space. However, to do this would require the boundedness of the parameter estimates, which may need some modification to the algorithm, e.g. (i) project $\hat{\theta}(t)$ to a bounded region at each time instant, or (ii) monitor P(t) such that $\sup_t \lambda_{max}(P(t))<\infty$. However, if the algorithm were modified it would be necessary to re-verify the conditions of Theorem 7.1.

7.4 Parameter Convergence with Constrained Overparameterized Models

In this section, it is assumed that some a priori knowledge of the plant is available such that the overparameterized model can be structurally constrained. We first present for this case a counterpart theorem of Theorem 7.1, then apply it to obtain parameter convergence results for the RLS algorithm applied to DARMA models. New approaches to some results of (Goodwin et al. 1985, Janecki 1987, Heymann 1988) are also discussed in this section.

Theorem 7.4:

Subject to assumptions (7.18), (7.20-7.22) of Theorem 7.1, if

$$V_0(t)\equiv\tilde{\theta}^\circ(t)^T S(t)\tilde{\theta}^\circ(t)\to 0 \quad \text{as } t\to\infty \tag{7.52}$$

and

$$(\theta^i-\theta^\circ)^T S(t)(\theta^i-\theta^\circ)\to 0 \quad \text{as } t\to\infty \quad i=1,\ldots r \tag{7.53}$$

then

$$V_i(t)\equiv\tilde{\theta}^i(t)^T S(t)\tilde{\theta}^{\,i}(t)\to 0 \quad \text{as } t\to\infty \quad i=0,1,\ldots,r \tag{7.54}$$

and thus

$$\liminf_{t\to\infty}\ \inf_{h\in H\cap\bar{C}} \|\hat{\theta}(t)-h\|=0, \tag{7.55}$$

if in addition $\{\theta^\circ,\hat{\theta}(t)\}\in D(t)\subset R^N \qquad \forall t \qquad (7.56)$

then obviously

$$\lim_{t\to\infty} \inf_{h\in H\cap\bar{C}\cap D(t)} \|\hat{\theta}(t)-h\|=0 \qquad (7.57)$$

which implies

$$\lim_{t\to\infty} \hat{\theta}(t)=\theta^\circ \qquad (7.58)$$

provided that $\lim_{t\to\infty} \sup_{h_1,h_2\in H\cap\bar{C}\cap D(t)} \|h_1-h_2\|=0. \qquad (7.59)$

For a stochastic system this result applies pointwise with respect to the probability space.

Proof:

Following the same line as the proof of Theorem 7.1 it is easy to show that every subsequence of $\{\hat{\theta}(t)\}$ has a subsequence converging to $\theta^\circ+\bar{Q}^T X\cap\bar{C}$, which always contains $\{\theta^i,i=0,1,...,r\}$. Since $\{\theta^i,i=0,1,...,r\}$ uniquely determines the r dimensional linear hypersurface H we have $\theta^\circ+\bar{Q}^T X\cap\bar{C}=H\cap\bar{C}$ even though $\bar{Q}^T$ varies with different subsequences. Thus the proof is done. ∎

The key point of applying Theorem 7.4 in the following subsections is to decompose the nonminimal model parameter convergence problem into two separate parts. The first part is to verify conditions for insuring (7.55), where the linear hypersurface H is usually used to represent the equivalent class set, each element of which is a true representation of the plant if no constraint is imposed. The second part is to find an explicit or implicit relation for the structured constraint such that the intersection of H and the constrained region can be easily determined.

7.4.1 RLS Applied to Overparameterized DARMA Models with Some Known Parameters

In this case $C(q^{-1})w(t)\equiv 0$. Let H_d be as defined in Theorem 7.2 and $\theta^\circ \in H_d$. Assume that $i_1<i_2<\ldots<i_f$ th elements of θ° are known a priori, where $1<f<N\equiv p+l+2r+1$. The parameter adaptation is

$$\hat{\theta}(t)=\hat{\theta}(t-1)+ P_s(t-2)\phi(t-1)[y(t)-\phi(t-1)^T\hat{\theta}(t-1)] /[1+\phi(t-1)^T P_s(t-2)\phi(t-1)] \qquad t\geq 1 \qquad (7.60)$$

$$P_{sij}(t-1)=P_{ij}(t-1) \qquad \text{if } i,j\neq i_1,\ldots,i_f$$
$$=0 \qquad \text{if } i \text{ or } j=i_1,\ldots,i_f \qquad (7.61)$$

where $P(t-1)$ and $\phi(t-1)$ are given by (7.34) and (7.36), M_{ij} denotes the element of the i th row and j th column of a matrix M. The initial values of the algorithm are

$$P(-1)=P_0>0 \qquad (7.62)$$

and $\hat{\theta}_i(0)=\theta_{i0}=\text{arbitrary} \qquad \text{if } i\neq i_1,\ldots,i_f$
$$=\theta^\circ_i \qquad i=i_1,\ldots,i_f \qquad (7.63)$$

where θ°_i denotes the i th component of θ° etc. . Equivalently (7.60-7.63) can be represented by

