# ON-LINE SYSTEM IDENTIFICATION

# ON-LINE SYSTEM IDENTIFICATION

By

ABHIJIT SEN, B.TECH.,M.S.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

May, 1976

DOCTOR OF PHILOSOPHY (1975)

(Electrical Engineering)

McMaster University
Hamilton, Ontario.

TITLE: On-line System Identification

AUTHOR: Abhijit Sen, B. Tech. (Indian Institute of Technology)
M.S. (University of California)

SUPERVISOR: N.K. Sinha, B.Sc. Eng. (Banaras)
Ph.D. (University of Manchester)

NUMBER OF PAGES: xi, 148.

## ABSTRACT

The problem of finding the characterizing parameters of an unknown linear discrete-time system "on-line" from the measurements of the input and output data is considered in detail. Two new algorithms for system identification have been proposed for the estimation of parameters of time-invariant single-input single-output systems. The first algorithm, called the Generalized Pseudoinverse, is the recursive version of the generalized least squares algorithm. The second algorithm, combining pseudoinverse and stochastic approx. algorithm, is an iterative scheme and found to be computationally more efficient than the first algorithm. The two algorithms have been used in a number of simulation problems to test the reliability and efficiency of the methods. A critical comparison of the new method with the existing algorithms has shown the new algorithm to be reliable in most of the problems considered. Also a new recursive pseudoinverse algorithm has been developed for identification of a multi-variable transfer function model.

## ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. N.K. Sinha for his guidance, encouragement, and assistance throughout the course of this work and in the preparation of this thesis.

He also wishes to thank Dr. E. Della Torre, Dr. J.D. Wright and Dr. J.F. MacGregor for many valuable discussions.

It is the author's pleasure to acknowledge the inspiring discussions with his colleagues. Without being exhaustive, he would like to name S. Abdel-Azim, R. Tang, A. Tom and J. Hickin.

Financial support from McMaster University in the form of scholarship and teaching assistantship is also gratefully acknowledged.

The author especially wishes to express his appreciation for the patient encouragement provided by his wife, Nandini.

The competence and reliability of Carolyn Zoeftig, who typed this thesis, has been much appreciated.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D)

TABLE OF CONTENTS (CONT'D)

## TABLE OF CONTENTS (CONT'D)

LIST OF TABLES

## LIST OF FIGURES

## LIST OF FIGURES (CONT'D)

# CHAPTER 1

# INTRODUCTION

The past two decades have seen tremendous advances in optimal control theory. However one sees very little of it being applied to industrial process control. One of the reasons for this lack of incentive in this direction is that most of the work on optimal control applies only to the deterministic systems for which the mathematical models are known precisely. In reality such systems do not exist. Real processes are almost always characterized by random disturbances, their mathematical models are only approximate and the values of many parameters are not precisely known. In such cases, an effective control requires that the parameters of the system be determined on the basis of observed input output data and the estimates be updated in the light of new observations. This clearly calls for the development of suitable algorithms that are recursive in nature. Moreover, for practical implementation of any identification scheme in real time using mini computers requires parameter estimators which are numerically economical and robust. This problem of recursive identification of linear dynamic systems from the observed input-output data has received considerable attention in recent years.

The lack of a suitable model still remains one of the major obstacles in the implementation of control systems in process industries. Many of the industrial process control systems require simple dynamic models involving only basic process variables. Therefore, the dynamic properties that are of interest for control purposes could probably be described by simple models. Application of system identification can indicate suitable approximations. Moreover theoretical modelling of disturbances is very difficult. Identification is a valuable tool to determine the characteristics of the disturbances.

A number of methods of identification have been proposed in the literature in recent years, but they require considerable computation, in addition to the fact that one has to wait for accumulation of sufficient data to make the computation possible. There is thus a need for a method of identification which allows one to start with an estimate of the parameters of the model even after the first set of data has been obtained, and then keep on improving it as more data arrive. Moreover, a realistic method must take into account the fact that all measurements are contaminated with noise.

The major effort in this thesis is directed towards the problem of system identification under open loop conditions when the observed data are corrupted with disturbances. Attention is restricted to the case where the model of the plant is linear and time-invariant, and the order of the plant is known.

Despite the fundamental importance of closed-loop identification, attention is restricted only to identification of system under open-loop condition. However, in practical situations it may not be possible to open the feedback loop for identification purpose and identification may have to be carried out with the normal operating conditions without disturbing the feedback path.

Chapter 2 discusses the problem associated with the modelling of an unknown process operating in stochastic environment. It is also concerned with the definition and classification of different methods for system identification.

Chapter 3 gives a brief description of different off-line identification methods.

In Chapter 4 the general on-line identification problem is discussed for the single-input single-output case. This problem is formulated by relating the input-output record of the discrete-time system with the unknown parameters in terms of a matrix equation. Several existing on-line identification algorithms are briefly discussed. It is shown that although the pseudoinverse algorithm is computationally simple it gives biased estimates. The instrumental variable method (Wong and Polak, 1967) does not guarantee uniqueness of the estimates. Stochastic approximation algorithms for linear discrete systems (Saridis and Stein, 1968) require a-priori knowledge of the noise variances for on-line unbiased identification. The Generalised least-squares method (Clark, 1967) requires identification of noise statistics and there is a considerable increase in computation time due to iterative nature of the algorithm. For the

of [illegible] ity is not guaranteed.

In the thesis, two new schemes are proposed for the estimation of parameters of single-input single-output system. These schemes essentially try to remove the bias associated with parameter estimates by incorporating a noise model and recursively estimating the parameters of the process model and noise model. The first method called the Generalised Pseudoinverse is a recursive version of the generalised least squares algorithm and is based on the properties of pseudoinverse. The second scheme combining the pseudoinverse and stochastic approximation algorithms results in an iterative procedure in which noise parameters are estimated by the stochastic approximation algorithm and the process parameters are estimated by the pseudoinverse algorithm. The second algorithm is shown to be computationally more efficient than the first algorithm based on the recursive version of generalised least squares algorithm. These algorithms are discussed in detail in this chapter.

Chapter 5 presents the results of identification experiments performed on data obtained from simulated processes as well as from processes under open-loop operation.

Chapter 6 is devoted to the problem of identification of a multivariable system. A new recursive pseudoinverse algorithm is proposed for identification of such a system.

Conclusions and suggestions for future investigation in the problem of identification are discussed in Chapter 7.

# CHAPTER 2

## MODELLING OF INDUSTRIAL PROCESSES

### 2.1. Introduction.

The choice of model structure is one of the basic ingredients in the formulation of the identification problem. This is because the accuracy and the quality of the control processes depend to a large extent on the availability of an adequate model of the dynamics of the controlled plant. Models derived from physical a-priori knowledge often have a wide range of validity. They usually provide a good insight into the behaviour of the process and can be used for designing new plants or redesigning existing ones. However, the governing equations describing the system may not be known or they may be highly complex in nature. Therefore, it is often necessary to take recourse to experimental data so that a dynamic model may be inferred. Three model building steps can be distinguished.

(i) Determination of the form of the model based on physical principles.

(ii) Estimation of the parameters of the model of known structure using experimental data.

(iii) Combination of physical knowledge and estimation techniques.

The physical knowledge of the process can be used to establish the structure of the model and to assign numerical values to some model parameters. Improved estimates of the parameter values are then obtained from experimental data.

As many types of model may be employed, it is preferable to choose a model based on the following properties:-

(i) It must be as simple as possible with minimal number of parameters.

(ii) It must be based on input-output measurements and should not depend on measurements that might be difficult or impossible to make directly.

(iii) The model must adequately represent the dynamic behaviour of the process and should be able to accommodate stochastic behaviour of the system due to random disturbances.

(iv) If possible the proposed model should simplify the subsequent operations, viz. identification, state estimation and controller design.

Several assumptions are also made about the process under consideration during modelling which will reduce the amount of work without seriously limiting the usefulness of the model. The process is known to have the following properties:

(a). Linear: We shall assume that the system is linear or can be adequately approximated by

a linearised model.

(b) Stationary: This implies that the process characteristics are time-invariant.

(c) Stable: It is assumed that the process under consideration is stable. Stability per se is not essential for identifying a system from a theoretical point of view.

(d) Finite Dimensional: The process is assumed to be finite dimensional, i.e. its behaviour is predictable from a finite set of system states and system inputs. Accuracy of any identification scheme depends on the correct value of the system order. However, in the absence of any prior knowledge about the process order, it may become necessary to fit a number of models, of different order, and select the one which provides the best fit as demonstrated by suitable statistical tests. For noise-free cases many methods are available for determining the system order, e.g. one may refer to Lee (1968) and Sinha and Pille (1971) among others. For cases where observations are corrupted by noise, Woodside (1971) and Chow (1972) have proposed a set of tests for a-priori estimation of process order. Box and Jenkins (1970) have given a through treatment of model-order determination.

The above assumptions are based in practice on preliminary experimentation, intuition, a-priori knowledge of the system or engineering judgement. When dealing with control problems, the nature of the model, its structure, its degree of complexity and its accuracy depend upon the type of the control, its goal and specifications, as well as upon the mathematical tools used to build it up. Since most of the theory on control systems is based on the state space or transform representation of the systems, the majority of the mathematical models will be either a set of state equations or transfer functions.

### 2.2.1. State Variable Model.

A finite dimensional linear time invariant discrete-time system can be adequately described by the state variable difference equations:

$$\underset{\sim}{x}(k+1) = A\,\underset{\sim}{x}(k) + B\,\underset{\sim}{u}(k) \qquad (2.1)$$

$$\underset{\sim}{C}(k) = C\,\underset{\sim}{x}(k) \qquad (2.2)$$

where $k \in Z$, the set of integers, and $\underset{\sim}{x}(k) \in E^n$, $\underset{\sim}{u}(k) \in E^p$ and $\underset{\sim}{C}(k) \in E^m$, where $E^n$ denotes the n-dimensional real vector space. The matrices A, B, C are of compatible dimensions. The dimension of $\underset{\sim}{x}$ namely n is called the order of the system.

2.2.2. Transfer Function Model.

The transfer function model of the system can be obtained from the state variable model by taking the z-transform of the variables in equations (2.1) & (2.2) with zero initial conditions. Thus we have:

$$z \underset{\sim}{X}(z) = A \underset{\sim}{X}(z) + B \underset{\sim}{U}(z) \qquad (2.3)$$

$$\underset{\sim}{C}(z) = C \underset{\sim}{X}(z) \qquad (2.4)$$

and hence input and output signals are related by:

$$\underset{\sim}{C}(z) = H(z) \underset{\sim}{U}(z) \qquad (2.5)$$

where

$$H(z) = C(z\,I - A)^{-1}B \qquad (2.6)$$

$$= h_{ij}(z) \quad , \; i = 1,2\ldots m, \; j = 1,2\ldots p \qquad (2.7)$$

H(z) is an mxp matrix of rational functions and is called the Transfer Function matrix of the system. For the work in this thesis the transfer function model is chosen as the working model in view of the fact that it gives a direct relationship between the inputs and the outputs.

2.3.1. Identification Problem.

For the purpose of this thesis the identification problem may be stated as follows:

Given an accurate mathematical structure - the transfer function relating input and output in our case - determine the values of the parameters of the transfer function from the record of input and output measurements over a finite period of time in such a way that the dynamic behaviour of the model accurately describes the dynamic behaviour of the process. In other words, the identification problem deals with the determination of those physical quantities that can not be measured from those that can be measured. So given an accurate mathematical structure the purpose of identification is to estimate the numerical co-efficients associated with such a structure.

2.3.2. Formulation of the Identification Problem.

The common procedure to express the dynamic behaviour of a process in quantitative form is to use some mathematical model for the process. We will assume that the type of equation describing the system is known. Once the decision regarding the topology of the model is made the next step is the evaluation of the parameters of the model viz. parameter estimation. The term identification used in its broadest sense may encompass parameter estimation, but more specifically the former is the determination of the topology of the system, while the latter is the determination of the parameter values assuming the topology to be known.

The problem of parameter estimation can be represented by the block diagram of Figure 2.1 where P is the process with parameters $\underset{\sim}{\phi}$ and M is the model with parameters $\underset{\sim}{\hat{\phi}}$. Since a physical process or a plant is never ideal, an exact mathematical model may not be obtained. A choice must be made such that the model represents or approximates the system in the best manner. This is achieved by minimisation of some cost functional dependent on the error.

In the literature a number of different estimation procedures have been developed. These methods differ predominantly in the criteria used for defining optimality and in the use of available a-priori knowledge. It is evident that the choice between these criteria has aspects that are more or less subjective and the mathematical approach is strongly dependent on these criteria. Furthermore if greater accuracy has to be achieved one cannot disregard the interaction between the physical process and the stochastic environment in which it usually operates. Moreover, any measurements made on the process are not exactly reproducible due to the random effects that are present within and without the process.

Since it is desired to estimate the process parameter vector $\underset{\sim}{\phi}$ the intuitive mathematical formulation of the problem is to minimise the difference between the process parameter vector $\underset{\sim}{\phi}$ and its estimate $\underset{\sim}{\hat{\phi}}$. As $\underset{\sim}{\phi}$ is inaccessible for direct measurements one can only minimize the expectation of this difference if sufficient a-priori knowledge is available.

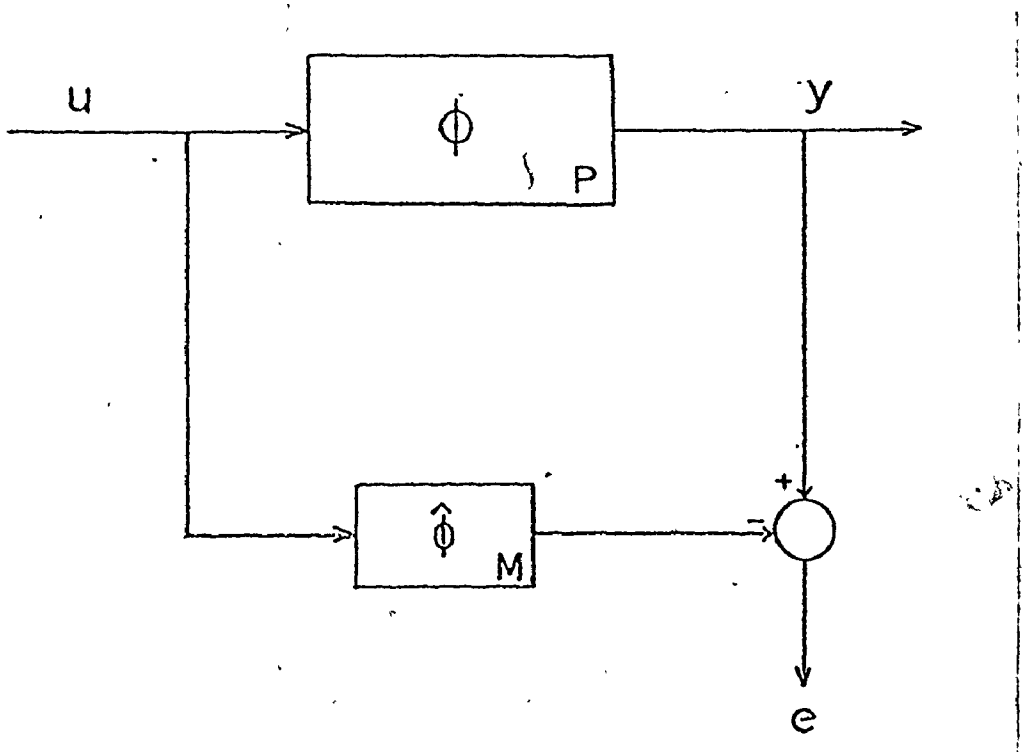



Figure 2.1 Parameter Estimation Problem

Another approach is to minimize some function or functional of

$$e = y - z \tag{2.8}$$

the error between the output of the process and the output of the model. The error e can provide some measure of the correspondence of the parameter vectors. The error e may be used because e can be made measureable and because in some cases the correspondence of input-output relations is more important than parameter correspondence, particularly if the model is simpler than the process.

2.3.3. Classification of Identification Methods.

The identification methods can be divided into two categories, namely off-line methods and on-line methods.

An off-line method is a one-shot approach whereby the solution is available after a fixed and finite number of steps. The processing of data can only start after all the measurements are completed, and estimates are calculated from the complete records of input and output observations.

But it is the characteristic of a real plant that its behaviour often changes gradually in time, therefore, a one-shot model does not yield positive results and so depending on the rate of variation, the model parameters and the structure should be updated continuously or at fixed times in the course of control. This clearly requires a

scheme for parameter estimation which can be implemented on-line. An identification scheme which is recursive in nature and which does not require that the whole string of input and output data be brought in at each step is called an 'on-line' method. The parameter estimates are updated as new measurements are made.

In order that the identification algorithm be truly on-line it should satisfy several requirements. First, it should be computationally efficient. This is because all the necessary calculations have to be completed in updating the estimates before the next set of new measurements are made. Moreover, in an industrial setting, identifier is only part of control loop and the computer is likely to be used for other purposes as well. Thus the ease of implementation and feasibility of realistic application depend heavily on computational simplicity. Secondly, it must not require any special input to be applied to the plant. Thirdly, it should use the samples of the input and output data under actual operating conditions. Fourthly, a realistic on-line method must take into account the fact that all measurements are contaminated with noise.

Finally, it must not require the storage of all the data to make the identification possible.

2.3.4. Problem in Practical Identification.

The term system identification includes three types of activities:

(i) Hypothesize a model based on physical data.

(ii) Parameter Estimation.

(iii) Performance Analysis.

Although the second step dealing with parameter estimation has received the greatest emphasis in the literature, in practice, it is often the case that all of the above mentioned steps are equally important. These steps may have to be repeated several times before the performance analysis shows that the identified model performs satisfactorily for the particular purpose for which it is intended to be used. There are many problems associated with the identification of real processes. The main reasons are:

(i) Limitations of data: Data from real processes are limited often in several senses. The data are not only contaminated by noise but also there may exist missing data or data with erroneous values. For many processes quantities that are important cannot be measured directly and these may cause additional measurement problems.

(ii) Sampling rate: The intended use of the model determines the frequency range of interest. The chosen sampling rate determines the highest sampling frequency that is possible to observe in the sampled signal. Gustavsson (1971) has shown that there exists an optimal sampling

rate with respect to parameter accuracy, but such a precise choice requires the process dynamics to be known. The rule of thumb for the choice of sampling frequency is to take the smallest time constant $T_{min}$ as the sampling interval $T_s$. Goodwin et al (1974) also considered the problem of determining optimal sampling interval in the context of design of experiments, and showed that to "achieve maximal return from an experiment it is essential to carry out a coupled design of pre-sampling filter, sampling intervals, and input signals". MacGregor and Huynh (1974) have considered in detail the effect of choice of sampling interval for process control. In general as the sampling interval is reduced the performance of the control system will improve with corresponding increase in computation time and effort.