$$\hat{\theta}_r(t)=\hat{\theta}_r(t-1)+ P_r(t-2)\phi_r(t-1)[\phi_r(t-1)^T\theta^\circ_r(t-1)-\phi_r(t-1)^T\hat{\theta}_r(t-1)] /[1+\phi_r(t-1)^T P_r(t-2)\phi_r(t-1)] \qquad t\geq 1 \qquad (7.64)$$

where θ°_r is an N-f dimensional vector containing the unknown components of θ° in their original order, $\hat{\theta}_r(t)$ is the estimate of θ°_r and $\phi_r(t-1)$ is the corresponding regressor, etc.,

$$P_r(t-1)=P_r(t-2)- P_r(t-2)\phi_r(t-1)\phi_r(t-1)^T P_r(t-2)/[1+\phi_r(t-1)^T P_r(t-2)\phi_r(t-1)]$$

$$t\geq 1, \tag{7.65}$$

the initial values are properly selected to match those of (7.62) and (7.63).

Theorem 7.5:

For the algorithm (7.60-7.63) or equivalently (7.64-7.65) applied to ARMAX model (7.1) with $C(q^{-1})w(t)\equiv 0$, if

$$\lim_{t\to\infty} \lambda_{N-r}(P(t-1)^{-1})=\infty \tag{7.66}$$

then
$$\liminf_{t\to\infty}\ _{h\in H_d\cap B_k\cap Y} \|\hat{\theta}(t)-h\|=0, \tag{7.67}$$

where $0<k<\infty$ and $Y\equiv\{z\in R^N:\ z_i=\theta_i^o,\ i=i_1,\ldots,i_\ell\}$, (7.68)

in addition if

$$H_d\cap Y=\{\theta^o\} \tag{7.69}$$

then $\lim_{t\to\infty} \hat{\theta}(t)=\theta^o$. (7.70)

Proof:

Let $H=H_d$, $S(t)=P(t-1)^{-1}/\lambda_{N-r}(P(t-1)^{-1})$, $D(t)=Y$ and $\bar{C}=B_k$ for k sufficiently large such that $\hat{\theta}(t)\in\bar{C}\ \forall t$ ($\{\hat{\theta}(t)\}$ is shown to be bounded in the following context) and $\{\theta^o,\theta^1,\ldots,\theta^r\}\subset\bar{C}$, where $\{\theta^1-\theta^o,\ldots,\theta^r-\theta^o\}$ forms a basis of $H-\theta^o$.

Step 1:

An argument similar to that in the proof of Lemma 3.3.6. of (Goodwin and Sin 1984) gives

$$\lim_{t\to\infty}(\hat{\theta}_r(t)-\theta_r^o)^T P_r(t-1)^{-1}(\hat{\theta}_r(t)-\theta_r^o)<\infty \tag{7.71}$$

and $\|\hat{\theta}_r(t)-\theta_r^o\|^2 \le \kappa \|\hat{\theta}_r(0)-\theta_r^o\|^2$ for some $0<\kappa<\infty$. (7.72)

(7.72) implies that $\{\theta(t)\}$ is bounded, i.e. $\{\hat{\theta}(t)\ \forall t\} \subset B_k$ for some $0<k<\infty$.

Step 2:

By inspecting $\phi(i)\sim\phi_r(i)$ and $P(t-1)^{-1}\sim P_r(t-1)^{-1}$ relations

$(P(t-1)^{-1}=P(-1)^{-1}+\sum_{i=0}^{t-1} \phi(i)\phi(i)^T$,

$P_r(t-1)^{-1}=P_r(-1)^{-1}+\sum_{i=0}^{t-1} \phi_r(i)\phi_r(i)^T)$ and noting that the $i_1,\ldots,i_f$ th components of $\tilde{\theta}^o(t)$ are all zeros

$$V_0(t) \equiv \tilde{\theta}^o(t)^T S(t) \tilde{\theta}^o(t) = \tilde{\theta}^o(t)^T [P(t-1)^{-1}/\lambda_{N-r}(P(t-1)^{-1})] \tilde{\theta}^o(t)$$
$$= \tilde{\theta}_r^o(t)^T [P_r(t-1)^{-1}/\lambda_{N-r}(P(t-1)^{-1})] \tilde{\theta}_r^o(t) \to 0 \quad \text{as } t\to\infty \qquad (7.73)$$

from (7.66) and (7.71).

Step 3:

Since $P(t-1)^{-1}/\lambda_{N-r}(P(t-1)^{-1})$

$$=[I+\sum_{i=1}^{t-1} \phi(i)\phi(i)^T]/\lambda_{N-r}(P(t-1)^{-1})$$
$$\to \sum_{i=1}^{t-1} \phi(i)\phi(i)^T/\lambda_{N-r}(P(t-1)^{-1}) \qquad \text{as } t\to\infty \qquad (7.74)$$

and $\phi(t)^T\theta^o=\phi(t)^T\theta^i \qquad i=1,\ldots,r \qquad \forall t \qquad (7.75)$

it follows that

$$(\theta^i-\theta^o)^T S(t)(\theta^i-\theta^o)\to 0 \quad \text{as } t\to\infty \quad i=0,1,\ldots,r. \qquad (7.76)$$

Step 4:

All the conditions of Theorem 7.4 have been verified. Thus applying the theorem yields the proof. ∎

7.4.2 Parameter Convergence to a Unique Point

Theorem 7.4 showed that in general, the parameter estimates converge to the intersection of the hypersurface, H, and the constrained region. Obviously if the intersection is a singleton then the parameter estimates converge to a unique point. An interesting special case is RLS applied to overparameterized DARMA models (cf. section 7.3.1) with $\kappa \equiv \lambda_1(P(-1)^{-1})/\lambda_N(P(-1)^{-1})=1$. In this case parameter convergence to a unique point (which depends on the initial parameter estimate) can be guaranteed no matter what the degree of overparameterization is. This example shows that the constrained region need not be explicitly given in the parameterized structure and can be automatically provided by the properties of some specific identification mechanisms.

Theorem 7.6:

Subject to the same conditions of Theorem 7.2 and $\kappa \equiv \lambda_1(P(-1)^{-1})/\lambda_N(P(-1)^{-1})=1$ we have

$$\hat{\theta}(t) \in G \subset R^N \qquad \forall t \tag{7.77}$$

and
$$\lim_{t\to\infty} \hat{\theta}(t)=\theta \tag{7.78}$$

where θ is the unique point on H_d closest to $\hat{\theta}(0)$ and

$$G \equiv \{g \in R^N:\ g-\theta \perp H_d-\theta \text{ and } \|g-\theta\| \le \|g-\hat{\theta}(0)\|\}. \tag{7.79}$$

Proof:

Refer to Chapter 6 to obtain (77). Then applying Theorem 7.2 and 4 yields the proof since $D(t) \equiv G \cap H_d = \{\theta\}$. ∎

The geometric interpretation of Theorem 7.6 is interesting. When $r \geq 2$, G is a straight line section which is normal to the linear hypersurface H_d with $\hat{\theta}(0)$ at one end and $\theta \in H_d$ as the symmetric point. This symmetric point θ is uniquely defined by the plant model (7.1), the degree of overparameterization r and the initial value of the parameter estimates $\hat{\theta}(0)$. Parameter convergence to θ is guaranteed if $\kappa=1$ even though there are no constraints imposed on the model structure. Indeed the constraint is implicitly provided by the nature of the algorithm which follows the closest path from the initial value to H_d.

7.4.3 Parameter Convergence for Nonminimal Systems Having Purely Deterministic Disturbances

Results similar to those obtained by Goodwin et al. (1985) and Janecki (1985) for parameter convergence of systems having purely deterministic disturbances can be obtained using Theorem 7.4.

Consider the following system

$$A(q^{-1})y_{id}(t)=B(q^{-1})u(t), \tag{7.80}$$

$$y(t)=y_{id}(t)+d_A(t) \tag{7.81}$$

where $d_A(t) \equiv d(t)/A(q^{-1})$, (7.82)

$A(q^{-1})$ and $B(q^{-1})$ are given in (7.2) and (7.3) and assumed to be coprime. d(t) is a periodic signal such that among all the monic polynomials of order$\leq$r there exists a unique $D(q^{-1})$ satisfying

$$D(q^{-1})d(t)=[1+\sum_{i=1}^{r} d_i q^{-i}]d(t)=0 \qquad \forall t. \tag{7.83}$$

Obviously $d_r \neq 0$ by the uniqueness.

Note that an equivalent definition for d(t) is: d(t) is a periodic signal which satisfies the condition for weakly persistent excitation of order r (Definition 3.4.B of Goodwin and Sin, 1984) but not of any order greater than r.

Theorem 7.7:

For the RLS (7.33-7.36) applied to the system (7.80-7.83) if

$$\limsup_{t\to\infty} [\lambda_1(\sum_{i=0}^{t} \phi_{id}(i)\phi_{id}(i)^T)/t]<\infty, \quad (7.84)$$

$$\liminf_{t\to\infty} [\lambda_{N-r}(\sum_{i=0}^{t} \phi_{id}(i)\phi_{id}(i)^T)/t]>0 \quad (7.85)$$

and $\lim_{t\to\infty} (\sum_{i=0}^{t} \phi_{id}(i)\phi_d(i)^T)/t=0$ (7.86)

where

$$\phi_{id}=[-y_{id}(t-1),\ldots,-y_{id}(t-p-r),u(t),\ldots,u(t-1-r)]^T, \quad (7.87)$$

and $\phi_d(t-1)=[-d_A(t-1),\ldots,-d_A(t-p-r),0,\ldots,0]^T$, (7.88)

then $\lim_{t\to\infty} \hat{\theta}(t)=\theta°$, (7.89)

where $\theta°$ yields the following model

$$A(q^{-1})D(q^{-1})y(t)=B(q^{-1})D(q^{-1})u(t). \quad (7.90)$$

Proof:

Step 1:

From (Goodwin and Sin 1984)

$$\lim_{t\to\infty} \tilde{\theta}°(t)^T P(t-1)^{-1}\tilde{\theta}°(t)<\infty \quad (7.91)$$

and $\hat{\theta}(t)\in\theta°+B_k$ $\forall t$ for some $0<k<\infty$. (7.92)