(iii) Feedback Structure: Most parameter estimation algorithms suffer severe difficulties in the presence of feedback. This is because, many of the assumptions made in identifying open-loop processes do not hold for systems operating under closed loop. An incorrect model form may

result if standard open-loop procedures for model identification are applied to such cases. For some processes it is often dangerous to carry out experiments under open-loop.

Fisher (1965) has shown that a process under closed loop control may be identified if the feedback is made non-linear. Bohlin (1971) has discussed the implications of closed loop control on the problem of model identifiability and has shown that such a process will be identifiable provided a small perturbation is superimposed at the input. Gustavsson et al (1974) have studied the identification of closed-loop systems and have deduced identifiability conditions for a variety of situations. Ljung et al (1974) have discussed in detail the concepts that are useful for the treatment of system identification under linear feedback control and suggested a method where a series of different feedback laws are used for obtaining parameter estimates. Box and MacGregor (1974, 1975) also discussed the problems associated with parameter estimation under closed-loop operating data and gave necessary and sufficient conditions for estimability using data collected under conditions of optimal control.

(iv) Choice of identification algorithm: Many techniques have been developed for identification of process dynamics as indicated in the survey paper by Astrom and Eykhoff (1971). Isermann et al (1974), Saridis (1974), Sinha and Sen (1975) have presented comparisons of different identification algorithms and evaluated the performance of each method. It is very difficult to give a general answer to the question of what identification method should be used for a particular case. For data with small disturbances any method works well. For highly disturbed data some elaborate method has to be tried to extract useful information

(v) Accuracy of identification: The problem of assigning accuracy to an identification experiment is a difficult one. The reason is that accuracy can be defined in several ways and the identification which may be considered accurate in one sense may be very inaccurate in many other senses. This problem is discussed extensively by Astrom (1967), Herles (1970) and Stepan (1970).

(vi) Identifiability Problems:- If the input sequence does not adequately excite some of the modes or if the model includes unknown dynamics that are fast compared to the sampling rate the associated parameters may not be identifiable. If the model chosen to get the input-output data is inadequate the parameters of that model may be forced to account for some major unmodeled effects. The estimated parameter values may therefore be quite different from the actual values. Finally if there are large unaccounted for instrumental errors or poor initial parameter estimates, non-physical parameter values may result.

All these difficulties have led to considerable research on the subject of system identification. It seems to be difficult to give a general answer to all of the above problems faced in practical identification.

# CHAPTER 3

# TIME SERIES AND MAXIMUM LIKELIHOOD METHOD FOR OFF-LINE IDENTIFICATION.

## 3.1. Introduction.

The term system identification includes three types of activities, (1) hypothesize a model based on physical data, (2) parameters estimation, and (3) performance analysis. The second step dealing with parameter estimation has for the most part received the greatest attention in the literature. However, in practice, all of the above mentioned steps are equally important. These steps in reality are not disjoint, for the results of any one may alter the results of another and, thus, require a repeated application of some or all of the previous steps. Experience has shown that all of the above three steps may have to be repeated several times before the performance analysis shows that the identified model is adequate for the particular purpose for which it is to be used. This chapter presents a discussion of the three steps in identification methodology which have been extensively treated by Box and Jenkins (1970). Also considered is the method of maximum likelihood for estimating the parameters of the model, as it is the most attractive method from the theoretical point of view and is applicable

to the widest range of situations.

3.2. Method of Box and Jenkins.

Box and Jenkins (1970) have proposed to using general linear difference equations to characterize the dynamic relationship between the process input and process output. The model can be represented by:

$$y_t = \frac{a_0+a_1 z^{-1}+\ldots.+ a_s z^{-s}}{1+b_1 z^{-1}+\ldots.+ b_r z^{-r}} u_{t-f} \qquad (3.1)$$

where f represents the number of whole periods of time lag or transport delay between making a change in the input and the beginning of its effect on the output and $z^{-1}$ is the backward shift operator such that $z^{-k}u_t = u_{t-k}$.

However, in practice, the simulation of the process based only on the transfer function model is not adequate because of the presence of noise in the process. The noise which comes from disturbances in the process of one kind or another, can be, in general, described by a time series model. A time series can be conceived as a realization of a white noise process $a_t$ passed through a linear filter. Box and Jenkins introduced the general form of the time series model called Autoregressive-Integrated-Moving Average models of order (p, d, q) defined by:

$$N_t = \frac{1+\theta_1 z^{-1} \ldots + \theta_q z^{-q}}{(1-z^{-1})^d (1+\phi_1 z^{-1} \ldots + \phi_p z^{-p})} a_t \quad (3.2)$$

$$= \frac{\theta_q(z^{-1})}{\nabla^d \phi_p(z^{-1})} a_t \quad (3.3)$$

where $\{a_t, t = 1,2\ldots\}$ is a white noise sequence with mean zero and variance $\sigma_a^2$. The general input denominator term $\nabla^d \phi_p(z^{-1})$ consists of two terms, a backward difference term $\nabla^d = (1 - z^{-1})^d$ containing d roots on the unit circle in the $z^{-1}$ plane to allow the process to be nonstationary, and a stationary autoregressive term $\phi_p(z^{-1})$ containing p roots outside the unit circle to satisfy stationarity conditions. For most stochastic disturbances the various polynomial orders p, d and q will rarely exceed two.

Any complete description of the process must include both a description of the dynamic elements relating input to output and the description of stochastic distrubance affecting the output. The dynamic stochastic model for a process may be written as:

$$y_t = \frac{a_o + a_1 z^{-1} \ldots + a_s z^{-s}}{1+b_1 z^{-1} \ldots + b_r z^{-r}} u_{t-f} + N_t \quad (3.4)$$

$$= \frac{w_s(z^{-1})}{\delta_r(z^{-1})} z^{-f} u_t + N_t \quad (3.5)$$

where

$$N_t = \frac{\theta_q(z^{-1})}{\nabla^d \phi_p(z^{-1})} a_t = \psi(z^{-1})\, a_t \qquad (3.6)$$

$$\frac{w_s(z^{-1})}{\delta_r(z^{-1})} = \frac{a_o + a_1 z^{-1} \ldots + a_s z^{-s}}{1 + b_1 z^{-1} \ldots + b_r z^{-r}} \qquad (3.7)$$

From (3.4)

$$y_t = V(z^{-1})\, u_t + N_t \qquad (3.8)$$

where

$$V(z^{-1}) \triangleq \frac{w_s(z^{-1}) z^{-f}}{\delta_r(z^{-1})} = (V_o + V_1 z^{-1} + V_2 z^{-2} + \ldots.) \qquad (3.9)$$

3.3. Model Building.

Box and Jenkins (1970) have proposed a very extensive body of techniques for building such a dynamic stochastic model of a process from the input-output measurements taken on the process. In brief, they advocate a three-step iterative procedure, viz. identification, estimation and diagnostic checking, to build models characterizing dynamic and stochastic behaviour of the system. This is illustrated in Figure 3.1 .

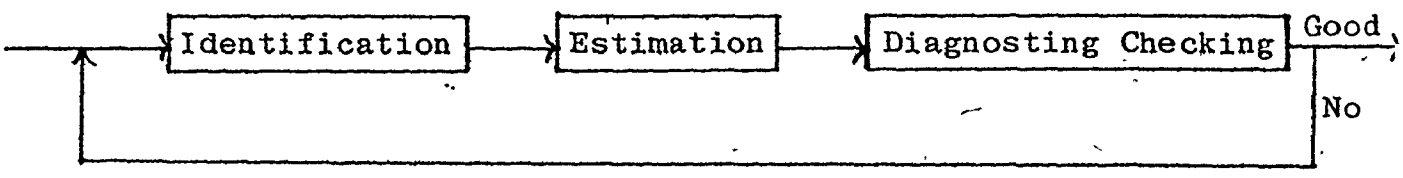



Figure 3.1.Three-Step iterative approach for model building.

3.4. Identification of Transfer Function Model.

Identification is necessary when the form of the transfer function model is unknown. The model for the process may be represented by:

$$y_t = V(z^{-1})\, u_t + N_t \qquad (3.10)$$

where $y_t$ and $u_t$ are outputs and inputs respectively and $N_t$ is the noise term. The basic tool used for identification of the transfer function model is the cross-correlation function between input and output. The autocovariance and the cross-covariance at lag K are defined as:

$$\nu_{uu}(K) \overset{\Delta}{=} E((u_t-\mu_u)\,(u_{t+K}-\mu_u)) \qquad (3.11)$$

$$\nu_{yy}(K) \overset{\Delta}{=} E((y_t-\mu_y)\,(y_{t+K}-\mu_y)) \qquad (3.12)$$

$$\nu_{uy}(K) \overset{\Delta}{=} E((u_t-\mu_u)\,(y_{t+K}-\mu_y)) \qquad (3.13)$$

$$\nu_{yu}(K) \overset{\Delta}{=} E((\ _t-\mu_y)\,(u_{t+K}-\mu_u)) \qquad (3.14)$$

where $\mu_u$ and $\mu_y$ are the mean values of $u_t$ and $y_t$ respectively. Similarly autocorrelation and cross-correlation co-efficients at lag K are defined as:

$$\rho_{uu}(K) \triangleq \frac{\nu_{uu}(K)}{\nu_{uu}(0)} \tag{3.15}$$

$$\rho_{yy}(K) \triangleq \frac{\nu_{yy}(K)}{\nu_{yy}(0)} \tag{3.16}$$

$$\rho_{uy}(K) \triangleq \frac{\nu_{uy}(K)}{\sqrt{\nu_{uu}(0)\nu_{yy}(0)}} \tag{3.17}$$

$$\rho_{yu}(K) \triangleq \frac{\nu_{yu}(K)}{\sqrt{\nu_{uu}(0)\nu_{yy}(0)}} \tag{3.18}$$

Assuming that the input $u_t$ is an uncorrelated white noise series, multiplying equation (3.10) by $u_{t-K}$ and taking expectations yields

$$\nu_{uy}(K) = V_K \nu_{uu}(0) + \nu_{un}(K) \quad \text{for } K \geq 0 \tag{3.19}$$

$$= \nu_{un}(K) \quad \text{for } K < 0. \tag{3.20}$$

Since the input is not correlated with noise

$$\nu_{un}(K) = 0 \tag{3.21}$$

and therefore from (3.19)

$$V_K = \frac{\nu_{uy}(K)}{\nu_{uu}(0)} \quad , \quad K \geq 0. \tag{3.22}$$

Knowing $\nu_{uu}(K)$, $\nu_{uu}(0)$ one can obtain estimates of $V_K$. But in practice one uses the estimates of autocovariance and cross-covariance coefficients as well as of auto-correlation and cross-correlation respectively. The estimates $\hat{\nu}_{uu}(K)$ and $\hat{\nu}_{uy}(K)$ of autocovariance and cross-covariance coefficients respectively are given by:

$$\hat{\nu}_{uu}(K) = \frac{1}{N} \sum_{t=1}^{N-K} (u_t - \bar{u})\,(u_{t+K} - \bar{u}) \tag{3.23}$$

and

$$\hat{\nu}_{uy}(K) = \frac{1}{N} \sum_{t=1}^{N-k} (u_t - \bar{u})\,(y_{t+K} - \bar{y}) \tag{3.24}$$

where

$$\bar{u} = \frac{1}{N} \sum_{t=1}^{N} u_t \tag{3.25}$$

$$\bar{y} = \frac{1}{N} \sum_{t=1}^{N} y_t \tag{3.26}$$

Substituting these estimates in (3.15) and (3.17) $\hat{\rho}_{uu}(K)$ and $\hat{\rho}_{uy}(K)$ of autocorrelation and cross-correlation co-efficients at lag K can be evaluated:

$$\hat{\rho}_{uu}(K) = \frac{\hat{\nu}_{uu}(K)}{\nu_{uu}(0)} \tag{3.27}$$

and

$$\hat{\rho}_{uy}(K) = \frac{\hat{\nu}_{uy}(K)}{\sqrt{\nu_{uu}(0)\nu_{yy}(0)}} \tag{3.28}$$

Thus from (3.22) the estimate $\hat{V}_K$ of $V_K$ is given by:

$$\hat{V}_K = \frac{\hat{\nu}_{uy}(K)}{\hat{\nu}_{uu}(0)} = \hat{\rho}_{uy}(K)\sqrt{\frac{\hat{\nu}_{yy}(0)}{\hat{\nu}_{uu}(0)}}, \quad K \geq 0 \qquad (3.29)$$

Using these estimates of $\hat{V}_K$ the orders of r, s and f will be identified and the preliminary estimates of the parameters $a_i$ and $b_i$ will be made.

## 3.5. Identification of the Noise Model.

Given the preliminary estimate $\hat{V}(z^{-1})$ of the transfer function $V(z^{-1})$ as discussed in the last section, then an estimate of the noise series is provided by:

$$\hat{N}_t = y_t - \hat{V}(z^{-1})\, u_t \qquad (3.30)$$

The study of the estimated auto-correlation function of $\hat{N}_t$ will lead to the identification of the noise model similar to the case of identification of transfer function model.

Once the tentative form of the transfer function of the process and the noise model is identified, the next step is the estimation of the parameters of the process and noise model.

## 3.6. Estimation.

In the previous section, the model forms were identified and rough estimates of the parameters were obtained. Since

rough estimates do not allow for checking for the inadequacy of the model, parameters should be estimated by more efficient methods. This is done by minimising the conditional sum of squares:

$$S(\underset{\sim}{\beta}) = \sum_{t=1}^{n} a_t^2 (\underset{\sim}{\beta}) \qquad (3.31)$$

where $\underset{\sim}{\beta}$ denotes the parameters to be estimated. $a_t$ can be computed recursively from the appropriate model, for any given choice of parameters and starting values.

3.7. Diagnostic Checking.

Diagnostic checking of the fitted models using residuals indicate not only whether the representation of the data provided by the fitted model is adequate, but if it is inadequate, how the original model form should be modified.

Let the overall model be given by:

$$y_t = V(z^{-1})\, u_t + \psi(z^{-1})\, a_t \qquad (3.32)$$

Suppose $a_{ot}$ be the residuals obtained from the incorrect model

$$y_t = V_o(z^{-1})\, u_t + \psi_o(z^{-1})\, a_{ot} \qquad (3.33)$$

Then:

$$a_{ot} = \psi_o^{-1}(z^{-1}) \, (V(z^{-1}) - V_o(z^{-1}) \, ) \, u_t + \psi_o^{-1}(z^{-1}) \, \psi(z^{-1}) a_t \tag{3.39}$$

It is apparent that $a_{ot}$ will be cross-correlated with $u_t$ when the identified transfer function model $V_o(z^{-1})$ is different from the true model $V(z^{-1})$. Furthermore, if the transfer function is correct $a_{ot}$ is given by:

$$a_{ot} = \psi_o^{-1}(z^{-1}) \, \psi(z^{-1}) \, a_t \tag{3.40}$$

It is easy to see if $\psi_o(z^{-1})$ is different from $\psi(z^{-1})$, $a_{ot}$ will not be white. Thus, one can check for model adequacy by testing for the presence of significant sample cross-correlations $\gamma_{uao}(K)$ and autocorrelations $\nu_{a_o a_o}(K)$. Finally, an overall check can be performed on the autocorrelation and cross-correlations of the residuals.

## 3.8. Maximum-Likelihood Algorithm for Parameter Estimation.

The algorithm discussed here was originally presented by Åström and Bohlin (1965). The linear system treated here is described by the following equation:

$$B(z^{-1}) \, y(t) = A(z^{-1}) \, u(t) + \lambda \, C(z^{-1}) \, e(t) \tag{3.41}$$

where {u(t)} is the input, {y(t)} the output and {e(t)} is a sequence of independent normal (0,1) random variables. $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ are polynomials:

$$A(z^{-1}) = a_o + a_1 z^{-1} .... + a_n z^{-n} \qquad (3.42)$$

$$B(z^{-1}) = 1 + b_1 z^{-1} .... + b_n z^{-n} \qquad (3.43)$$

$$C(z^{-1}) = 1 + c_1 z^{-1} .... + c_n z^{-n}. \qquad (3.44)$$

The system model (3.41) contains 3n + 1 parameters $a_o, a_1 \ldots a_n, b_1 \ldots b_n$ and $c_1, c_2 \ldots c_n$. So our problem is to estimate the parameters of the model from the input {u(t), t=1,2,...N} and observation of the output {y(t), t=1,2,...N}. Let $p(\{y(t)\} \mid \{u(t)\}, a,b,c,\lambda)$ be the probability density function of the outputs {y(t)}. given the inputs {u(t)} and the parameters $a,b,c,\lambda$, where we introduce the row vectors a,b,c whose components are $a_i, b_i$ and $c_i$, respectively.

Let us define the function $\varepsilon(t)$ by:

$$\begin{aligned} C(z^{-1})\ \varepsilon(t) &= B(z^{-1})\ y(t) - A(z^{-1})\ u(t) \\ &= \lambda\ C(z^{-1})\ e(t) \end{aligned} \qquad (3.45)$$

It follows that $\varepsilon(t)$ defined in (3.45) is an independent and normal process $(0,\lambda)$. The logarithm of the probability density function of $\{\varepsilon(t)\}$ now becomes:

$$L(\theta,\lambda) = -\frac{1}{2\lambda^2}\sum_{t=1}^{N}\varepsilon^2(t) - N\log\lambda + \text{const.} \quad (3.46)$$

Let

$$\theta = \{a_o \ldots a_n,\ b_1 \ldots b_n\ c_1 \ldots c_n\}$$

Maximising $L(\theta,\lambda)$ implies:

$$\frac{\partial L(\theta,\lambda)}{\partial\lambda^2} = 0 \Rightarrow \lambda^2 = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t). \quad (3.47)$$

$$\text{and } \max_{\theta} L(\theta\ \lambda) \Rightarrow \min \frac{1}{2}\sum_{t=1}^{N}\varepsilon^2(t) = \min_{\theta} J(\theta) \quad (3.48)$$

3.9. Numerical Algorithm.

To minimize $J(\theta)$ the following algorithm is used:

$$\theta^{k+1} = \theta^k - (J_{\theta\theta}(\theta^k))^{-1} J_\theta(\theta^k) \quad (3.49)$$

where $J_\theta$ denotes the gradient and $J_{\theta\theta}$ the matrix of second partial derivatives of $J(\theta)$. The partial derivatives are given by:

$$\frac{\partial J}{\partial\theta_i} = \sum_{t=1}^{N}\varepsilon(t)\frac{\partial\varepsilon(t)}{\partial\theta_i} \quad (3.50)$$

$$\frac{\partial^2 J}{\partial\theta_i\partial\theta_j} = \sum_{t=1}^{N}\frac{\partial\varepsilon(t)}{\partial\theta_i}\frac{\partial\varepsilon(t)}{\partial\theta_j} + \sum_{t=1}^{N}\varepsilon(t)\frac{\partial^2\varepsilon(t)}{\partial\theta_i\partial\theta_j} \quad (3.51)$$

The derivatives of $\varepsilon(t)$ are given by:

$$C(z^{-1}) \frac{\partial \varepsilon(t)}{\partial a_i} = -z^{-i} u(t) \qquad (3.52)$$

$$C(z^{-1}) \frac{\partial \varepsilon(t)}{\partial b_i} = z^{-i} y(t) \qquad (3.53)$$

$$C(z^{-1}) \frac{\partial \varepsilon(t)}{\partial c_i} = -z^{-i} \varepsilon(t) \qquad (3.54)$$

$$C(z^{-1}) \frac{\partial^2 \varepsilon(t)}{\partial a_i \partial c_j} = -z^{-(i+j)} \frac{\partial \varepsilon(t)}{\partial a_1} \qquad (3.55)$$

$$C(z^{-1}) \frac{\partial^2 \varepsilon(t)}{\partial b_i \partial c_j} = -z^{-(i+j)} \frac{\partial \varepsilon(t)}{\partial b_1} \qquad (3.56)$$

$$C(z^{-1}) \frac{\partial^2 \varepsilon(t)}{\partial c_i \partial c_j} = -2z^{-(i+j)} \frac{\partial \varepsilon(t)}{\partial c_i} \qquad (3.57)$$

The algorithm for likelihood function is thus:

(1) Put $\theta^k = \theta^o$ starting value of $\theta$.