Since $\phi(t)=\phi_{id}(t)+\phi_d(t)$ (7.93)

we have $P(t-1)^{-1}=P(-1)^{-1}+\sum_{i=0}^{t-1} \phi(i)\phi(i)^T$

$$=P(-1)^{-1}+$$

$$\sum_{i=0}^{t-1} [\phi_{id}(i)\phi_{id}(i)^T+\phi_{id}(i)\phi_d(i)^T+\phi_d(i)\phi_{id}(i)^T+\phi_d(i)\phi_d(i)^T] \tag{7.94}$$

From (7.84-7.86), (7.91) and (7.94) we have

$$\tilde{\theta}^\circ(t)^T[\sum_{i=0}^{t-1} \phi_{id}(i)\phi_{id}(i)^T/(t-1)]\tilde{\theta}^\circ(t)\to 0 \quad \text{as } t\to\infty, \tag{7.95}$$

and $\tilde{\theta}^\circ(t)^T[\sum_{i=0}^{t-1} \phi_d(i)\phi_d(i)^T/(t-1)]\tilde{\theta}^\circ(t)\to 0 \quad \text{as } t\to\infty.$ (7.96)

Note that in this example (7.96) defines the constrained region (cf. (7.56) of Theorem 7.4) for the parameter estimates.

Step 2:

By following the same argument as in the proof of Theorem 7.5, (7.92) and (7.95) imply that

$$\lim_{t\to\infty} \inf_{h\in H_d\cap\theta^\circ+B_k} \|\hat{\theta}(t)-h\|=0 \tag{7.97}$$

i.e.

$$\hat{\theta}(t)=h(t)+\Delta h(t), \tag{7.98}$$

where $h(t)\in H_d\cap\theta^\circ+B_k \qquad \forall t$ (7.99)

and $\lim_{t\to\infty} \|\Delta h(t)\|=0,$ (7.100)

H_d is as defined in Theorem 7.2. According to the definition of H_d, $h(\tau)$ gives a model of the following form

$$A(q^{-1})D_h(q^{-1},\tau)y(t)=B(q^{-1})D_h(q^{-1},\tau)u(t) \tag{7.101}$$

where $D_h(q^{-1},\tau)=1+\sum_{i=1}^{r} d_{hi}(\tau)q^{-i}$ (7.102)

is τ dependent.

Step 3:

Let $\theta^{d0}\equiv[d_1,\ldots,d_r]^T,$ (7.103)

$\theta^{dh}(t)\equiv [d_{h1}(t),\ldots,d_{hr}(t)]^T,$ (7.104) and $\delta(t-1)\equiv[d(t-1),\ldots,d(t-r)]^T$. (7.105)

From (7.90) and (7.101) $h(t)-\theta^\circ$ is a linear function of $\theta^{dh}(t)-\theta^{d0}$. By expressing this function in terms of a matrix equation it is not difficult to show that

$$(h(t)-\theta^\circ)^T[\sum_{i=0}^{t-1} \phi_d(i)\phi_d(i)^T/(t-1)](h(t)-\theta^\circ)$$
$$=(\theta^{dh}(t)-\theta^{d0})^T[\sum_{i=0}^{t-1} \delta(i)\delta(i)^T/(t-1)](\theta^{dh}(t)-\theta^{d0}) \quad (7.106)$$

Step 4:

(7.98), (7.100) and (7.84) imply

$$\tilde{\theta}^\circ(t)^T[\sum_{i=0}^{t-1} \phi_{id}(i)\phi_{id}(i)^T/(t-1)]\tilde{\theta}^\circ(t)$$
$$\to(h(t)-\hat{\theta}^\circ(t))^T[\sum_{i=0}^{t-1} \phi_{id}(i)\phi_{id}(i)^T/(t-1)](h(t)-\hat{\theta}^\circ(t)), \quad (7.107)$$

which in turn implies from (7.95) and (7.106)

$$(\theta^{dh}(t)-\theta^{d0})^T[\sum_{i=0}^{t-1} \delta(i)\delta(i)^T/(t-1)](\theta^{dh}(t)-\theta^{d0})\to 0$$
$$\text{as } t\to\infty. \quad (7.108)$$

Since $d(t)$ is weakly persistently exciting of order r

$$\liminf_{t\to\infty} \sum_{i=0}^{t-1} \delta(i)\delta(i)^T/(t-1)>0, \quad (7.109)$$

which yields from (7.108)

$$\lim_{t\to\infty} \theta^{dh}(t)=\theta^{d0}, \quad (7.110)$$

which in turn implies that $\lim_{t\to\infty} \hat{\theta}(t)=\theta^\circ$. (7.111)

Thus the proof is complete. ∎

Condition (7.86) has a much more explicit physical interpretation than conditions 5.15-5.16 of Goodwin et al. (1985) or the conditions of Lemma 2.3 of Janecki (1987): the filtered values of u(t) and d(t) should be uncorrelated and a very special case for (7.86) to hold is that u(t) does not contain those frequencies contained by d(t).

If $d(t)$ is weakly persistently exciting of order$<r$, the limiting set of the parameter estimates can be obtained as well by using Theorem 7.1.

When $A(q^{-1})$ is stable, for conditions (7.84) or (7.85) to hold it is sufficient that $u(t)$ be weakly persistently exciting of order $2n+r$ (Definition 3.4.B of Goodwin and Sin, 1984), where $n \equiv \max(p, l+1)$.