(2) Evaluate $J_\theta(\theta^k)$ and $J_{\theta\theta}(\theta^k)$ using (3.50) and (3.51).

(3) Calculate $\theta^{k+1}$ from (3.49) and repeat 2.

It is shown by Åström and Bohlin (1965) that the maximum likelihood estimates are consistent, asymptotically normal and efficient for increasing sample size. But on the other hand it is extremely slow, time consuming and difficult to implement.

# CHAPTER 4

# ON-LINE IDENTIFICATION METHODS

## 4.1. Introduction.

Once a suitable model for the process has been postulated, the next step is to estimate the parameters associated with this model. In this chapter we will consider a single-input, single-output discrete-time system which is characterized by its pulse transfer function of the form:

$$H(z) = \frac{a_o + a_1 z^{-1} + \ldots + a_m z^{-m}}{1 + b_1 z^{-1} \ldots + b_n z^{-n}} \qquad (4.1)$$

where m and n are known. The identification problem to be treated here consists of determining the coefficients $\{a_o, a_1 \ldots a_m, b_1, b_2 \ldots b_n\}$ from the equation (4.1) on the basis of input and output records taken over some arbitrary time interval.

In this chapter attention is restricted to the estimation of the process parameters. Various methods which have been proposed in the literature for on-line identification of discrete-time systems are discussed. Two new algorithms are proposed which are particularly well-suited for handling the inevitable measurement noise inherent in most practical situations.

4.2. Model for Identification.

The linear discrete-time single-input, single-output system characterized by the pulse transfer function H(z) given by equation (4.1) can be equivalently expressed in the following difference equation form:

$$c_i = \sum_{j=0}^{m} a_j r_{i-j} - \sum_{j=1}^{n} b_j c_{i-j} \qquad (4.2)$$

where

$c_i$ = c(iT) = Output of the system at t = iT

$r_i$ = r(iT) = Input of the system at t = iT.

T = Sampling period

i = An integer

Letting i range from 1 to some integer k equation (4.2) can be concatenated in the form of the matrix equation:

$$A_k' \underset{\sim}{\phi} = \underset{\sim}{c}_k \qquad (4.3)$$

where

$$A_k' = \begin{bmatrix} r_1 & r_o & \cdots & r_{1-m} & -c_o - c_{-1} & \cdots & -c_{1-n} \\ r_2 & r_1 & \cdots & r_{2-m} & -c_1 - c_o & \cdots & -c_{2-n} \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ r_k & r_{k-1} & \cdots & r_{k-m} & -c_{k-1} & -c_{k-2} \cdots & -c_{k-n} \end{bmatrix} \qquad (4.4)$$

$$c_k^T = (c_1\ c_2\ \ldots\ c_k) \tag{4.5}$$

$$\underset{\sim}{\phi} = (a_o a_1\ \ldots\ a_m\ \ b_1\ b_2\ \ldots\ b_n)^T \tag{4.6}$$

and the superscript T represents transposition.

Equation (4.3) can be solved analytically if m+n+1 pairs of $r_i$ and $c_i$ are known exactly.

However, the equation (4.3) does not supply details of any unmeasurable disturbances or measurement noise that may affect the process. It is assumed that the observed output is given by:

$$y_i = c_i + n_i \tag{4.7}$$

where $n_i$ represents the combined effect of all unmeasureable disturbances and measurement noise affecting the process output $c_i$. The overall signal topology of the system as described by equations (4.2) and (4.7) is shown in Figure 4.1. From equations (4.3) and (4.7) we have:

$$\underset{\sim}{y}_k = A_k \underset{\sim}{\phi} + \underset{\sim}{e}_k \tag{4.8}$$

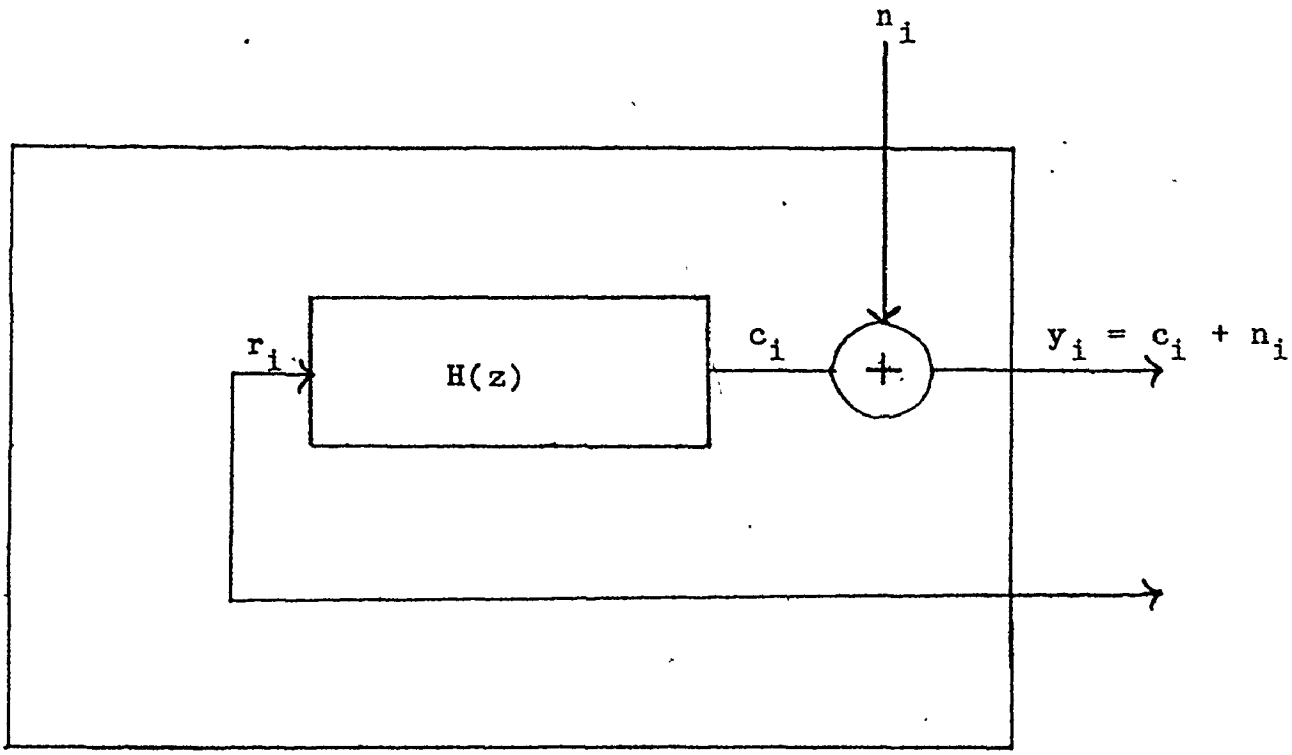



Figure 4.1: Signal topology of a general single-input, single-output process.

where

$$A_k = \begin{bmatrix} r_1 & r_o & \cdots & r_{1-m} & -y_0 & -y_{-1} & \cdots & -y_{1-n} \\ r_2 & r_1 & \cdots & r_{2-m} & -y_1 & -y_0 & \cdots & -y_{2-n} \\ \cdot & & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & & \cdot \\ r_k & r_{k-1} & \cdots & r_{k-m} & -y_{k-1} & -y_{k-2} & \cdots & -y_{k-n} \end{bmatrix} \tag{4.9}$$

$$\underset{\sim}{y}_k^T = (y_1 \; y_2 \; \cdots \; y_k) \tag{4.10}$$

$$\underset{\sim}{e}_k^T = (e_1 \; e_2 \; \cdots \; e_k) \tag{4.11}$$

and

$$e_k = n_k + \sum_{j=1}^{n} b_j \, n_{k-j} \tag{4.12}$$

Our objective is to derive an estimate $\hat{\underset{\sim}{\phi}}$ of $\underset{\sim}{\phi}$ from the sequence of observations $r_k$ and $y_k$ such that the resulting estimated model adequately describes the dynamic characteristics of the process.

4.3. Remarks.

1. In general, $k > p$, i.e. the number of measurements exceeds the number of parameters $p = m+n+1$ to be estimated.

2. For a single equation or one of the rows in equation (4.8):

$$y_k = \sum_{j=0}^{m} a_j \, r_{k-j} - \sum_{j=1}^{n} b_j \, y_{k-j} + e_k \tag{4.13}$$

the independent variables $r_k$, $r_{k-1} \cdots r_{k-m}$, $y_{k-1}$ $y_{k-2} \cdots y_{k-n}$ are referred as the 'regressors' and the dependent variable $y_k$ on the left hand side, is referred as the 'regressand'.

3. In the vector matrix model (4.8) the following terminology is often used.

$\underset{\sim}{y}_k$ = Observation vector

$A_k$ = Information matrix

$\underset{\sim}{\phi}$ = Parameter vector

$\underset{\sim}{e}_k$ = Disturbance vector.

4. The following partitioning of the information matrix $A_k$ into its rows is convienient.

$$A_k = \begin{bmatrix} \underset{\sim}{a}_1^T \\ \underset{\sim}{a}_2^T \\ \vdots \\ \underset{\sim}{a}_k^T \end{bmatrix} \tag{4.14}$$

where $\underset{\sim}{a}_k^T = (r_k \; r_{k-1} \cdots r_{k-m} \; -y_{k-1} \; -y_{k-2} \cdots -y_{k-n})$ (4.15)

so that

$$y_k = \underset{\sim}{a}_k^T \; \underset{\sim}{\phi} + e_k \tag{4.16}$$

Needless to say, the model (4.8) and its variations presented above will be frequently used.

4.3.1. Least-Squares Estimation of Process Parameters using Pseudoinverse Algorithm.

To find the least-squares estimator $\hat{\underset{\sim}{\phi}}_k$ of the parameters $\underset{\sim}{\phi}$ an error vector $\varepsilon$ may be defined as:

$$\underset{\sim}{\varepsilon} = \underset{\sim}{y}_k - A_k \hat{\underset{\sim}{\phi}}_k \tag{4.17}$$

The least squares estimators imply a minimisation of:

$$E = \underset{\sim}{\varepsilon}^T \underset{\sim}{\varepsilon} = ||\underset{\sim}{\varepsilon}||^2$$

$$= ||\underset{\sim}{y}_k - A_k \hat{\underset{\sim}{\phi}}_k||^2.$$

$$= \underset{\sim}{y}_k^T \underset{\sim}{y}_k + \hat{\underset{\sim}{\phi}}_k^T A_k^T A_k \hat{\underset{\sim}{\phi}}_k - 2 \hat{\underset{\sim}{\phi}}_k^T A_k^T \underset{\sim}{y}_k \tag{4.18}$$

Differentiation of equation (4.18) with respect to $\hat{\underset{\sim}{\phi}}$ gives:

$$\frac{\partial E}{\partial \underset{\sim}{\phi}_k} = \nabla_{\hat{\underset{\sim}{\phi}}_k} E = 2 A_k^T A_k \hat{\underset{\sim}{\phi}}_k - 2 A_k^T \underset{\sim}{y}_k \tag{4.19}$$

For the optimal model parameter

$$\nabla_{\hat{\underset{\sim}{\phi}}_k} E = 0 \tag{4.20}$$

This implies:

$$\hat{\underset{\sim}{\phi}}_k = (A_k^T A_k)^{-1} A_k^T \underset{\sim}{y}_k \tag{4.21}$$

or

$$\hat{\phi}_k = A_k^+ \underset{\sim}{y}_k \tag{4.22}$$

where $A_k^+$ is the Pseudoinverse of $A_k$ defined as:

$$A_k^+ = (A_k^T A_k)^{-1} A_k^T \quad \text{for } k > p \tag{4.23}$$

The properties of the pseudoinverse are discussed in Appendix I. The pseudoinverse defined by equation (4.23) exists and is unique if the matrix $A_k$ possesses full rank, i.e. it is of rank p. This condition is assured if the input sequence satisfies any of the following set of conditions (Liff, 1966).

1. $r_k$ is a random sequence
2. $r_k = 0$ for $k < m$
   $= 1$ for $k > m$
3. $r_k$ is a sequence composed of m+1 discrete Fourier components and all natural modes are present in the output sequence $c_i$.

Model (4.8), which is the basis of the least-squares estimator and which will dictate its statistical properties, has the following characteristics:

1. It will be assumed that:

$$\underset{k\to\infty}{\text{p.lim}} \left(\frac{1}{k} A_k^T A_k\right) \tag{4.24}$$

exists and is nonsingular. This condition assures

that the measured signals contain sufficient information to make the estimation possible.

2. The asymptotic property of the estimator can be found by taking the probability limit of (4.22).

$$\text{p.}\lim_{k\to\infty} \hat{\underset{\sim}{\phi}}_k = \text{p.}\lim_{k\to\infty} A_k^{+} \underset{\sim}{y}_k$$

$$= \text{p.}\lim_{k\to\infty} \{(A_k^T A_k)^{-1} A_k^T \underset{\sim}{y}_k\}$$

$$= \text{p.}\lim_{k\to\infty} \{(A_k^T A_k)^{-1} A_k^T (A_k \underset{\sim}{\phi} + e_k)\}$$

$$= \text{p.}\lim_{k\to\infty} \{\underset{\sim}{\phi} + (A_k^T A_k)^{-1} A_k^T \underset{\sim}{e}_k\}$$

$$= \underset{\sim}{\phi} + \text{p.}\lim_{k\to\infty} (A_k^T A_k)^{-1} \cdot \text{p.}\lim (A_k^T \underset{\sim}{e}_k)$$

$$= \underset{\sim}{\phi} + \text{p.}\lim_{k\to\infty} (\tfrac{1}{k} A_k^T A_k)^{-1} \text{p.}\lim_{k\to\infty} (\frac{A_k^T \underset{\sim}{e}_k}{k}) \quad (4.25)$$

In equation (4.25) the first limit on the right hand side exists by assumption (4.24) and the second limit does not vanish since $e_k$ at any given sampling instant k is correlated with the regressors: $y_{k-1}\ y_{k-2} \cdots y_{k-n}$ This confirms that:

$$\text{p.}\lim_{k\to\infty} \hat{\underset{\sim}{\phi}}_k \neq 0 \quad (4.26)$$

In other words the estimators will yield inconsistent results,

4.3.2. Recursive Pseudoinverse Algorithm for Least Squares Parameter Estimation:

The computation of least squares estimation of $\hat{\underset{\sim}{\phi}}_k$ as given by equation (4.22) requires computation of pseudo-inverse of $A_k^+$ for each estimate $\hat{\phi}_k$. This requires matrix multiplication and inversion of a matrix of order $p = m+n+1$ for $k > p$. Clearly if the parameter estimator is on-line, matrix inversion is undesirable, particularly for a system of higher order. Indeed on-line estimation may not be possible as k increases, since the number of operations required to form the estimate $\hat{\underset{\sim}{\phi}}_k$ becomes prohibitively large. Moreover, another additional disadvantage of using the equation (4.22) is that all the past data in the interval over which the estimates are made must be stored. These problems suggest the necessity of a recursive scheme where it would be possible to update the estimate $\hat{\underset{\sim}{\phi}}_k$ in real time at each stage k based upon $\hat{\underset{\sim}{\phi}}_{k-1}$ and the kth data pair of the input and output.

Greville (1959, 1960) developed a recursive algorithm for the pseudoinverse by considering a matrix A and the addition of a single column to A. The problem of additional rows, which is the effect of an additional set of data, was considered by Wells (1967), Albert and Sittler (1966), and Sinha and Pille (1971). The information matrix $A_{k+1}$ is formed in the following manner:

$$A_{k+1} \triangleq \begin{bmatrix} A_k \\ \underset{\sim}{a}_{k+1}^T \end{bmatrix} \tag{4.27}$$

where

$$\underset{\sim}{a}_{k+1}^{T} = (r_{k+1}\ r_k \cdots r_{k+1-m} - y_k - y_{k-1} \cdots - y_{k+1-n}) \tag{4.28}$$

is a row vector containing the latest set of measurements. Similarly the output vector $\underset{\sim}{y}_{k+1}$ is of the form:

$$\underset{\sim}{y}_{k+1} = \begin{bmatrix} \underset{\sim}{y}_k \\ y_{k+1} \end{bmatrix} \tag{4.29}$$

where $y_{k+1}$ is the latest measurement of the output of the system at (k+1)th instant corresponding to the input $r_{k+1}$ The result is that, when a new pair of input-output measurements is made, a new row is added to the information matrix $A_k$ and a new element is added to the output vector $\underset{\sim}{y}_k$. With these arrangements we are now in a position to derive the iterative algorithm (Sinha and Pille 1971). The main results for the kth variation are summarized below:

For k> m+n+1

$$\underset{\sim}{\hat{\phi}}_{k+1} = \underset{\sim}{\hat{\phi}}_k + \frac{P_k\ \underset{\sim}{a}_{k+1}\ (y_{k+1} - \underset{\sim}{a}_{k+1}^{T}\ \underset{\sim}{\hat{\phi}}_k)}{1 + \underset{\sim}{a}_{k+1}^{T}\ P_k\ \underset{\sim}{a}_{k+1}} \tag{4.30}$$

$$P_{k+1} = P_k - \frac{P_k\ \underset{\sim}{a}_{k+1}\ (P_k\ \underset{\sim}{a}_{k+1})^{T}}{1 + \underset{\sim}{a}_{k+1}^{T}\ P_k\ \underset{\sim}{a}_{k+1}} \tag{4.31}$$

The values of $P_{m+n+1}$ and $\hat{\underset{\sim}{\phi}}_{m+n+1}$ are determined from the square submatrix $A_{m+n+1}$ as:

$$\hat{\underset{\sim}{\phi}}_{m+n+1} = (A_{m+n+1})^{-1} \underset{\sim}{y}_{m+n+1} \qquad (4.32)$$

$$P_{m+n+1} = (A_{m+n+1})^{-1} (A^{-1}_{m+n+1})^{T} \qquad (4.32)$$

The formulation of the estimates for the parameter vector $\underset{\sim}{\phi}$ in this way avoids the matrix inversion required in equation (4.22) and also reduces storage requirements.

### 4.4.1. Transformation of Data - The Generalized Least-Squares Estimators.

The application of least-squares method to the model to estimate its parameters $\underset{\sim}{\phi}$ leads to inconsistent estimates of the process parameters when the observed data are corrupted by noise. However $\underset{\sim}{\phi}$ can be consistently estimated by the least-squares method provided the elements of the disturbance vector $\underset{\sim}{e}_k$ are independently and identically distributed. Many methods have been proposed in the literature to overcome this difficulty (Wong and Polak 1967, Young 1970, Clarke 1967). One of these is the generalised least-squares method (Clarke 1967). In order to arrive at an uncorrelated residual, a noise model is introduced in the system equation and the system parameters are estimated along with the noise model by the method of least-squares. The basic idea is as follows. From equation (4.13) the process can be expressed as:

$$y_k + \sum_{j=1}^{n} b_j y_{k-j} = \sum_{J=0}^{n} a_j r_{k-j} + e_k \qquad (4.34)$$

or $$B(z^{-1})\, y_k = A(z^{-1})\, r_k + e_k \qquad (4.35)$$

where $$B(z^{-1}) = 1 + b_1 z^{-1} + \ldots + b_n z^{-n} \qquad (4.36)$$

$$A(z^{-1}) = a_0 + a_1 z^{-1} + \ldots + a_m z^{-m} \qquad (4.37)$$

and $\{e_k\}$ is the sequence of correlated residuals. Supposing $\{e(k)\}$ can be generated as an autoregressive process of order s:

$$e_k = - f_1 e_{k-1} - f_2 e_{k-2} \ldots - f_s e_{k-s} + w_k \qquad (4.38)$$

where $\{w_k\}$ is assumed to be an uncorrelated random sequence. Thus from (4.38):

$$e_k = \frac{w_k}{F(z^{-1})} \qquad (4.39)$$

where $$F(z^{-1}) = 1 + f_1 z^{-1} \ldots + f_s z^{-s} \qquad (4.40)$$

From equations (4.35) and (4.39) we can write:

$$B(z^{-1})\, y_k^* = A(z^{-1})\, r_k^* + w_k \qquad (4.41)$$

where

$$y_k^* = F(z^{-1})\, y_k$$

$$r_k^* = F(z^{-1})\, r_k \tag{4.42}$$

Hence if the signals $r_k^*$ and $y_k^*$ are considered as the inputs and outputs we have the ordinary least-squares problem. The autoregressive parameters $\psi = \left[f_1, f_2, \ldots f_s\right]^T$ are estimated by:

$$\hat{\underset{\sim}{\psi}}_k = G_k^+ \hat{\underset{\sim}{e}}_k \tag{4.43}$$

where $\hat{\underset{\sim}{\psi}}_k$ = Estimators of $\psi$ at kth iteration

$$= \left[\hat{f}_1 \;\; \hat{f}_2 \ldots \hat{f}_s\right]^T \tag{4.44}$$

$$\hat{e}_k = y_k - \underset{\sim}{a}_k^T \hat{\underset{\sim}{\phi}}_k \tag{4.45}$$

$$\underset{\sim}{a}_k = \left[r_k\, r_{k-1} \ldots r_{k-m} - y_{k-1} - y_{k-2} \ldots - y_{k-n}\right]^T \tag{4.46}$$

$$G_k = \begin{bmatrix} -\hat{e}_o & -\hat{e}_{-1} & \ldots & -\hat{e}_{1-s} \\ -\hat{e}_1 & -\hat{e}_o & \ldots & -\hat{e}_{2-s} \\ -\hat{e}_{k-1} & -\hat{e}_{k-2} & \ldots & -\hat{e}_{k-s} \end{bmatrix} \tag{4.47}$$

and $G_k^+$ is the pseudoinverse of $G_k$ i.e.

$$G_k^+ = \left[G_k^T \; G_k\right]^{-1} G_k \quad \text{for } k > s \tag{4.48}$$

Utilising this estimate $\hat{\underset{\sim}{\psi}}_k$, the noisy input-output sequences $r_k$ and $y_k$ are filtered to obtain:

$$r_k^* = r_k + \sum_{j=1}^{s} \hat{f}_j \, r_{k-j} \qquad (4.49)$$

$$y_k^* = y_k + \sum_{j=1}^{s} \hat{f}_j \, y_{k-j} \qquad (4.50)$$

From these filtered sequences a corrected matrix $A_k^*$ is obtained instead of $A_k$ in equation (4.9) using $r_k^*$ and $y_k^*$ instead of $r_k$ and $y_k$ respectively. Equation (4.22) can now be used with $A_k^*$ to obtain unbiased estimations.

Due to highly nonlinear relationships between the parameters in the estimating equation, a theoretical analysis of convergence is not yet available. Experimental results reported in the literature have shown, however, that this algorithm works quite well. In the following section, a recursive algorithm which is particularly well suited for on-line implementation will be developed.

### 4.4.2. Recursive Algorithm for Generalized Least-Squares (Generalized Pseudo-inverse) Parameter Estimation.

A recursive version of the generalised least-squares estimation procedure, based on the recursive algorithm for computing the matrix pseudoinverse, is proposed. Because of its recursive nature, the algorithm can be used for on-line identification.

Defining:

$$G_{k+1} = \begin{bmatrix} G_k \\ \underset{\sim}{g}_{k+1}^T \end{bmatrix} \tag{4.51}$$

$$\underset{\sim}{g}_{k+1}^T = \left[-\hat{e}_k \; -\hat{e}_{k-1} \; \cdots \; -\hat{e}_{k+1-s}\right] \tag{4.52}$$

$$A_{k+1}^* = \begin{bmatrix} A_k^* \\ \underset{\sim}{a}_{k+1}^{*T} \end{bmatrix} \tag{4.53}$$

$$\underset{\sim}{a}_{k+1}^{*T} = \left[r_{k+1}^* \; r_k^* \; \cdots \; r_{k+1-m}^* \; -y_k^* \; -y_{k-1}^* \; \cdots \; -y_{k+1-n}^*\right] \tag{4.54}$$

the following equations can easily be derived for $k>m+n+1$:

$$\underset{\sim}{\hat{\phi}}_{k+1} = \underset{\sim}{\hat{\phi}}_k + \frac{P_k \, \underset{\sim}{a}_{k+1}^* \, (y_{k+1}^* - \underset{\sim}{a}_{k+1}^{*T} \, \underset{\sim}{\hat{\phi}}_k)}{1 + \underset{\sim}{a}_{k+1}^{*T} \, P_k \, \underset{\sim}{a}_{k+1}^*} \tag{4.55}$$

$$P_{k+1} = P_k - \frac{P_k \, \underset{\sim}{a}_{k+1}^* \, (P_k \, \underset{\sim}{a}_{k+1}^*)^T}{1 + \underset{\sim}{a}_{k+1}^{*T} \, P_k \, \underset{\sim}{a}_{k+1}^*} \tag{4.56}$$

$$\underset{\sim}{\hat{\psi}}_{k+1} = \underset{\sim}{\hat{\psi}}_k + \frac{R_k \, \underset{\sim}{g}_{k+1} \, (\hat{e}_{k+1} - \underset{\sim}{g}_{k+1}^T \, \underset{\sim}{\hat{\psi}}_k)}{1 + \underset{\sim}{g}_{k+1}^T \, R_k \, \underset{\sim}{g}_{k+1}} \tag{4.57}$$

and

$$R_{k+1} = R_k - \frac{R_k \, \underset{\sim}{g}_{k+1} (R_k \, \underset{\sim}{g}_{k+1})^T}{1 + \underset{\sim}{g}_{k+1}^T \, R_k \, \underset{\sim}{g}_{k+1}} \tag{4.58}$$

The values of $P_{m+n+1}$, $R_s$, $\hat{\phi}_{m+n+1}$, $\hat{\psi}_s$ are determined from the following:

$$\hat{\underset{\sim}{\phi}}_{m+n+1} = (A^*_{m+n+1})^{-1} \underset{\sim}{y}^*_{m+n+1} \tag{4.59}$$

$$\hat{\psi}_s = G_s^{-1} \hat{\underset{\sim}{e}}_s$$

$$P_{m+n+1} = \left[A^*_{m+n+1}\right]^{-1} \left[A^{*-1}_{m+n+1}\right]^T \tag{4.60}$$

and

$$R_s = G_s^{-1} \left[G_s^{-1}\right]^T \tag{4.61}$$

The entire process may have to be repeated a number of times in order that the residuals may be uncorrelated.

The generalised pseudoinverse algorithm proposed here makes it much more efficient to iteratively estimate the parameters of the system model as well as the auto-regressive parameters of the noise model. The results of applying this algorithm for system identification will be presented in a later chapter.

### 4.5.1. Instrumental Variable Method for Process Parameter Estimation.

The Instrumental Variable technique which provides consistent estimates of structural parameters of the model was first used in engineering context by Joseph, Lewis, and Tow (1961). Since then it has been discussed by Wong

and Polak (1967), Rowe (1970), Young (1970) and Pandya (1972) among others.

The Instrumental Variable approach aims at removing the source of inconsistency in ordinary least-squares estimaters in a conceptually simple manner. The principle is easily understood by rewriting the equation (4.8):

$$\underset{\sim}{e}_k = \underset{\sim}{y}_k - A_k \underset{\sim}{\phi}_k \tag{4.62}$$

Consistent estimates can be obtained by premultiplying equation (4.62) with $W_k^T$

$$W_k^T \underset{\sim}{e}_k = W_k^T \underset{\sim}{y}_k - W_k^T A_k \underset{\sim}{\phi}_k \tag{4.63}$$

where $W_k^T$ is called the instrumental matrix which satisfies:

$$\underset{k\to\infty}{p.lim} \left[\frac{1}{k} W_k^T \underset{\sim}{e}_k\right] = 0 \tag{4.64}$$

and

$$\underset{k\to\infty}{p.lim} \left[\frac{1}{k} W_k^T A_k\right] \text{ nonsingular} \tag{4.65}$$

The elements of $W_k$ are, therefore, chosen to be uncorrelated with the residuals but correlated with the uncorrupted signals of the process. Then, from equation (4.63) we have:

$$\underset{\sim}{\hat{\phi}}_k = \left[W_k^T A_k\right]^{-1} W_k^T \underset{\sim}{y}_k \tag{4.67}$$

Wong and Polak (1967) and Young (1970) showed that there exist optimal instrumental variables, and they used samples of calculated undisturbed output signal as instrumental variables, taking the parameter estimates as the parameters of an auxilary model. If h is the output of the auxiliary model then $W_k$ becomes:

$$W_k = \begin{bmatrix} r_1 & r_o & \cdots & r_{1-m} & -h_0 & -h_{-1} & \cdots & -h_{1-n} \\ r_2 & r_1 & & r_{2-m} & -h_1 & -h_0 & \cdots & -h_{2-n} \\ \cdot & \cdot & & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & & \cdot \\ r_k & r_{k-1} & \cdots & r_{k-m} & -h_{k-1} & -h_{k-2} & & -h_{k-n} \end{bmatrix} \qquad (4.68)$$

The rate of convergence of the estimates to the true parameter value depends on the choice of the instrumental variables and there are many sets of instrumental variables that satisfy conditions (4.64) and (4.65). Thus one of the basic problems in this method is to find a suitable set of instrumental variables. Also, since the optimal instrumental variables cannot be generated directly, extra effort has to be made to compute approximately the optimal instrumental variables.

4.5.2. Recursive Instrumental Variable Algorithm:

It has been established by Wong and Polak (1967) and Young (1970), that the instrumental variable estimator has the following recursive form:

$$\underset{\sim}{\hat{\phi}}_{k+1} = \underset{\sim}{\hat{\phi}}_{k} + \frac{P_k \, \underset{\sim}{\omega}_{k+1} \, (y_{k+1} - \underset{\sim}{a}^{T}_{k+1} \, \underset{\sim}{\hat{\phi}}_{k})}{1 + \underset{\sim}{a}^{T}_{k+1} \, P_k \, \underset{\sim}{\omega}_{k+1}} \tag{4.69}$$

$$P_{k+1} = P_k - \frac{P_k \, \underset{\sim}{\omega}_{k+1} \, (P_k \, \underset{\sim}{\omega}_{k+1})^{T}}{1 + \underset{\sim}{a}^{T}_{k+1} \, P_k \, \underset{\sim}{\omega}_{k+1}} \tag{4.70}$$

where

$$P_k = A_k^{-1} \, (W_k^{-1})^{T} \tag{4.71}$$

and

$$\underset{\sim}{\omega}^{T}_{k+1} = \left[ r_{k+1} \, r_k \, \cdots \, r_{k+1-m} \, -h_k \, -h_{k-1} \, \cdots \, -h_{k+1-n} \right] \tag{4.72}$$

Young (1970) introduced a time delay and a low-pass filter, before updating the auxiliary model, to ensure that auxiliary model parameters are not correlated with $\underset{\sim}{e}_k$ at the same instant and to smooth the estimates. The following low-pass filter is used:

$$\underset{\sim}{\hat{\phi}}^{aux}_{k} = (1-\nu) \, \underset{\sim}{\hat{\phi}}^{aux}_{k-1} + \nu \, \underset{\sim}{\hat{\phi}}_{k} \tag{4.74}$$

where $0.03 \leq \nu \leq 1$ is chosen to prevent instability in the estimation.

The recursive form not only permits the updating of the parameters in real time, but also generates the instrumental variables in real time. Thus it is not essential that all the elements of the instrumental matrix $W_k$ be known a-priori.

4.6.1. Bootstrap Estimator.

Bootstrap estimators proposed by Rowe (1970) and extended by Pandya (1972) belong to the class of on-line estimators that utilize an auxiliary dynamic model to generate recursively the estimates of noise-free output. Needless to say, one of the estimators in this set of estimators is the instrumental variable estimator. An on-line algorithm for this estimator was given in the previous section. The bootstrap estimators are characterised by the fact that the structural parameters $\underset{\sim}{\phi}$ are estimated recursively in such a way that the latest estimates of the structural parameters are used in estimating the incidental parameters and vice versa. Thus the basic bootstrap estimator consists of two recursive estimators — one for the structural parameters and the other for the incidental parameters.

The recursive algorithm as proposed by Pandya (1972) has the following form:

$$\hat{\underset{\sim}{\phi}}_{k+1} = \hat{\underset{\sim}{\phi}}_{k} + \frac{P_k \, \underset{\sim}{z}_{k+1} \left( y_{k+1} - \underset{\sim}{a}^{T}_{k+1} \, \hat{\underset{\sim}{\phi}}_{k} \right)}{1 + \underset{\sim}{a}^{T}_{k+1} \, P_k \, \underset{\sim}{z}_{k+1}} \qquad (4.78)$$

$$P_{k+1} = P_k - \frac{P_k \underset{\sim}{z}_{k+1} \underset{\sim}{a}^T_{k+1} P_k}{1 + \underset{\sim}{a}^T_{k+1} P_k \underset{\sim}{z}_{k+1}} \qquad (4.79)$$

$$\underset{\sim}{z}^T_{k+1} = \left[ r_{k+1}\ r_k \cdots r_{k+1-m} - \hat{\chi}_{k/k} \cdots - \hat{\chi}_{k-n+1/k-n+1} \right] \qquad (4.80)$$

where

$$\hat{\chi}_{k/k} = \hat{\chi}_{k/k-1} + \nu(k) \left[ y_k - \hat{\chi}_{k/k-1} \right] \quad 0 \leq \nu(k) \leq 1 \qquad (4.81)$$

and $$\hat{\chi}_{k/k-1} = \underset{\sim}{z}^T_k \hat{\underset{\sim}{\phi}}_k \qquad (4.82)$$

The gain sequence $\nu(k)$ is chosen in such a way as to prevent instability in the estimation. By introducing the filter the algorithm becomes more flexible. This is because the estimator now depends on the nature of filter gain sequence $\nu(k)$. There are many choices of $\nu(k)$ as discussed by Pandya (1972) but no general theory is available that dictates the optimal choice of $\nu(k)$. Moreover stability of the procedure cannot be guaranteed.