7.4.4 Other Applications of Theorem 7.4

Theorem 7.4 can be used to obtain the parameter convergence properties for RLS applied to the deterministic, structured, nominimal model presented by Heymann (1988). An equivalent algorithm like (7.60-7.63) can be formulated and the constrained region Y defined as the set, each point of which gives a model containing the structurally specified common factor polynomial. Then a result like Theorem 7.5 can be stated and similarly proven. Convergence to a unique point is only a special case, i.e. when $H_d \cap Y$ is a singleton.

Xia et al. (1987) use RLS with an overparameterized model for an adaptive pole placement control algorithm and add an auxiliary signal to the regressor in order to force the parameters to convergence to a unique point. A simpler alternative approach based on the results in this chapter would be to use a constrained, overparameterized model with some prespecified (known) parameters, e.g. $\{b_{n+i}=0,\ i=1,\ldots,r\}$. Theorem 7.5 will guarantee parameter convergence to a unique point if the prespecified parameters uniquely determine the common factor polynomial. Adaptive control schemes can use overparameterized models in this manner to gain the degrees of freedom in the model structure necessary to cope with uncertain disturbances without losing the guarantee of parameter convergence to a unique point if the uncertain disturbances vanish.

Another application for Theorem 7.4 and 7.5 is to prove parameter convergence for the multirate identification algorithm developed in Chapter 3. In this multirate, discrete system, $J>1$ is a positive integer,

$$A(q^{-1})=1+\sum_{i=1}^{nJ} a_i q^{-i},$$

$$B(q^{-1})=\sum_{i=1}^{nJ} b_i q^{-i},$$

and $\{a_i \ i\neq J,2J,\ldots,nJ\}$ are structurally constrained to zero.

7.5 Conclusions

Generalized parameter convergence properties were derived for overparameterized deterministic and stochastic systems with (Theorem 7.4) or without (Theorem 7.1) structured constraints. These results apply to any identification algorithm and model for which an appropriate Lyapunov type function of parameter estimates can be found and do not require the quasi-stationary data assumption nor that the degree of overparameterization be known. Parameters are shown to converge to a set H, each element of which contains the irreducible plant model plus a common factor polynomial, or to the intersection of H and the constrained region. The application of Theorem 7.1 and 7.4 is illustrated by using them to: prove parameter convergence of RLS with overparameterized DARMA models (Theorem 7.2, 7.5 and 7.6) and ELS with ARMAX models (Theorem 7.3); and provide an alternative proof for the results of Goodwin et al. (1985) and Janecki (1987) for nonminimal systems with purely deterministic disturbances (Theorem 7.7). Theorem 7.1 and 7.4 are also shown to be applicable to several other problems in identification and adaptive control (e.g. Heymann 1988, Chapter 3 of this thesis, Xia et al. 1987).

7.6 References

Anderson B. D. O. and Johnson Jr. C. R., "Exponential Convergence of Adaptive Identification and Control Algorithms," Automatica, vol. 18, pp. 1-13, Jan. 1982.

Becker Jr. A. H., Kumar P. R., and Wei C. Z., "Adaptive control with the stochastic approximation algorithm: geometry and convergence," IEEE Trans. Automat. Contr., vol. AC-30, pp. 330-338, Apr. 1985.

Goodwin G. C., Norton J. P., and Viswanathan M. N., "Persistency of excitation for nominimal models of systems having purely deterministic disturbances," IEEE Trans. Automat. Contr., vol. AC-30, pp. 589-590, June 1985.

Goodwin G. C. and Sin K. S., Adaptive Filtering Prediction and Control, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

Heymann M., "Persistency of Excitation Results for Structured Nonminimal Models," IEEE Trans. Automat. Contr., vol. AC-33, pp. 112-116, 1988.

Janecki D., "Remarks on Persistency of Excitation for Nonminimal Models of Systems Having Purely Deterministic Disturbances," IEEE Trans. Automat. Contr., vol. AC-32, pp. 1123-1125, 1987.

Lai T. L. and Wei C. Z., "Extended Least Squares and Their Applications to Adaptive Control and Prediction in Linear Systems," IEEE Trans. Automat. Contr., vol. AC-31, pp. 898-906, 1986.

Lai T. L. and Wei C. Z., "On the Concept of Excitation in Least Squares Identification and Adaptive Control," Stochastics, vol. 16, pp. 227-254, 1986.

Ljung L., "Consistency of Least-Squares Identification Method," IEEE Trans. Automat. Contr., vol. AC-21, 779-781, 1976.

Ljung L., "Analysis of Recursive Stochastic Algorithms," IEEE Trans. Automat. Contr., vol. AC-22, pp. 551-575, 1977.

Ljung L., System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

Moore J. B., "Persistence of Excitation in Extended Least Squares," IEEE Trans. Automat. Contr., vol. AC-28, pp. 60-68, Jan. 1983.

Rudin W., Functional Analysis. McGraw-Hill, New York, 1973.

Shah S. and Franklin G. F., "On Satisfying Strict Positive Real Condition for Convergence by Overparameterization," IEEE Trans. Automat. Contr., vol. AC-27, pp.715-716, June 1982.

Solo V., "The Convergence of AML," IEEE Trans. Automat. Contr., vol. AC-24, pp. 958-962, 1979.

Xia L., Moore J. B. and Gevers M., "On adaptive estimation and pole assignment of overparameterized systems", Int. J. of Adaptive Contr. and Signal Processing, vol. 1, pp. 143-160, 1987.

8. CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

This thesis makes several contributions related to multirate systems and the identification of overparameterized models. The two areas are closely related because overparameterized models are generally used in the formulation of multirate systems. Although the results are mainly theoretical, remarks and/or simulation examples are included to indicate the practicality and area of application of the results. The main contributions are summarized below in the order of presentation in the thesis.

Multirate Parameter and Output Estimation: (cf. Chapters 2 and 3)

A multirate estimation model is formulated from a discretized continuous plant model by two approaches. The obtained estimation model is overparameterized. Both projection and recursive least squares algorithms are developed to identify the parameters of the model and to generate the estimates of $\{y(k), k=1,2,\ldots\}$ using $\{y(kJ), u(k), k=1,2,\ldots\}$. It is proven that the estimated output value, $\hat{y}(k)$, converges to the real output $y(k)$ at each sampling interval although the real output measurement is available only every J sampling intervals. Simulation results show that the least squares algorithm converges much faster than the projection algorithm.

This work appears to be the first in the literature to show the existence of an intersampling output estimation algorithm with asymptotically zero estimation error for processes with unknown parameters.

Adaptive Control of Multirate Systems: (cf. Chapter 4)

The recursive least squares algorithm of Chapter 3 is used to formulate an adaptive control law with input constraints for application to multirate systems. The error between the actual and a reference performance index is shown to be bounded by the product of a finite gain and the parameter estimation error in the limit sense. Sufficient conditions for parameter convergence are proven and thus, under these conditions the performance of the multirate, adaptive constrained control system asymptotically behaves like that of the analogous single (fast) rate, constrained control system with known, constant parameters.

This work illustrates the use of the intersampling output estimation techniques in control applications and a different approach to defining the convergence properties of adaptive controllers.

Multirate Adaptive Inferential Estimation: (cf. Chapter 5)

An adaptive, inferential algorithm for estimation and control of multirate systems is derived. A secondary process output, v(t), sampled at the same rate as the process input, is included in the multirate estimation scheme to produce process output estimates at the same sampling rate as the process input u(t). Compared to previous work (Guilandoust et al. 1987) on multirate inferential systems, the proposed algorithm has a more formal theoretical basis, e.g. y(t) is related to v(t) not simply through an external stochastic disturbance but also through the internal system structure. Thus the working equation of the algorithm can be related more quantitatively to the characteristics of the actual process. Convergence properties are formally proven and a simplified algorithm is proposed for practical applications (but without formal convergence proofs). Simulated results illustrate the convergence properties of the algorithm and the improvement in simple feedback control systems that can be obtained by using estimated values y_e, calculated at every input sampling interval, rather than just the slowly measured output values of y.

This work provides a solid theoretical basis for multirate inferential adaptive estimation and control.

Parameter Convergence with Overparameterized Models: (cf. Chapters 6 and 7)

Through the concept of quotient space, parameter convergence with overparameterized models is defined. Under a set of defined sufficient conditions, which do not restrict system signals to being quasi-stationary as assumed by Ljung (1987), the parameter estimates of a broad class of identification algorithms including RLS and ELS, applied to overparameterized (nonminimal) models, are shown to converge to a set H, each element of which defines a model containing the irreducible plant model plus a common factor polynomial. This result is extended to show that when the overparameterized model is structurally constrained the parameter estimates converge to the intersection of H and the constrained region. The results are shown to be applicable to the multirate identification algorithms presented in Chapters 2 and 3 and provide an alternative approach to obtain some recently published results (Goodwin et al. 1985, Janecki, Heymann 1988) on overparameterization and parameter convergence.

This work represents an essentially complete solution to the problem of parameter convergence with overparameterized or nonminimal models.

8.2 Recommendation for the Future Work

(1) Multirate Parameter and Output Estimation of Stochastic Systems:

The work on multirate parameter and output estimation in this thesis is essentially complete for deterministic systems. However, stochastic systems were not considered in the theoretical analysis. It is therefore desirable to extend the work to include stochastic systems. For example, consider the following system

$$A_1(q^{-1})y(k)=B_1(q^{-1})u(k)+x(k), \tag{8.1}$$

where A_1 and B_1 are given in (3.3), $x(k)$ is a stochastic process and the measured data are $\{y(kJ)\}$ and $\{u(k)\}$, i.e. y is sampled J times slower than u. The objective would be to study the convergence properties of parameter and output estimation algorithms presented in this thesis for this stochastic system and/or develop new algorithms to handle a specific structure of the term x(t), e.g. where x(k) is a moving average of white noise.

(2) Parameter and Output Estimation of Systems with Continuously Measured Inputs and Irregularly Measured Outputs:

The multirate work in this thesis assumes an integer ratio, J, of the two sampling rates and that the input value between two fast sampling instants is constant and known. This is actually a special case of the more general problem involving parameter and output estimation of systems with continuously measured inputs and irregularly measured outputs. The latter can be handled by using continuous models since they are invariant to the output measurement sampling rate. The solution of this problem would result in some fundamental improvements in the adaptive identification and control of a large class of systems, e.g. nonsynchronized sampled systems or two time scale systems.