4.7.1. Correlation as Applied to the Estimation of Parameters.

The method essentially consists of first determining correlation functions and then estimating the parameters of the desired parametric model by the method of least-squares. If the input is a stationary random variable, then its autocorrelation function is given by:

$$R_{uu}(\tau) = \lim_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} r_k \; r_{k-\tau} \qquad (4.83)$$

and the cross-correlation between the input and output signals is:

$$R_{uy}(\tau) = \lim_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} r_{k-\tau} \; y_k \qquad (4.84)$$

The convolution equation

$$R_{uy}(\tau) = \sum_{\nu=0}^{\infty} g(\nu) \; R_{uu}(\tau-\nu) \qquad (4.85)$$

relates the input-output through the impulse response $g(\nu)$ of the process. From (4.85) we get:

$$\underset{\sim}{R}_{uy}(M) = \underset{\sim}{R}_{uu}(M) \; \underset{\sim}{g} \qquad (4.86)$$

where

$$\underset{\sim}{R}_{uy}(M) = \begin{bmatrix} R_{uy}(-M) \\ \vdots \\ R_{uy}(-1) \\ R_{uy}(0) \\ \vdots \\ R_{uy}(M) \end{bmatrix} \qquad (4.87)$$

$$\underset{\sim}{R}_{uu}(M) = \begin{bmatrix} R_{uu}(-M) & \cdots & R_{uu}(-M-\ell) \\ \vdots & & \vdots \\ R_{uu}(-1) & \cdots & R_{uu}(-1-\ell) \\ R_{uu}(0) & \cdots & R_{uu}(-\ell) \\ \vdots & & \vdots \\ R_{uu}(M) & \cdots & R_{uu}(M-\ell) \end{bmatrix} \tag{4.88}$$

$$\underset{\sim}{g} = \begin{bmatrix} g(0) & \cdots & g(\ell) \end{bmatrix}^T \tag{4.89}$$

The least-squares estimate $\hat{\underset{\sim}{g}}(M)$ of $\underset{\sim}{g}$ is given by:

$$\hat{\underset{\sim}{g}}(m) = \underset{\sim}{R}^{+}_{uu}(M)\, R_{uy}(M) \tag{4.90}$$

where $\underset{\sim}{R}^{+}_{uu}(M)$ is the pseudoinverse of $\underset{\sim}{R}_{uu}(M)$ defined by:

$$\underset{\sim}{R}^{+}_{uu}(M) = \left[\underset{\sim}{R}^{T}_{uu}(M)\, \underset{\sim}{R}_{uu}(M)\right]^{-1} \underset{\sim}{R}^{T}_{uu}(M) \tag{4.91}$$

One may, again, use the recursive pseudoinverse algorithm with the correlation ordinates and thereby get an on-line procedure. If there are $\ell = m+n+1$ unknown parameters then $\ell = m+n+1$ values of the impulse response $g(i)$ are estimated by equation (4.90) and the parameters $a_i$ and $b_i$ are then calculated using

the estimated g(i)'s. The autocorrelation and the cross-correlation function can also be obtained recursively using:

$$R_{uu}(\tau,k) = R_{uu}(\tau,k-1) + \frac{1}{k+1}\left[r_{k-\tau}\, r_k - R_{uu}(\tau,k-1)\right] \tag{4.92}$$

$$R_{uy}(\tau,k) = R_{uy}(\tau,k-1) + \frac{1}{k+1}\left[r_{k-\tau}\, y_k - R_{uy}(\tau,k-1)\right] \tag{4.93}$$

The correlation method is relatively easy to implement because computation time and storage requirements are small. The accuracy of the result can be controlled by changing the number $\ell$ of impulse response samples. Thus the computational expense can be reduced by accepting a loss of accuracy of the model. For practical implementation, the estimates can sometimes be very sensitive to the choice of M in equation (4.91) and there is no a-priori way of choosing the optimum value of M.

### 4.8.1. Stochastic Approximation Algorithm.

One of the most popular methods for parameter identification is the Stochastic Approximation method. It was first introduced by Robbins and Monro (1951) and generalized by Dvoretzky (1956). An extensive treatment of this method was given by Albert and Gardner in (1967). Comprehensive survey papers by Sakrison (1966) and Saridis (1974) give a good general picture of the various aspects of the subject. Applications of Stochastic approximation algorithms have been proposed in adaptive and learning control by Sklansky (1966),

in system identification by Saridis, Nikolic and Fu (1969), and in adaptive communication and pattern recognition by Cooper (1964).

Stochastic approximation may be defined as a scheme for successive approximation of a sought quantity, when the observations involve random errors due to the stochastic nature of the problem. It can be applied to any problem that can be formulated as some form of regression in which repeated observations are made. It has the following advantages:

(a) Only a small amount of data needs processing.

(b) Only simple computations are required, even when the actual functional dependence of the regression function on the parameters of interest is nonlinear.

(c) A-priori knowledge of the process statistics is not necessary nor is the detailed knowledge of the functional relationship between the desired parameters and the observed data. The only requirements are that the regression function satisfy certain regularity conditions and that the regression problem have a unique solution.

Let us first look at the Robbins-Monro (1951) approach to search for the roots of an unknown function $f(x)$ of random variable $x$ corrupted with measurement noise $\nu$:

$$z(x) = f(x) + \nu \tag{4.94}$$

The following iterative algorithm is proposed:

$$x_{k+1} = x_k - \gamma_k z(x_k) \tag{4.95}$$

Robbins and Monro showed that the equation (4.95) will converge with probality one and in the mean square sense, if the following conditions are met:

$$\lim_{k\to\infty} \gamma_k = 0 , \tag{4.96}$$

$$\sum_{k=1}^{\infty} \gamma_k = \infty , \tag{4.97}$$

and

$$\sum_{k=1}^{\infty} \gamma_k^2 < \infty \tag{4.98}$$

A simple sequence which meets these requirements is:

$$\gamma_k = \frac{\alpha}{\beta+k} \tag{4.99}$$

where $\alpha$ and $\beta$ are positive constants. It is also required that $f(x)$ be bounded on either side of a true solution by straight lines such that it is not possible to overshoot the solution $x$ which cannot be corrected by a $\gamma_k$ satisfying (4.99).

Kiefer and Wolfowitz (1952) proposed the second important stochastic approximation algorithm, suitable for the search of an extremum of the function f(x) of a random variable x corrupted with measurement noise v of the Robbins-Monro method. The following iterative algorithm is proposed:

$$x_{k+1} = x_k + \gamma_k \frac{z(x_k + C_k) - z(x_k - C_k)}{2 C_k} \tag{4.100}$$

Kiefer and Wolfowitz (1952) showed that the equation will converge with probability one and in the mean square sense if the following conditions are met:

$$\lim_{k \to \infty} \gamma_k = 0 \tag{4.101}$$

$$\lim_{k \to \infty} C_k = 0 \tag{4.102}$$

$$\lim_{k \to \infty} \sum_{n=1}^{k} \left(\frac{\gamma_n}{C_n}\right)^2 < \infty \tag{4.103}$$

$$\lim_{k \to \infty} \sum_{n=1}^{k} \gamma_n = \infty \tag{4.104}$$

Detailed presentations of the mathematical background, applications and development of stochastic approximation algorithms are given by Wasan (1969), Mendel and Fu (1970), Albert and Gardner (1967) and Holmes (1969).

4.8.2. Stochastic Approximation Algorithms as Applied to Parameter Estimation.

Most of the stochastic approximation algorithms used for recursive identification of the dynamic systems are based on the Robbins-Monro or Kiefer-Wolfowitz algorithms. Saridis and Stein (1968) have proposed such an algorithm for obtaining estimation of the impulse response. It is of the form:

$$\hat{\underset{\sim}{G}}(k+\ell) = \hat{\underset{\sim}{G}}(k-1) + \nu(k)\, \underset{\sim}{U}(k+\ell) \left[ y(k+\ell) - \underset{\sim}{U}^T(k+\ell)\, \hat{\underset{\sim}{G}}(k-1) \right] \quad (4.105)$$

with

$$\hat{\underset{\sim}{G}}^T = \left[ g_1\ g_2\ \ldots\ g_\ell \right] \quad (4.106)$$

where $g_i$ are the samples of impulse response

$$\underset{\sim}{U}^T(k) = \left[ r_k\ r_{k-1}\ \ldots\ r_{k-\ell-1} \right] \quad (4.107)$$

$$\nu(k) = \frac{1}{\xi(k)} , \quad (4.108)$$

$$\xi(k) = \frac{k-\ell}{\ell+1} \quad k=1,\ \ell+2,\ 2\ell+3\ldots \quad (4.109)$$

If there are m+n+1 unknown parameters $\ell$ = m+n+1 values of the impulse response $g_i$ are estimated by equation (4.105) and the parameters $a_i, b_i$ are then calculated using the estimated $g_i$'s updating algorithm (4.105) after each sampling interval leads:

$$\hat{\underset{\sim}{G}}(k+1) = \hat{\underset{\sim}{G}}(k) + \nu(k+1)\, \underline{U}(k+1) \left[ y_{k+1} - \underline{U}^T(k+1)\, \hat{\underset{\sim}{G}}(k) \right] \quad (4.110)$$

with $$\nu(k) = 1/k \quad k = 1, 2, 3.... \qquad (4.111)$$

This algorithm gives biased estimates, but the bias is relatively small even for large noise-to-signal ratio. This algorithm is simple and easy to implement but demonstrates rather slow convergence.

### 4.9.1. Proposed Algorithm for On-Line System Identification Combining Stochastic Approximation and Pseudoinverse.

In a recent survey (Sinha and Sen 1974) of various methods for on-line identification of a simulated system, under realistic conditions of noisy measurement of input-output data, it has been observed that whereas recursive pseudoinverse algorithm is computationally efficient, the resulting parameter estimates are biased and inconsistent. On the other hand, although the ordinary stochastic approximation algorithm of the form described in equation (4.110) gives biased estimates, it is simple and easy to implement requiring less computation per iteration. The aim of the proposed algorithm is to combine pseudoinverse and stochastic approximation to overcome the problem of bias encountered in the pseudoinverse algorithm in the presence of large noise while requiring less computation to make it practical for use in on-line identification. The method is based on getting generalized least squares estimates of the parameters of the system model as well as the auxiliary noise parameters in such a way as to obtain uncorrelated residuals. The main difference between the proposed

method and the generalized least squares method is that a Dvoretzky-type (1956) stochastic approximation algorithm presented by Kwatny (1972) is used for estimating the parameters of the noise model, whereas recursive pseudoinverse algorithm is used for estimating the parameters of the system model. This solution, which is iterative is shown to be computationally more efficient than the generalised least squares algorithm.

In the next section the noise model is formulated in terms of which the observed input-output data are filtered to get uncorrelated residuals. This results in an iterative procedure by which the filter parameters and system parameters are alternately estimated.

4.9.2. Formulation of the Noise Model.

Let the system dynamic equation be represented as in equation (4.13) in the following form:

$$B(z^{-1})\, y_k = A(z^{-1})\, r_k + e_k \qquad (4.112)$$

where

$$e_k = (1 + \sum_{j=1}^{n} b_j\, z^{-j})\, n_k = B(z^{-1})\, n_k \qquad (4.113)$$

and $n_k$ is a zero mean random noise sequence. To guarantee stability of the process, the roots of $B(z^{-1})$ are assumed to lie outside the unit circle.

It is assumed that the noise sequence $\{n_k\}$ can be described as a linear transformation of a well behaved zero mean white noise signal $\xi_i$, i.e.

$$n_k + \sum_{j=1}^{p} d_j\, n_{k-j} = \xi_k + \sum_{i=1}^{q} g_i\, \xi_{k-i} \qquad (4.114)$$

or equivalently,

$$n_k = \frac{G(z^{-1})}{D(z^{-1})}\, \xi_k \qquad (4.115)$$

where $$D(z^{-1}) = 1 + d_1 z^{-1} \ldots + d_p z^{-p} \qquad (4.116)$$

and $$G(z^{-1}) = 1 + g_1 z^{-1} \ldots + g_q\, z^{-q} \qquad (4.117)$$

Again the roots of $D(z_i^{-1})$ are assumed to be outside the unit circle. Combining equations (4.113) and (4.115), we have:

$$e_k = \frac{B(z^{-1})\ G(z^{-1})}{D(z^{-1})}\ \xi_k. \tag{4.118}$$

$$= \frac{C(z^{-1})}{D(z^{-1})}\ \xi_k \tag{4.119}$$

where we write:

$$C(z^{-1}) = B(z^{-1})\ G(z^{-1}) \tag{4.120}$$

In (4.119) $e_k$ is described as a sample of mixed autoregressive moving-average time series. The above residual error sequence $\{e_k\}$ is now approximated by a low order linear process. Two possible processes are suitable for this purpose. They are the moving average process of the form:

$$e_k = \xi_k + \sum_{r=1}^{p} m_r\ z^{-r}\ \xi_k \tag{4.121}$$

and the autoregressive process of the form:

$$e_k + \sum_{i=1}^{p} f_i\ z^{-i}\ e_k = \xi_k. \tag{4.122}$$

The true process can be approximated to any degree of accuracy by choosing an appropriate sequence $\{f_i\}$. Using this principle equation (4.119) can therefore be approximated by:

$$e_k = \frac{\xi_k}{F_s(z^{-1})} \tag{4.123}$$

where

$$F_s(z^{-1}) = 1 + f_1 z^{-1} \ldots + f_s z^{-1} \qquad (4.124)$$

This implies that:

$$F_s(z^{-1}) \approx D(z^{-1}) \left[B(z^{-1})\, G(z^{-1})\right]^{-1} \qquad (4.125)$$

This is true if sufficient number of $f_i$'s of the filter are used. Substituting (4.123) into (4.112), we have:

$$B(z^{-1})\, y_k = A(z^{-1})\, r_k + \frac{\xi_k}{F_s(z^{-1})} \qquad (4.126)$$

$$B(z^{-1})\, F_s(z^{-1})\, y_k = A(z^{-1})\, F_s(z^{-1})\, r_k + \xi_k \qquad (4.127)$$

$$B(z^{-1})\, y_k^* = A(z^{-1})\, r_k^* + \xi_k \qquad (4.128)$$

where

$$y_k^* = (1 + f_1 z^{-1} + \ldots + f_s z^{-s}) y_k \qquad (4.129)$$

$$r_k^* = (1 + f_1 z^{-1} + \ldots + f_s z^{-s}) r_k \qquad (4.130)$$

and $\xi_k$ is a white noise sequence.

If the filter is known, the measurements $r_k$, $y_k$ will be processed as in equations (4.129) and (4.130) to obtain the filtered input and output pair $r_k^*$ and $y_k^*$. These filtered quantities are then used to calculate the estimates. Comparing equations (4.128) and (4.112), the residual error sequence has now been changed to a white noise sequence. Hence the

least-squares estimates will be consistent and unbiased because the white residual error is not correlated with the input and output.

Once the structure of the filter is known, the next step is to find the parameters of the filter $F_s(z^{-1})$. The residuals are estimated by assuming the model:

$$\hat{e}_k = \underset{\sim}{\psi}^T \underset{\sim}{\varepsilon}_k + \xi_k \tag{4.131}$$

$$\underset{\sim}{\psi} = \left[f_1\ f_2\ \ldots\ f_s\right]^T, \text{ the filter parameter vector} \tag{4.132}$$

$$\underset{\sim}{\varepsilon}_k = \left[-\hat{e}_{k-1}\ \ -\hat{e}_{k-2}\ \ldots\ -\hat{e}_{k-s}\right]^T, \text{ the error vector} \tag{4.133}$$

$$\hat{e}_k = y_k - \underset{\sim}{a}_k^T \underset{\sim}{\hat{\phi}}_k \tag{4.134}$$

To obtain the estimate of $\underset{\sim}{\psi}$, it is proposed to use a stochastic approximation algorithm of the same form as suggested by Kwatny (1972), i.e.

$$\underset{\sim}{\hat{\psi}}_{k+1} = \underset{\sim}{\hat{\psi}}_k - \frac{\nu}{k+1} \frac{\left[\underset{\sim}{\psi}_k^T \underset{\sim}{\varepsilon}_k - \hat{e}_k\right] \underset{\sim}{\varepsilon}_k}{||\underset{\sim}{\varepsilon}_k||^2} \tag{4.135}$$

where $\nu$ is a positive constant and

$$\underset{\sim}{\hat{\psi}}_k = \text{kth estimate of } \underset{\sim}{\hat{\psi}}$$

$$= \left[f_1(k),\ f_2(k)\ \ldots\ \hat{f}_s(k)\right]^T \tag{4.136}$$

Using the estimated filter $\hat{\underset{\sim}{\psi}}_k$, the input and output measurements are filtered to obtain:

$$r_k^* = r_k + \sum_{i=1}^{s} \hat{f}_i(k)\, r_{k-i} \tag{4.137}$$

$$y_k^* = y_k + \sum_{i=1}^{s} \hat{f}_i(k)\, y_{k-i} \tag{4.138}$$

These filtered quantities are used in updating the information matrix $A_k$ in the pseudoinverse algorithm. Thus the complete algorithm reads as follows:

$$r_{k+1}^* = r_{k+1} + \sum_{i=1}^{s} \hat{f}_i(k) r_{k+1-i} \tag{4.139}$$

$$y_{k+1}^* = y_{k+1} + \sum_{i=1}^{s} \hat{f}_i(k) y_{k+1-i} \tag{4.140}$$

$$\hat{\underset{\sim}{\phi}}_{k+1} = \hat{\underset{\sim}{\phi}}_k + \frac{P_k\, \underset{\sim}{a}^*_{k+1}(y^*_{k+1} - \underset{\sim}{a}^{*T}_{k+1}\, \hat{\underset{\sim}{\phi}}_k)}{1 + \underset{\sim}{a}^{*T}_{k+1}\, P_k\, \underset{\sim}{a}^*_{k+1}} \tag{4.141}$$

$$P_{k+1} = P_k - \frac{P_k\, \underset{\sim}{a}^*_{k+1} (P_k\, \underset{\sim}{a}^*_{k+1})^T}{1 + \underset{\sim}{a}^{*T}_{k+1}\, P_k\, \underset{\sim}{a}^*_{k+1}} \tag{4.142}$$

$$\hat{e}_k = y_k - \underset{\sim}{a}^{*T}_k\, \hat{\underset{\sim}{\phi}}_k \tag{4.143}$$

$$\hat{\underset{\sim}{\psi}}_{k+1} = \hat{\underset{\sim}{\psi}}_k - \frac{\nu}{k+1} \frac{\left[\underset{\sim}{\psi}_k^T\, \underset{\sim}{\varepsilon}_k - \hat{e}_k\right] \underset{\sim}{\varepsilon}_k}{||\underset{\sim}{\varepsilon}_k||^2} \tag{4.144}$$

where

$$\underset{\sim}{a}^{*T}_{k+1} = \left[r^*_{k+1}\, r^*_k \cdots r^*_{k+1-m} - y^*_k - y^*_{k-1} \cdots - y^*_{k-1-n}\right] \tag{4.145}$$

$$\underset{\sim}{\varepsilon}_k = \left[-\hat{e}_{k-1} - \hat{e}_{k-2} \cdots - \hat{e}_{k-s}\right]^T \tag{4.146}$$

Thus, the proposed algorithm consists of using stochastic approximation to obtain the autoregressive noise model parameters, and, recursive pseudoinverse to determine the process model parameters, after the input-output data are suitably filtered utilising the noise model. Obviously the algorithm as represented by equations (4.139) to (4.146) consists of two recursive estimators - one for the process parameter $\phi$ and the other for the noise parameter, where the latest estimates of the noise parameters are used in estimating the process parameters and vice versa.

The convergence of the algorithm has not been theoretically justified. When these two estimators are treated seperately, the convergence of each can be argued if the other one satisfies certain properties. Kwatny (1972) has presented conditions under which algorithm of equation (4.135) converge in the mean. However, the convergence of the overall estimator is not obvious. This is because the behavior of one affects the operation of the other.