(3) Multirate Parameter and Output Estimation of MIMO Systems:

The multirate inferential system treated in Chapter 5 is essentially a very special case of MIMO systems, i.e. a system with two outputs and one input. Unlike the intuitive and adhoc solution given by Guilandous et. al (1987), the work in Chapter 5 provides a reliable approach to inferential estimation with slow output measurements, i.e. in this work the inferential schemes are developed structurally from a linear model framework and the associated convergence problem is addressed. Due to the generality of the linear model framework the methodology for deriving the inferential schemes is not limited to the special case treated in Chapter 5. It is also applicable to more general MIMO systems for which there may be several slowly measured outputs and several fast measured inputs and secondary outputs. The work in Chapter 5 would thus be a good starting point to study multirate inferential and control of such MIMO systems.

(4) Implementation of the Multirate Adaptive Control Systems:

The multirate work in this thesis is oriented towards process control applications. Since most of work done so far is theoretical and fundamental it is important in the future to do some experimental and/or extensive simulation work and to address practical implementation issues which may arise in industrial applications. For example, for the adaptive and inferential systems studied in Chapters 4 and 5 it is necessary to develop some adhoc methods to strengen the output feedback and to improve the robustness of the system to unmeasured external disturbances. Such an adhoc method could be a cascade control loop in which an outer loop using the slowly measured output as feedback signal is added such that the outer loop operates with a sampling interval JT and the inner loop with T; or, in terms of Figure 4.1 or Figure 5.1 this would include improvements in the feedback of y(kJ) to the output estimation block so that there is a more direct feedback of the disturbances via y_e.

(5) Parameter Convergence with MIMO Models:

One problem associated with parameter convergence of MIMO systems is determining the existence of a unique point in the parameter vector space to which the parameter estimates converge. In general the uniqueness property does not hold. The quotient space concept used in Chapters 6 and 7 for parameter convergence with overparameterized SISO models is also applicable to MIMO systems. The technique used for the proof of theorems in Chapters 6 and 7 could be applied to determine the limiting set of parameter estimates. Knowing the limiting set of estimates it would then be possible to find ways to structurally constrain the identified model such that an appropriate unique convergence point in the parameter vector space would exist for the identification.

(6) Applications of Overparameterized Models in Adaptive Predictive and Control Application:

Given the proof of parameter convergence for overparameterized models, the use of structurally overparameterized models may be attractive to specific adaptive predictive control applications, since extra design freedom could be gained by using overparameterized models. However it requires further study to see the 'gain' which can be obtained by the use of overparameterized models.

8.3 References

Goodwin G. C., Norton J. P., and Viswanathan M. N., "Persistency of excitation for nominimal modles of systems having purely deterministic disturbance," IEEE Trans. Automat. Contr., vol. AC-30, pp. 589-594, June 1985.

Guilandoust M. T., Morris A. J. and Tham M. T. "Adaptive Estimation in Inferential Process Control," Proc. ACC, pp. 1743-1748, 1987.

Heymann M., "Persistency of Excitation Results for Structured Nonminimal Models," IEEE Trans. Automat. Contr., vol. AC-33, pp. 112-116, 1988.

Janecki D., "Remarks on Persistency of Excitation for Nonminimal Models of Systems Having Purely Deterministic Disturbances," IEEE Trans. Automat. Contr., vol. AC-32, pp. 1123-1125, 1987.

Ljung L., System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

APPENDIX A

DISCRETE PASSIVITY AND POSITIVE REALNESS

The concept of passivity is important in the design and analysis of many adaptive identification and control schemes and is therefore widely discussed in textbooks, e.g. Desoer and Vidyasagar (1975), Landau (1979), Goodwin and Sin (1984), Anderson et al. (1986), Caines (1988) and Astrom and Wittenmark (1989). However, the only terminology used in this thesis that is related to passivity is the very strictly passive (VSP) condition. It is used in Chapters 2 to 5 of this thesis during the development of output error identification methods. This appendix is intended as a brief introduction to the definition, interpretation and application of the VSP condition.

This thesis deals only with proper and rational transfer function systems. For these systems the passivity conditions are equivalent to those of positive realness. These conditions and the equivalence of the two concepts, passivity and positive realness, are discussed below.

A.1 Positive Realness

Consider the following discrete transfer function system in the time domain

$$y(t)=H(q^{-1})u(t) \tag{A.1a}$$

or in the frequency domain

$$y(z)=H(z^{-1})u(z). \tag{A.1b}$$

For convenience in the following discussion let

$$H^*(z)\equiv H(z^{-1}) \tag{A.2}$$

and assume that the transfer function $H^*(z)$ is always proper and rational.

Definition A.1:

The discrete transfer function $H^*(z)$ is positive real (PR) if:

(i) $H^*(z)$ is analytic in $|z|>1$ and the poles of $H^*(z)$ on $|z|=1$ are simple, with nonnegative residues; and

(ii) $Re\{H^*(e^{jw})\}\geq 0$ for all $w\in(-\pi,\pi]$ at which $H^*(e^{jw})$ exists.

Definition A.2:

The discrete transfer function $H^*(z)$ is strictly positive real (SPR) if:

(i) $H^*(z)$ is analytic in $|z|\geq 1$; and

(ii) $Re\{H^*(e^{jw})\}>0$ for all $w\in(-\pi,\pi]$.

Theorem A.1: The following statements are equivalent to each other for system (A.1):

(i) $H^*(z)$ is PR.

(ii) System (A.1) is passive (P).

(iii) The input and output sequences of system (A.1) always satisfy

$$\sum_{i=0}^{t} y(i)u(i)\geq 0 \qquad \forall t\geq 0. \tag{A.3}$$

Theorem A.2: The following statements are equivalent to each other for system (A.1):

(i) $H^*(z)$ is SPR.

(ii) System (A.1) is very strictly passive (VSP).

(iii) There exists a $\delta>0$ such that the input and output sequences of system (A.1) always satisfy

$$\sum_{i=0}^{t} y(i)u(i) \geq \delta \sum_{i=0}^{t} u(i)^2 \qquad \forall t \geq 0. \qquad (A.4)$$

<u>Remarks:</u>

(1) Condition (i) of Definition A.1 which states that $H^*(z)$ is PR implies that $H^*(z)$ is stable but not necessarily asymptotically stable, while Condition (i) of Definition A.2 which states that $H^*(z)$ is SPR is equivalent to asymptotic stability condition for $H^*(z)$.

(2) Condition (ii) of Definition A.1 or Condition (ii) of Definition A.2 is essentially nothing more than a condition on $\phi(w)$ the phase of $H^*(e^{jw})$, i.e. for a PR $H^*(z)$

$$-\pi/2 \leq \phi(w) \leq \pi/2 \qquad w \in (-\pi,\pi] \qquad (A.5)$$

which implies that the Nyquist plot of $H^*(z)$ lies in the closed right complex plane, and for a SPR $H^*(z)$

$$-\pi/2 < \phi(w) < \pi/2 \qquad w \in (-\pi,\pi] \qquad (A.6)$$

which implies that the Nyquist plot of $H^*(z)$ lies in the open right complex plane. To see this note that

$$\phi(w) = \tan^{-1}\{Im(H^*(e^{jw}))/Re\{H^*(e^{jw}))\} \qquad (A.7)$$

and $Re\{H^*(e^{jw})\} \geq 0$ for a PR $H^*(z)$

and $Re\{H^*(e^{jw})\} > 0$ for a SPR $H^*(z)$.

(3) In the textbook by Goodwin and Sin (1984) there are also two other types of strict passivity: input strict passivity (ISP) and output strict passivity (OSP). For proper and rational transfer function systems ISP, OSP and VSP are equivalent.

(4) Theorems A.1 and A.2 follow directly from results in

texts such as Desoer and Vidyasagar (1975), Landau (1979) and Goodwin and Sin (1984). The information in Remarks 1 and 3 can also be obtained from these texts but requires some further analysis and/or rationalization of nomenclature. For example, equation (A.7) is very helpful for one to relate the phase condition of (A.5) or (A.6) and the conditions of PR or SPR but does not appear explicitly anywhere in the above cited texts.

A.2 Choice of D Filter

As discussed in Chapters 2 to 5 the VSP condition (or equivalently the SPR condition, cf. Theorem A.2) is not due to the multirate sampling. It is inherent in the output error identification method (Landau 1979). This condition is not required if an equation error method is used. However the two methods result in different intersampling output estimation error behaviour and it is shown in Chapter 3 that the output error method can have a smaller bound on the error when the input is not rich in excitation and slowly changing.

The choice of the $D(q^{-J})$ filter can play an important role in satisfying the VSP condition associated with the output error method. In many cases a D filter is not required, i.e. $D(q^{-J})=1$. The first example of such cases is where the input u(t) only contains low frequencies since the VSP condition is usually violated only at the high frequencies for most asymptotically stable systems. The second example is where J is large since $A(q^{-J})$ would then be very close to 1 and the VSP condition would therefore always be satisfied. Generally, for asymptotically stable systems it is recommended that $D(q^{-J})=\hat{A}(0,q^{-J})$, where $\hat{A}(0,q^{-J})$ is a good estimate of $A(q^{-J})$. Such an estimate can be obtained by using an equation error method since it does not need the VSP condition.

Other methods of selecting D filters are discussed in the book by Landau (1979).

A.3 REFERENCES

Anderson B. D. O., Bitmead R. R., Johnson Jr. C. R., Kokotovic P. V., Kosut R. L., Mareels M. Y., Praly L. and Riedle B. D., Stability of Adaptive Systems: Passivity and Averaging Analysis. The MIT Press, Cambridge, Ma., 1986.

Astrom K. J. and Wittenmark B., Adaptive Control. Addison-Wesley Publishing Company, 1989.

Caines P. E., Linear Stochastic Systems. John Wiley & Sons, Inc. 1988.

Desoer C. A. and Vidyasagar M., Feedback Systems: Input-Output Propewrties. Academic Press, Inc., London, 1975.

Goodwin, G. C. and Sin K. S. Adaptive Filtering, Prediction and Control, Prentice Hall, New Jersey, 1984.

Landau, I. D., Adaptive Control: the Model Reference Approach, Marcel Dekker, Inc., New York, 1979.