Ljung (1974) has given conditions under which general recursive algorithms with stochastic observations converge. The

recursive algorithm can in general be put in the following form

$$\hat{\phi}_n = \hat{\phi}_{n-1} + \nu_n H_n (\hat{\phi}_{n-1}, y_n)$$

where $\hat{\phi}_n$ is nth estimate of $\phi_n$. The correction term $\gamma_n H_n (\hat{\phi}_{n-1}, y_n)$ is function of previous estimate $\hat{\phi}_{n-1}$ and observation $y_n$. Ljung (1974) has suggested that the problem of convergence can be studied by convergence and stability analysis of an ordinary differential equation. The problem of convergence of the proposed algorithm combining pseudoinverse and stochastic approximation algorithm, it is suggested, may be treated along the lines of Ljung (1974) and is left for future investigation.

The algorithm is tested extensively on simulated data and on data from real processes under open loop operation. Since the stochastic approximation algorithm used to estimate the noise parameters requires very little computation per iteration the proposed algorithm is computationally more efficient relative to the recursive version of generalised least squares algorithm. This advantage is further enhanced as data length and the dimension of $\phi$ increases since increased data filtering time is required for generalised least squares. This proposed algorithm also retains the advantage of generalised least squares algorithm in that the bias is considerably reduced because a noise model is incorporated in the algorithm in the same way as in the generalised least squares algorithm. The performance of the algorithm is illustrated in the next chapter.

Because of the relatively less computational requirement the proposed algorithm can easily be implemented in real-time. The implementation of this algorithm using PDP-11/45 has been reported by Tang (1975).

# CHAPTER 5

# COMPARISON OF DIFFERENT ON-LINE METHODS FOR SINGLE-INPUT SINGLE-OUTPUT SYSTEM

## 5.1. Introduction.

The various identification algorithms that perform on-line by sequentially updating the parameter estimates from noisy measurements are discussed in Chapter 4. In this chapter the performance of each on-line method of identification is evaluated and compared. For comparison four test cases are considered (i) a simulated second-order system with output noise added to obtain different noise-to-signal ratios, (ii) a two-stage heat-exchanger system, (iii) a continuous stirred tank process, and (iv) a dual-input heat exchanger system. The test problems will now be described.

### 5.1.1. The Simulated Second-order Process.

The process is described by:

$$x_{k+1} = a_1 r_k + a_2 r_{k-1} - b_1 x_k - b_2 x_{k-1} \qquad (5.1)$$
$$y_k = x_k + \alpha n_k$$

where $a_1 = 0.079$, $a_2 = 0.047$, $b_1 = -0.975$, and $b_2 = 0.223$. The input sequence $r_k$ was a zero mean white Gaussian noise sequence with unit variance. Two different cases of the output

noise sequence, $n_k$, were considered, (i) when it was a zero-mean white Gaussian noise sequence of unit variance, uncorrelated with $r_k$, and (ii) when it was a coloured noise sequence, obtained as the output of a first-order digital filter with white-noise input, i.e.

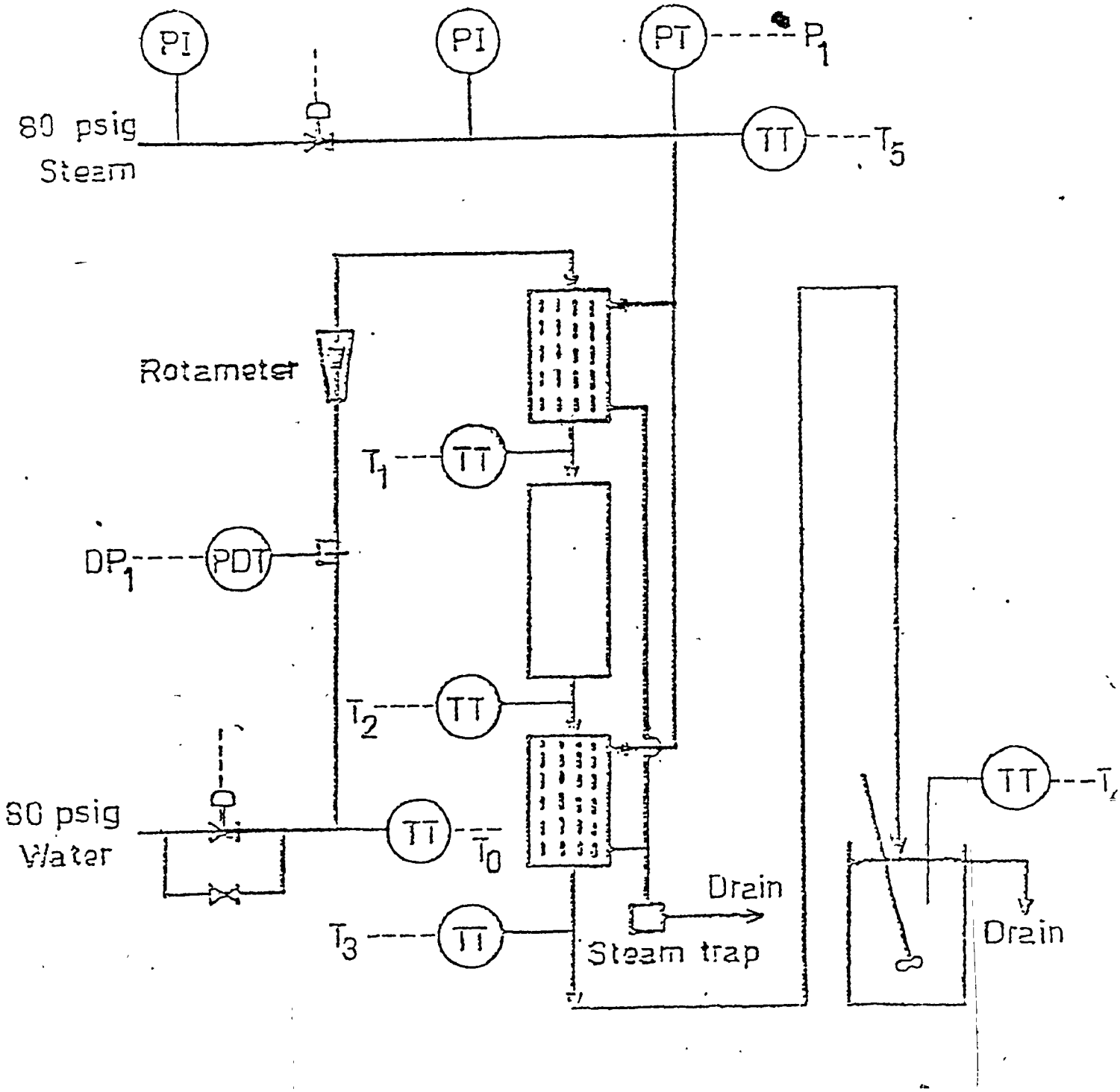
$$n_{k+1} = 0.8\ n_k + 0.7\ w_k \qquad (5.3)$$

where $w_k$ is a unit variance white noise sequence uncorrelated with $r_k$. The value of the constant $\alpha$ was adjusted for different noise-to-signal ratios, varying between 20% and 100%. In particular, with $\alpha$ equal to 0.04, 0.08 and 0.2, the ratios of the standard deviation of the noise and the signal correspond to 20%, 40% and 100%, respectively.

### 5.1.2. The Two-Stage Heat-Exchanger Process

The heat exchanger system selected for this example has been studied extensively by Wright and Bacon (1975), who have obtained models using time series analysis of Box and Jenkins (1970). For detail description of the process, reference may be made to the paper by Wright and Bacon (1975). A schematic diagram of the process was shown in Figure 5.1.

For our purpose, the process may be considered as a system with one input and one output. The input variable is the steam valve position and the output variable is the water temperature $T_e$. For our identification purpose 300 pairs of input-output




† Figure 5.1 Schematic diagram of a two-stage heat-exchanger system.

† The Figure 5.1 has been reproduced from the paper of Wright and Bacon (1975).

data, taken at 5 second sampling interval have been supplied to the author by Dr. J.D. Wright.

Following the procedure of Box and Jenkins (1970) it was found that the system can be adequately represented, within 95% confidence limits, by the following difference equation model:

$$Y_k = bY_{k-1} + a_1 u_{k-1} + a_2 u_{k-2}, \text{ where}$$

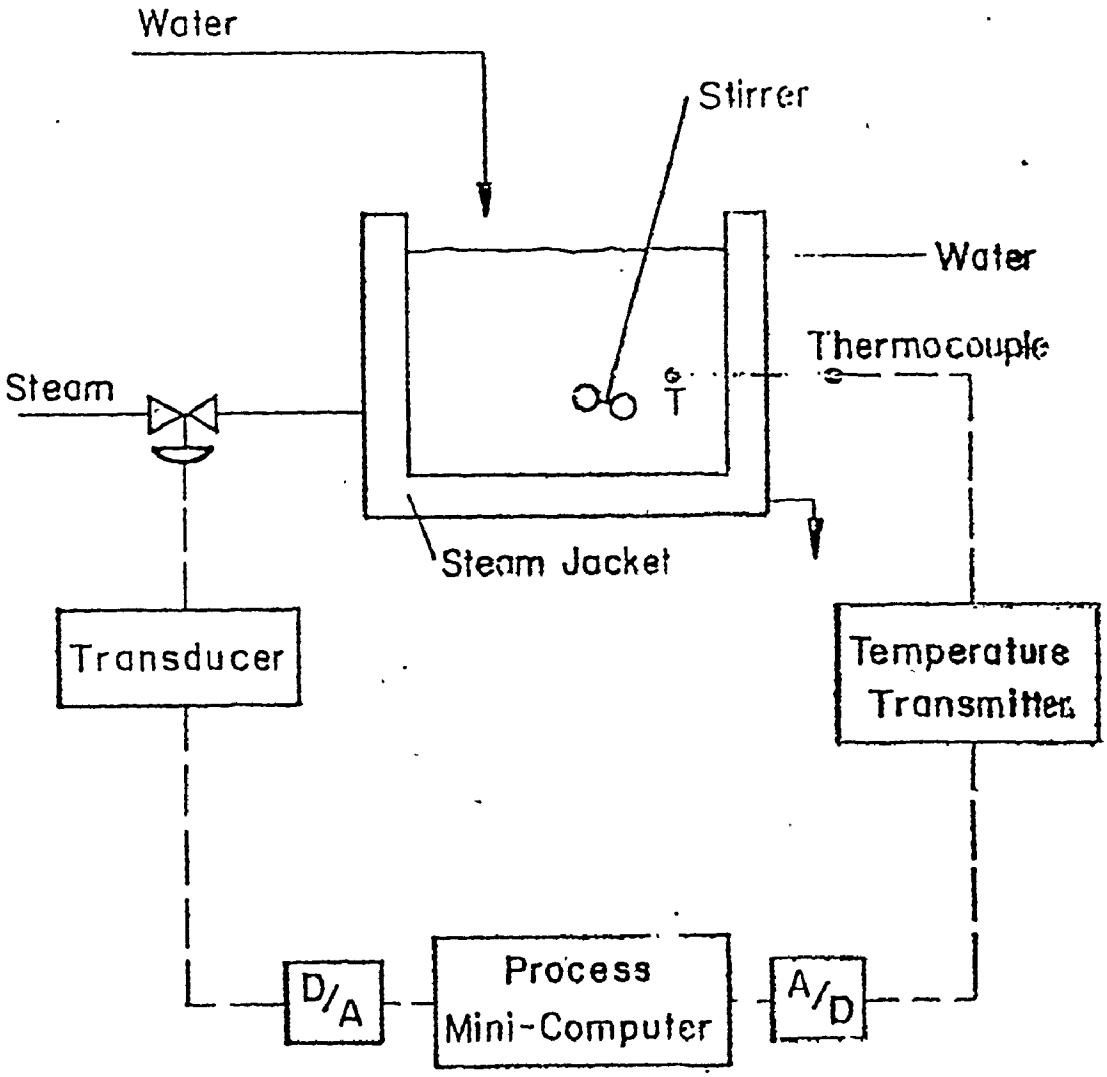
$$b = 0.829 \quad a_1 = 0.0148 \quad a_2 = 0.022.$$

These parameters, therefore, were used as the basis for comparison with the values obtained using the on-line methods.

It may be pointed out that since no noise was added externally, as in the simulated problems, it is difficult to state the noise-to-signal ratio.

### 5.1.3 The Continuous Stirred Tank Process

A schematic diagram of the process is shown in Figure 5.2. It consists of a steam-jacketed tank which is continuously stirred and interfaced to a remote Supernova mini-computer. Water flowing through the steam is being heated by low pressure steam in the jacket. The output of the system is the temperature of the water in the tank, which is controlled by regulating the steam flow to the tank jacket. Uncontrolled fluctuations in the inlet water flow and temperature as well as in the stream pressure have the effect of adding a random disturbance noise to the system. The temperature of the water in the tank is measured by a thermocouple and relayed to the A/D input of the mini-computer by a [illegible]. The steam flow to the




†Figure 5.2. Schematic Diagram of the continuous stirred tank reactor

---

†The Figure 5.2 has been reproduced from the report of MacGregor and Huynh (1974).

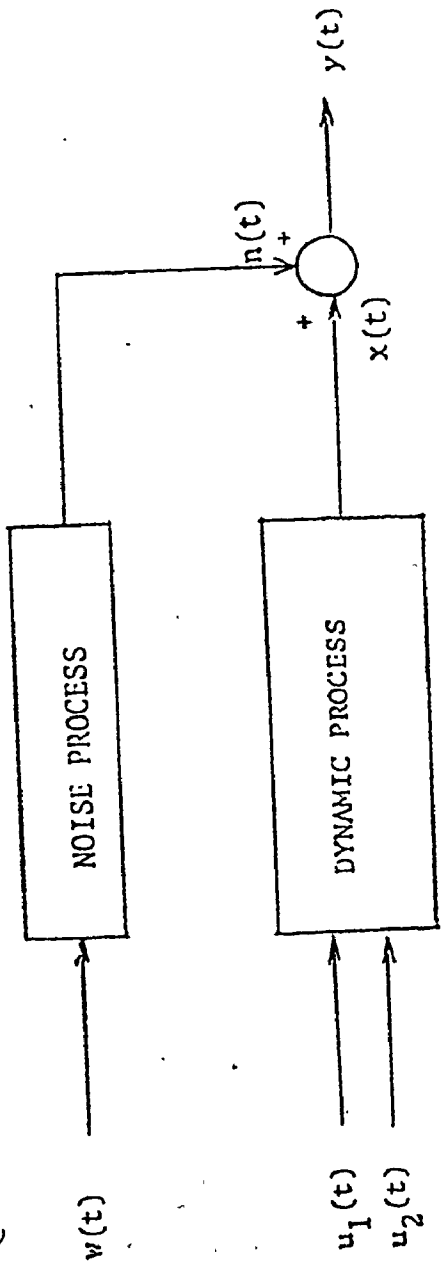



Figure 5.3 Block diagram of the dual input heat exchanger system

tank jacket is regulated by the voltage output from the D/A unit of the mini-computer to an electro-pneumatic transducer and a pneumatic control valve. The stream pressure upstream of the control valve was manually controlled at a nominal pressure of 80 p.s.i.g. For further details, reference may be made to the report by Hong and McGregor (1974) and to Hong's thesis (1974). Hong and McGregor (1974) using the time-series analysis of Box and Jenkins (1970) have determined that the system can be adequately represented within 95% confidence limits by the following difference equation model:

$$y_k = a_1 u_{k-1} - b_1 y_{k-1} \qquad (5.5)$$

where

$$a_1 = 0.162 \text{ and } b_1 = -0.873$$

These parameters were used as the basis for the comparison with the values obtained using on-line methods.

5.1.4. Dual-Input Heat-Exchanger System.

The same two stage heat exchanger discussed in section 5.1.2 was considered. For our purpose, the process may be considered as a system with two inputs and one output. The two input variables are the steam valve and the water valve positions, which control the rate of flow of steam and water, respectively. The output variable is the water temperature $T_3$, at the outlet from the second heat exchanger. In addition, there is an unwanted disturbance or noise input

due to random variations in the water supply pressure. The block diagram of the system is shown in Figure 5.3. The effect of the disturbance is to introduce an equivalent noise n(t) at the output, and this may be regarded as the output of a "noise process" subject to white Gaussian noise input. Wright and Bacon (1974) have determined the following discrete transfer function for the process using the time-series analysis method.

$$H_1(z^{-1}) = \frac{a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}}{1 + cz^{-1}} \qquad (5.6)$$

$$H_2(z^{-1}) = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + cz^{-1}} \qquad (5.7)$$

where

$a_1 = 0.00885$ $\quad a_2 = 0.00623$

$a_3 = -0.00283$ $\quad b_1 = -0.0295$

$b_2 = -0.0152$ $\quad c = -0.314$

In the absence of noise the output is given by:

$$C(z^{-1}) = H_1(z^{-1})\, U_1(z^{-1}) + H_2(z^{-1})\, U_2(z^{-1}) \qquad (5.8)$$

Since the output samples c(iT) are contaminated with noise we only have available:

$$y(iT) = c(iT) + n(iT) \qquad (5.9)$$

From (5.8) and (5.9) we have:

$$\underset{\sim}{y}_k = A_k \underset{\sim}{\phi} + \underset{\sim}{e}_k \tag{5.10}$$

where

$$\underset{\sim}{y}_k = [y(3T),\ y(4T) \ldots y(kT + T)]^T \tag{5.11}$$

$$\underset{\sim}{e}_k = [e(3T),\ e(4T) \ldots e(kT)]^T \tag{5.12}$$

$$e(kT) = n(kT) + c\, n(kT-T) \tag{5.13}$$

$$A_k = \begin{bmatrix} u_1(2T) & u_1(T) & u_1(0) & u_2(2T) & u_2(T) & -y(2T) \\ u_1(3T) & u_1(2T) & u_1(T) & u_2(3T) & u_2(2T) & -y(3T) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ u_1(kT) & u_1(kT-T) & u_1(kT-2T) & u_2(kT) & u_2(kT-T) & -y(kT) \end{bmatrix} \tag{5.14}$$

and

$$\underset{\sim}{\phi} = [a_1\ a_2\ a_3\ b_1\ b_2\ c]^T \tag{5.15}$$

Equation (5.15) is of exactly the same form as equation (4.8) of chapter 4, although the latter was derived from a single-input, single-output system. Hence the various algorithms given there can be directly used for the present problem.

## 5.2. Results and Comparison of Methods:

The results obtained using various methods of identification will now be presented for each case.

5.2.1. Second-order Process with White Measurement Noise.

For this case, 500 samples of input-output data were used for recursively estimating the four parameters of the system for three different values of noise-to-signal ratio. The normalised estimation error, defined as:

$$\frac{||\underset{\sim}{\phi} - \underset{\sim}{\hat{\phi}}_k||^2}{||\underset{\sim}{\phi}||^2}$$

has been plotted against the number of iterations k, in figures 5.4, 5.5, and 5.6, for noise to signal ratios of 20%, 40% and 100% respectively. The final values of the parameters, obtained at the end of 500 iterations, are shown in Table 5.1.

It will be seen although all the methods give reasonably good results for low noise, when the noise-to-signal ratio is increased to about 100%, the correlation method, the instrumental variable method and the combined pseudoinverse and stochastic approximation method only give good results. Of these three methods, the combined pseudo-inverse and stochastic approximation algorithm appears to have fastest rate of convergence.

5.2.2. Second-order Process with Coloured Measurement Noise.

The computations were repeated with 500 samples of input-output data with coloured output noise. The normalized

estimation error

$$\frac{||\underset{\sim}{\phi} - \underset{\sim}{\hat{\phi}}||^2}{||\underset{\sim}{\phi}||^2}$$

has been plotted against the number of iterations in figures 5.7, 5.8, and 5.9, for noise-to-signal ratios of 20%, 40% and 100% respectively. The final values of the estimates of the parameters, obtained at the end of 500 iterations, are shown in Table 5.2.

In this case again, all the methods work well for the low noise case, but for 100% noise-to-signal ratio, good results are obtained only with correlation method, the instrumental variable method, the generalised pseudoinverse method, and the combined pseudoinverse and stochastic approximation algorithm. Of these the combined pseudoinverse and stochastic approximation algorithm appears to have the fastest rate of convergence.

5.2.3. The Two-Stage Heat-Exchanger Process.

In this case, 300 pairs of input-output data, obtained at sampling intervals of 5 seconds, were used for recursive estimation of the three parameters of the model. These estimates, along with the parameter values obtained using the time series method are shown in Table 5.3.

The results indicate considerable bias in the estimates obtained using the pseudoinverse method. Of all the methods considered, that combining pseudoinverse and stochastic

approximation appears most promising. It gives better estimates in fewer iterations.

5.2.4. The Continuous Stirred Tank Process.

Using 100 samples of the input-output data obtained experimentally, the parameters were estimated recursively. Table 5.4 gives the estimates after different number of iterations. Figure 5.10 shows the plot of the normalized error

$$\frac{||\underset{\sim}{\phi} - \underset{\sim}{\hat{\phi}}||^2}{||\underset{\sim}{\phi}||^2}$$

against the number of iterations.

It will be seen that pseudoinverse and stochastic approximation methods give estimates which are slightly biased. The combined pseudoinverse and stochastic approximation algorithm gives the best results.

5.2.5. The Dual-Input Heat Exchanger Process.

In this case, 280 samples of input-output data, obtained experimentally, are used for recursive estimation of parameters. A plot of the normalized error

$$\frac{||\underset{\sim}{\phi} - \underset{\sim}{\hat{\phi}}||^2}{||\underset{\sim}{\phi}||^2}$$

against the number of iterations is shown in figure 5.11. The final estimates of the parameters as obtained after 280 iterations are shown in Table 5.5.

It will be seen that the best results are obtained by the algorithm combining pseudoinverse and stochastic approximation and the generalized pseudoinverse algorithm gives the next best rate of convergence. Both of these algorithms show the estimates converging to their final values in less than 140 iterations.

Table (5.6) shows the total number of arithmetic operations (addition, substraction, multiplication and division) required per iteration of different algorithms. The number p denotes the total number of parameters of the process model and s denotes the number of parameters of the corresponding noise models. Table (5.6) also summarizes the execution time required by different algorithms (based on CDC 6400 computer and programmed in Fortran) for 300 iterations in the case of the two-stage heat exchanger system.

5.2.6. Discussions.

Different methods for on-line identification have been compared for four test cases. For the simulated second order process, three different noise-to-signal ratios have been used with white measurement noise as well as coloured measurement noise. The other test cases, a two stage heat exchanger system and a continuous stirred tank reactor, have sufficient inherent noise to make the problem realistic. On the basis of the comparison, the following conclusions may be formed.

(a) The pseudoinverse algorithm is most efficient for low-noise level. Not only does it require least time for computation (with the exception of stochastic approximation method), but estimates converge to the correct value very fast.

(b) The stochastic approximation requires the least amount of computation per iteration, but it has a very slow rate of convergence.

(c) The correlation method is quite efficient even for large noise, but one must wait for sufficient amount of data to get reasonable initial estimates of the correlation ordinates. It is a good "off-line" method, but its usefulness as an "on-line" method may be questionable.

(d) All the other methods considered try to develop a model for the noise parameters in an on-line manner. Although the estimates obtained are satisfactory with most of them, it was found that the method using pseudoinverse for the model parameters and stochastic approximation for the auxiliary noise parameters yields faster convergence and better accuracy. This is achieved at the cost of increased computational effort compared to other methods (except generalised pseudoinverse algorithm). This proposed method is found to be computationally more efficient relative to the recursive version of generalised least squares algorithm. The main reason for this is the incorporation of much simpler stochastic approximation algorithm to estimate the

parameters of the noise model. This inclusion of noise model, similar to the case of generalized least squares, tends to remove the bias of the estimates generated by the pseudoinverse algorithm. When operating with a finite data set, an effective estimator should not only converge to the true parameter values but does so at a rate that is faster than any other estimator. Moreover, for any identification scheme to be of any practical value it should determine the parameters as quickly as possible. For all the cases considered the algorithm combining pseudoinverse and stochastic approximation yields faster convergence and better accuracy with some increase in computational effort.

Finally, it should be added that no theoretical justification has yet been offered for the observed improvement in convergence obtained with the proposed method. But simulation results indicate the method to be quite promising. The theoretical aspect of convergence of this algorithm needs to be further explored.

| Method | Noise to signal ratio | $\hat{a}_1$ | $\hat{a}_2$ | $\hat{b}_1$ | $\hat{b}_2$ |
|---|---|---|---|---|---|
| Pseudoinverse | 20% | 0.0790 | 0.0750 | -0.6810 | -0.0310 |
| | 40% | 0.0780 | 0.0890 | -0.4530 | -0.2020 |
| | 100% | 0.0800 | 0.1080 | -0.2410 | -0.2270 |
| Generalized | 20% | 0.0630 | 0.0550 | -0.7160 | 0.1810 |
| Pseudoinverse | 40% | 0.0410 | 0.0520 | -0.4530 | 0.3090 |
| | 100% | 0.1390 | 0.0450 | -0.1330 | 0.5610 |
| Instrumental | 20% | 0.0780 | 0.0580 | -0.8520 | 0.1150 |
| Variable | 40% | 0.0790 | 0.0610 | -0.8310 | 0.1000 |
| | 100% | 0.0800 | 0.0630 | -0.8290 | 0.0970 |
| Pandya's | 20% | 0.0790 | 0.0590 | -0.8200 | 0.0950 |
| Bootstrap | 40% | 0.0770 | 0.0770 | -0.6030 | -0.0820 |
| Estimator | 100% | 0.0820 | 0.1020 | -0.3110 | -0.2550 |
| Correlation | 20% | 0.0710 | 0.0405 | -0.8820 | 0.3110 |
| Method | 40% | 0.0710 | 0.0480 | -0.8110 | 0.2650 |
| | 100% | 0.0703 | 0.0663 | -0.6520 | 0.1860 |
| Ordinary | 20% | 0.0790 | 0.0520 | -0.8950 | 0.1550 |
| Stochastic | 40% | 0.0820 | 0.0540 | -0.8830 | 0.1540 |
| Approximation | 100% | 0.0870 | 0.0580 | -0.8480 | 0.1530 |
| Combined Pseudo- | 20% | 0.0810 | 0.0470 | -0.9610 | 0.2110 |
| inverse&stochastic | 40% | 0.0820 | 0.0460 | -0.9560 | 0.2070 |
| Approximation Alg. | 100% | 0.0850 | 0.0560 | -0.8690 | 0.1640 |
| True Values | | 0.0790 | 0.0470 | -0.9750 | 0.2230 |

Table 5.1: Final estimates of parameters of the second-order system for white measurement noise.

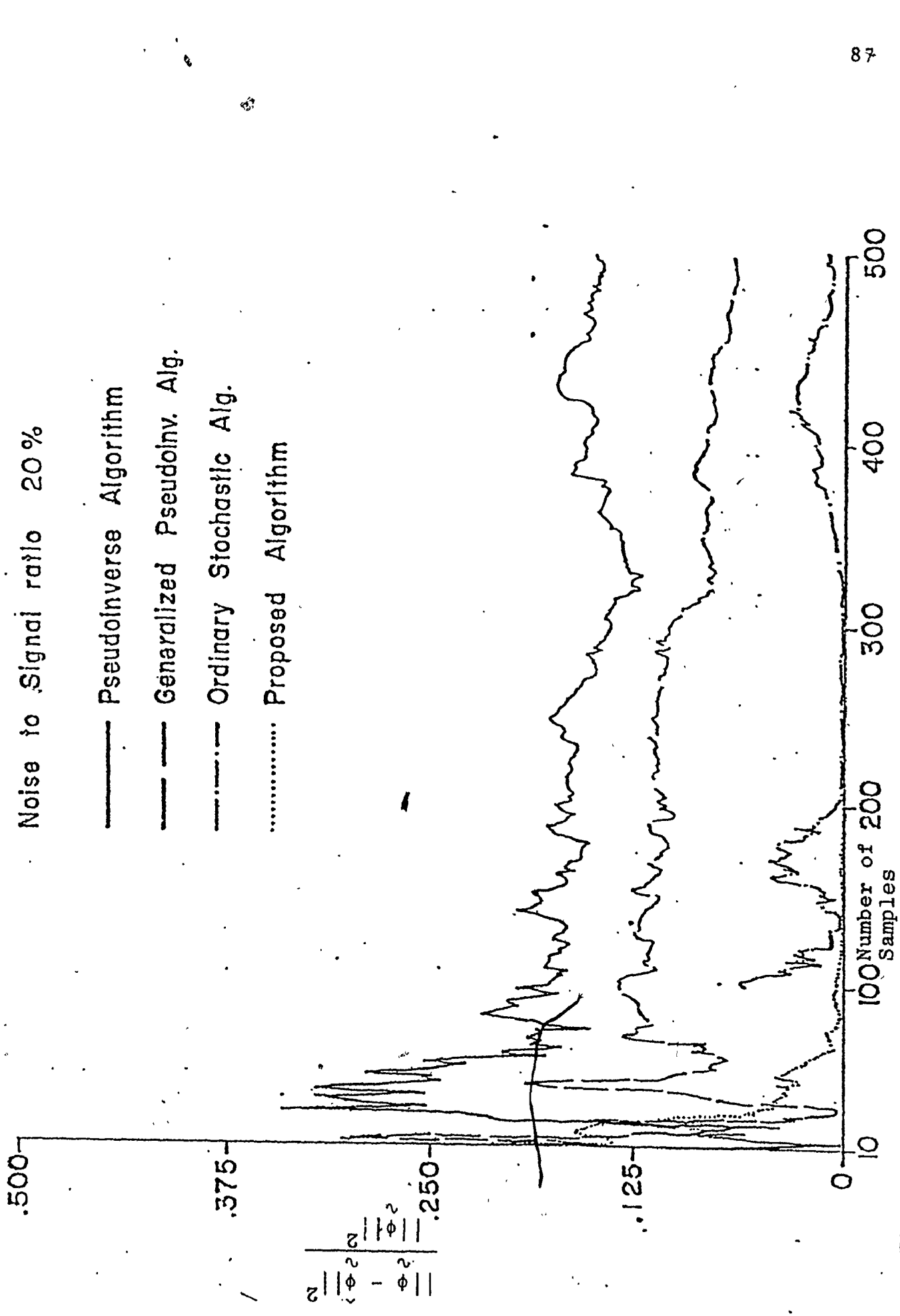



Figure 5.4. Convergence rates for 20% noise-to-signal ratio (white measurement noise)

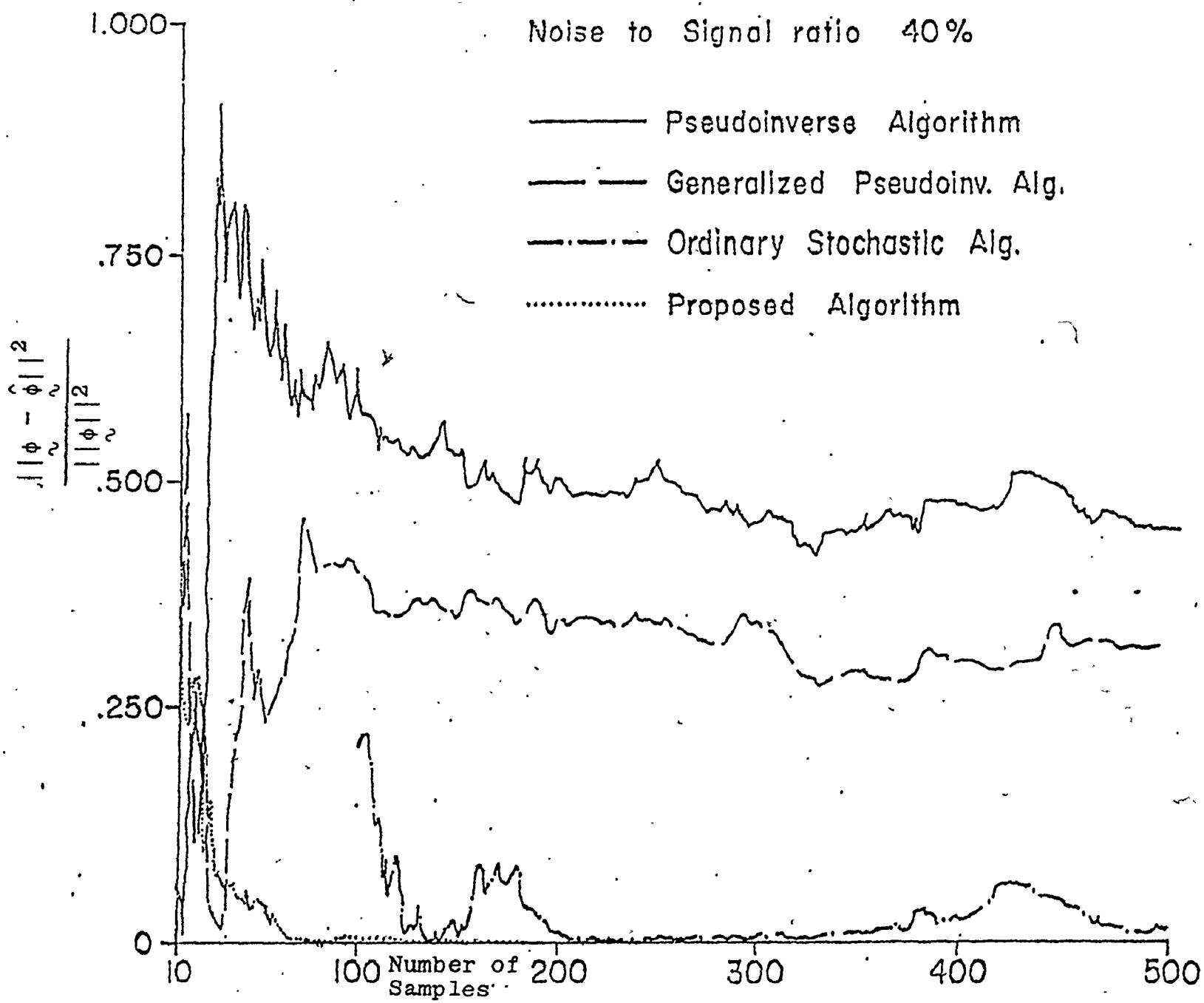



Figure 5.5. Convergence rates for 40% noise-to-signal ratio (white measurement noise)

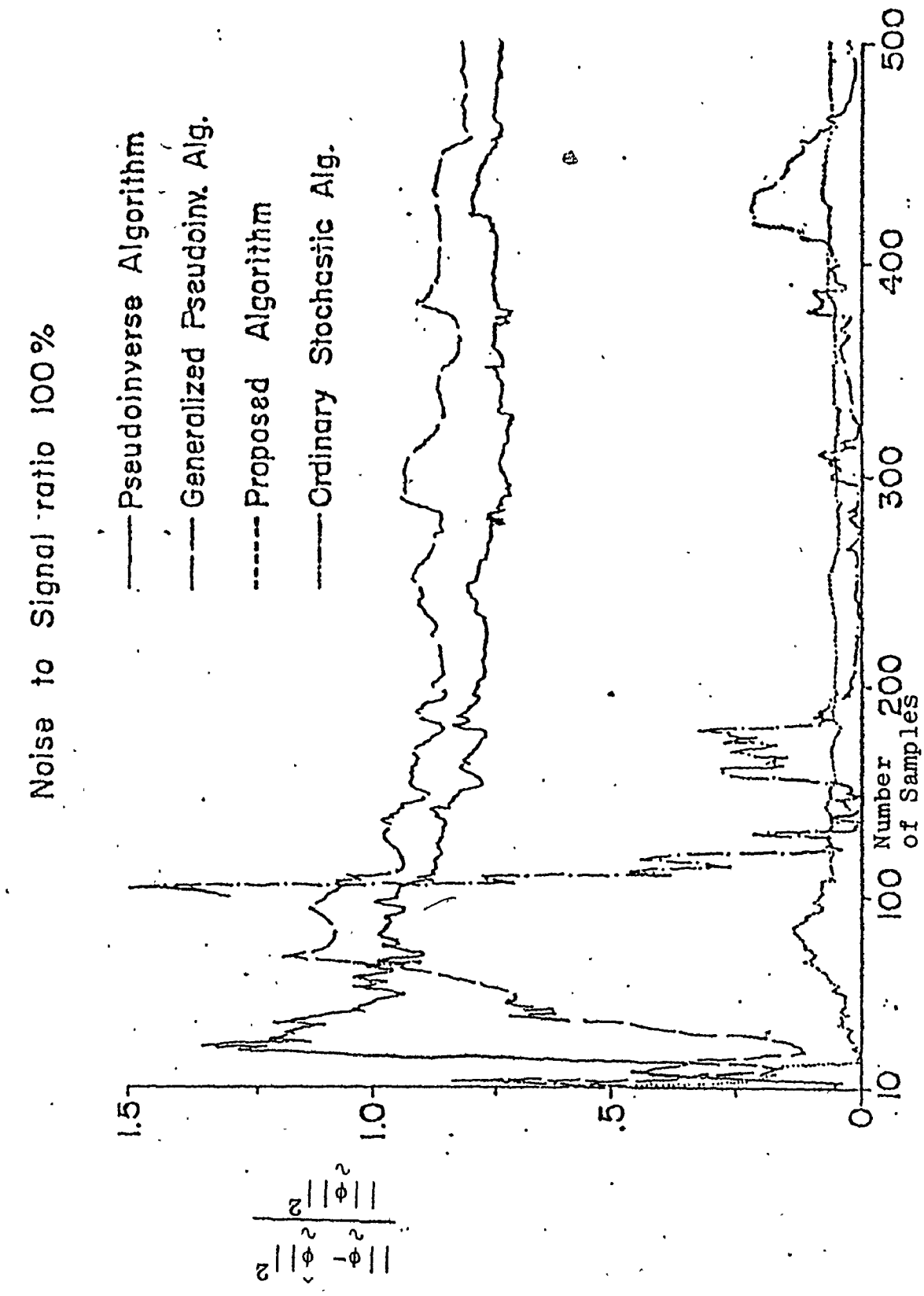



Figure 5.6. Convergence rates for 100% noise-to-signal ratio (white measurement noise)

| Method | Noise to signal ratio | $\hat{a}_1$ | $\hat{a}_2$ | $\hat{b}_1$ | $\hat{b}_2$ |
|---|---|---|---|---|---|
| Pseudoinverse | 20% | 0.0772 | 0.0483 | -0.9440 | 0.1910 |
| | 40% | 0.0756 | 0.0514 | -0.8900 | 0.1330 |
| | 100% | 0.0716 | 0.0543 | -0.8040 | 0.0121 |
| Generalized Pseudoinverse | 20% | 0.0775 | 0.0445 | -0.9820 | 0.2350 |
| | 40% | 0.0775 | 0.0423 | -0.9880 | 0.2530 |
| | 100% | 0.0695 | 0.0385 | -0.9790 | 0.2750 |
| Instrumental Variable | 20% | 0.0772 | 0.0475 | -0.9540 | 0.2030 |
| | 40% | 0.0755 | 0.0470 | -0.9510 | 0.1980 |
| | 100% | 0.0707 | 0.0452 | -0.9250 | 0.1710 |
| Pandya's Bootstrap Estimator | 20% | 0.0772 | 0.0490 | -0.9340 | 0.1850 |
| | 40% | 0.0750 | 0.0497 | -0.9110 | 0.1580 |
| | 100% | 0.0713 | 0.0523 | -0.8310 | 0.0540 |
| Correlation Method | 20% | 0.0705 | 0.0321 | -0.9520 | 0.3540 |
| | 40% | 0.0690 | 0.0325 | -0.9350 | 0.3360 |
| | 100% | 0.0640 | 0.0336 | -0.8790 | 0.2750 |
| Ordinary Stochastic Approximation | 20% | 0.0790 | 0.0520 | -0.8950 | 0.1550 |
| | 40% | 0.0820 | 0.0540 | -0.8830 | 0.1540 |
| | 100% | 0.0870 | 0.0580 | -0.8480 | 0.1530 |
| Combined Pseudoinverse&Stochastic Approximation Alg. | 20% | 0.0771 | 0.0459 | -0.9740 | 0.2201 |
| | 40% | 0.0765 | 0.0447 | -0.9750 | 0.2158 |
| | 100% | 0.0729 | 0.0418 | -0.9660 | 0.1776 |
| True Values | | 0.0790 | 0.0470 | -0.9750 | 0.2230 |

Table 5.2 Final estimates of the parameters of the second-order process for coloured measurement noise.

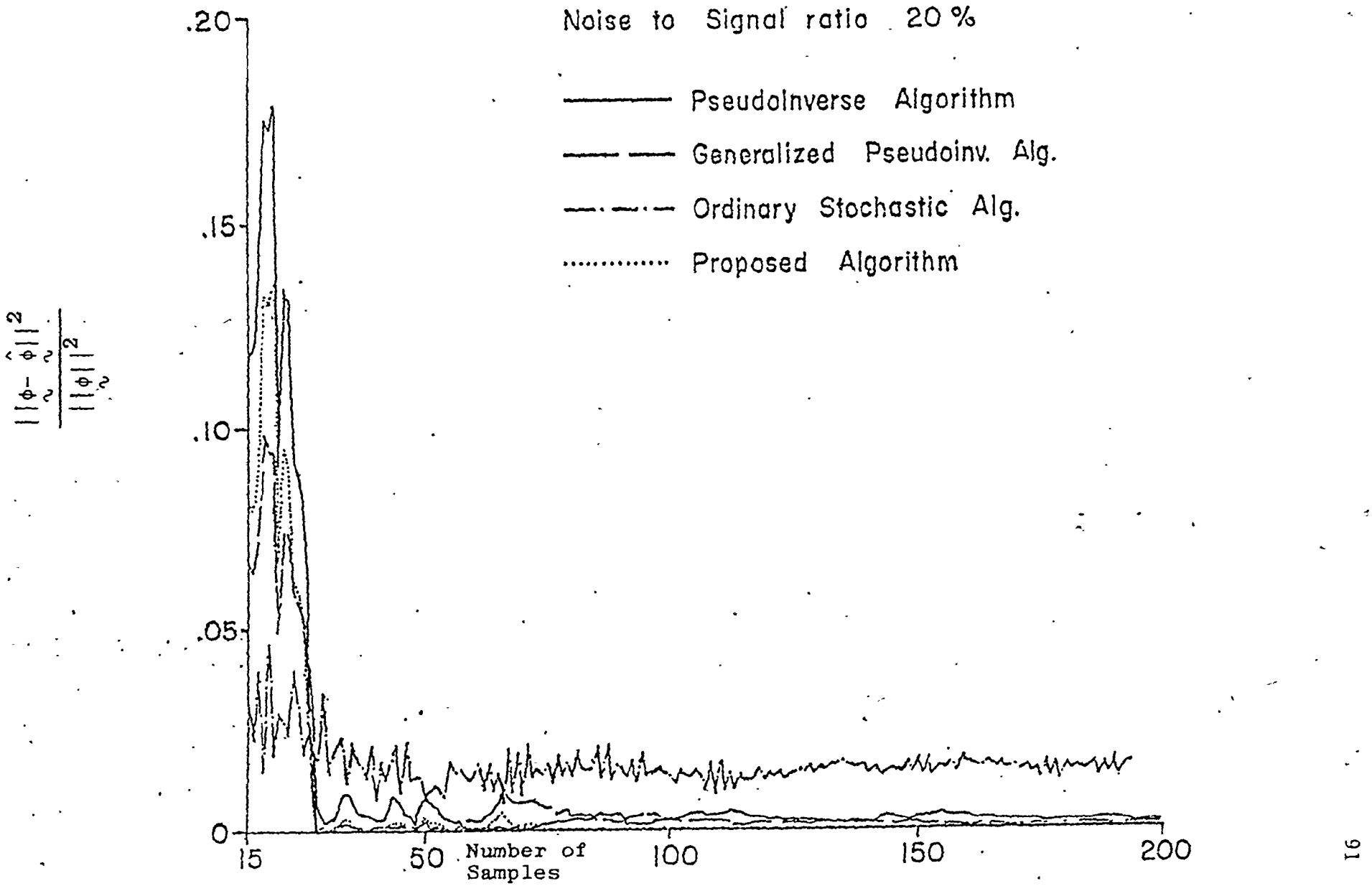



Figure 5.7. Convergence rates for 20% noise-to-signal ratio (coloured measurement noise)

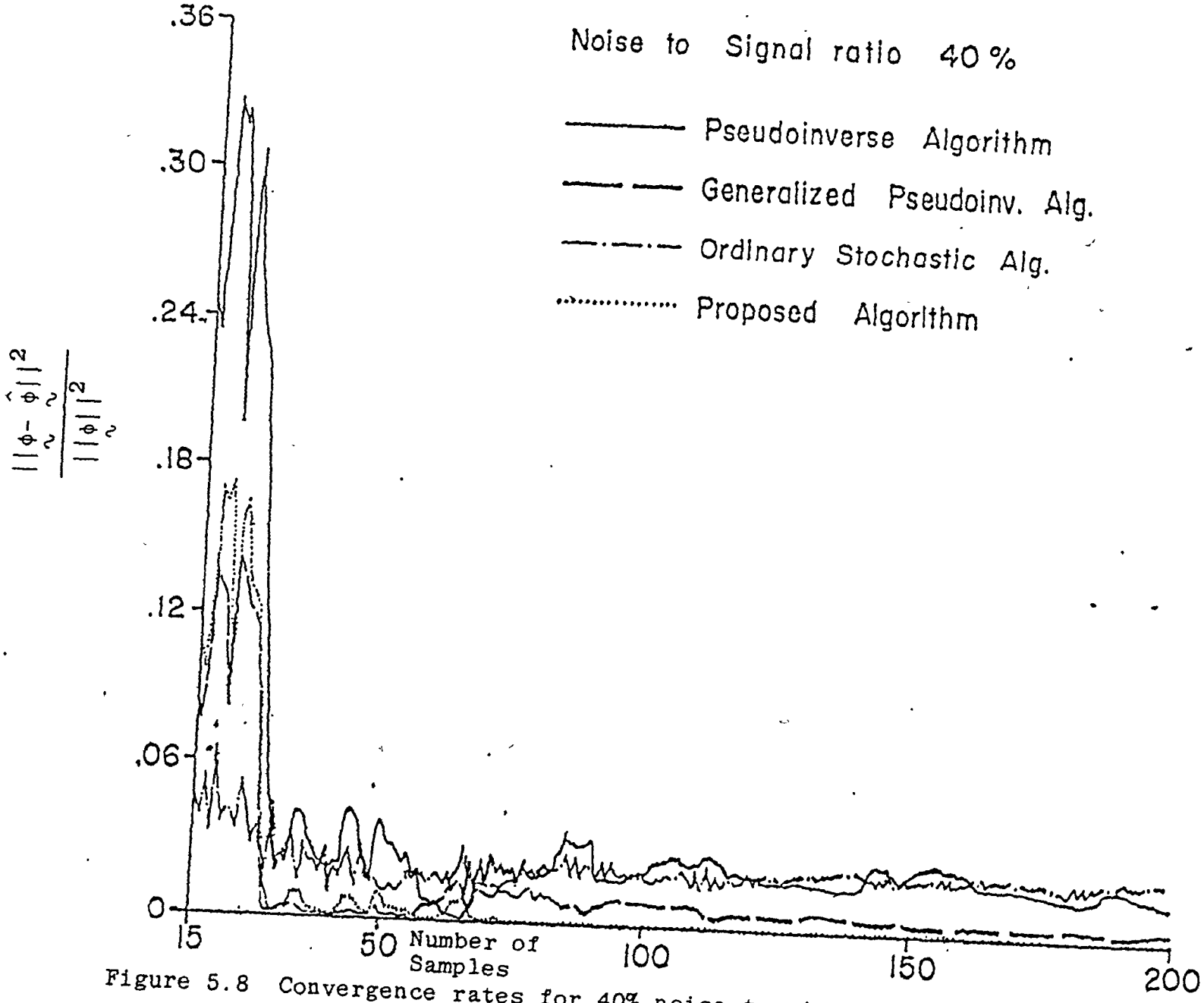



Figure 5.8 Convergence rates for 40% noise-to-signal ratio (coloured measurement noise)

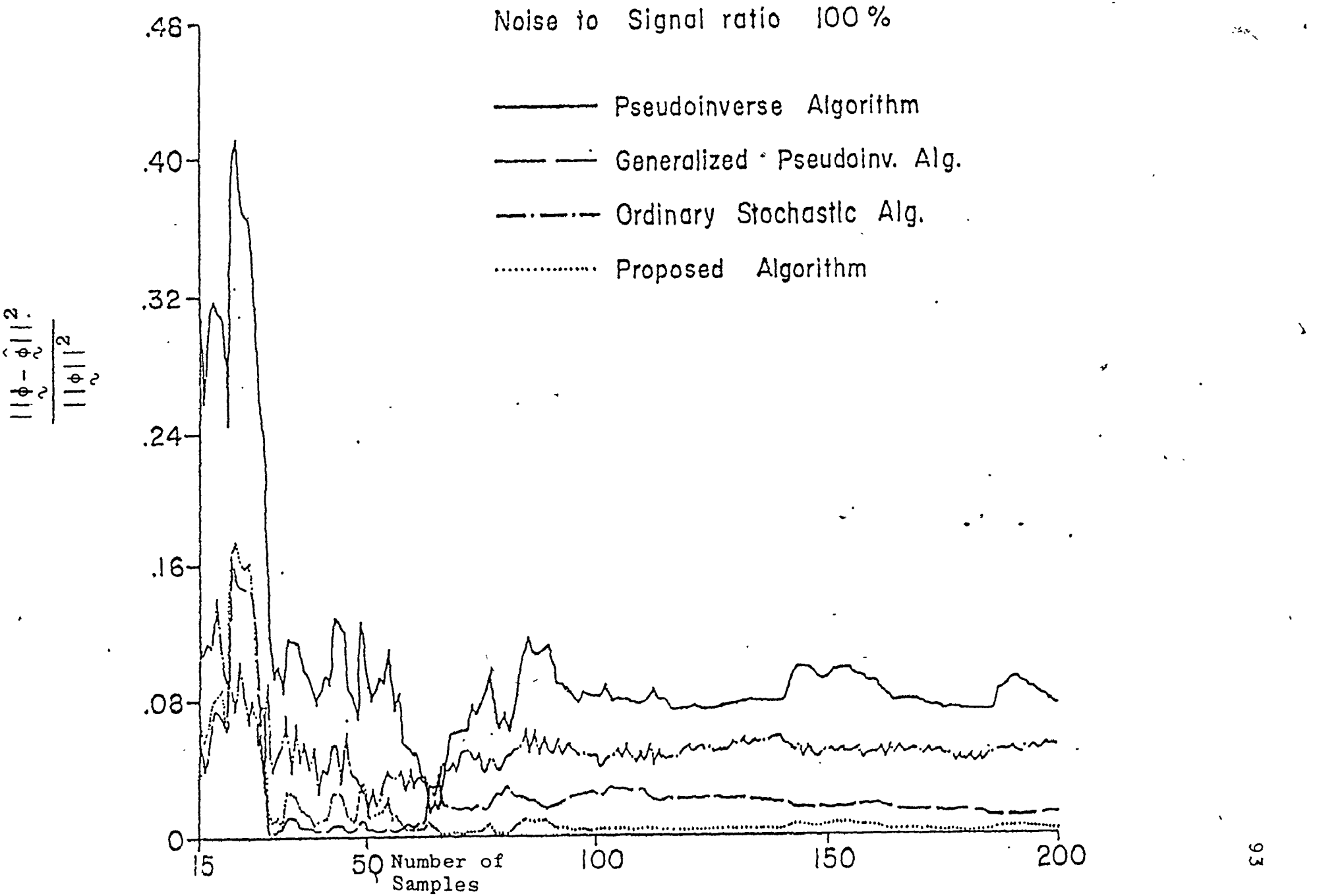



Figure 5.9. Convergence rates for 100% noise-to-signal ratio (coloured measurement noise)

| Method | Parameter | Estimates after different numbers of iterations | | | | | |
|---|---|---|---|---|---|---|---|
| | | 50 | 100 | 150 | 200 | 250 | 300 |
| Pseudoinverse | $\hat{a}_1$ | 0.0103 | 0.0179 | 0.0186 | 0.0162 | 0.0156 | 0.0154 |
| | $\hat{a}_2$ | 0.0118 | 0.0227 | 0.0234 | 0.0222 | 0.0239 | 0.0239 |
| | $\hat{b}$ | 0.8754 | 0.8525 | 0.8526 | 0.8574 | 0.8554 | 0.8582 |
| Generalized Pseudoinverse | $\hat{a}_1$ | 0.0107 | 0.0182 | 0.0178 | 0.0156 | 0.0149 | 0.0148 |
| | $\hat{a}_2$ | 0.0089 | 0.0297 | 0.0216 | 0.0205 | 0.0225 | 0.0228 |
| | $\hat{b}$ | 0.8184 | 0.8186 | 0.8218 | 0.8266 | 0.8193 | 0.8202 |
| Instrumental Variables | $\hat{a}_1$ | 0.0006 | 0.0100 | 0.0129 | 0.0122 | 0.0124 | 0.0125 |
| | $\hat{a}_2$ | 0.0140 | 0.0229 | 0.0228 | 0.0221 | 0.0241 | 0.0240 |
| | $\hat{b}$ | 0.7604 | 0.7973 | 0.8285 | 0.8030 | 0.8502 | 0.8600 |
| Pandya's Bootstrap Estimator | $\hat{a}_1$ | 0.0147 | 0.0192 | 0.0195 | 0.0169 | 0.0161 | 0.0157 |
| | $\hat{a}_2$ | 0.0126 | 0.0231 | 0.0237 | 0.0227 | 0.0245 | 0.0243 |
| | $\hat{b}$ | 0.7870 | 0.7887 | 0.8153 | 0.8165 | 0.8215 | 0.8296 |
| Ordinary Stochastic Approximation | $\hat{a}_1$ | 0.0169 | 0.0168 | 0.0169 | 0.0169 | 0.0168 | 0.0167 |
| | $\hat{a}_2$ | 0.0142 | 0.0143 | 0.0143 | 0.0143 | 0.0143 | 0.0143 |
| | $\hat{b}$ | 0.8468 | 0.8471 | 0.8469 | 0.8470 | 0.8484 | 0.8490 |
| Combined Pseudo-inverse&Stochastic Approximation Alg. | $\hat{a}_1$ | 0.0116 | 0.0183 | 0.0177 | 0.0152 | 0.0148 | 0.0148 |
| | $\hat{a}_2$ | 0.0104 | 0.0199 | 0.0209 | 0.0196 | 0.0220 | 0.0225 |
| | $\hat{b}$ | 0.7233 | 0.7692 | 0.7845 | 0.7915 | 0.7818 | 0.7805 |
| Time-series Method | | | | | | | 0.0148<br>0.022<br>0.829 |

Table 5.3. Estimates of the parameters of a two-stage heat-exchanger system.

| Method | Parameter | Estimates after different number of iterations | | | | |
|---|---|---|---|---|---|---|
| | | 20 | 40 | 60 | 80 | 100 |
| Pseudoinverse | $\hat{a}_1$ | 0.179 | 0.167 | 0.178 | 0.169 | 0.169 |
| | $\hat{b}_1$ | -0.895 | -0.904 | -0.917 | -0.908 | -0.894 |
| Generalized Pseudoinverse | $\hat{a}_1$ | 0.178 | 0.153 | 0.166 | 0.157 | 0.164 |
| | $\hat{b}_1$ | -0.807 | -0.842 | -0.867 | -0.862 | -0.861 |
| Instrumental Variables | $\hat{a}_1$ | 0.018 | 0.157 | 0.178 | 0.169 | 0.174 |
| | $\hat{b}_1$ | -1.114 | -0.917 | -0.937 | -0.900 | -0.862 |
| Pandya's Bootstrap Estimator | $\hat{a}_1$ | 0.018 | 0.167 | 0.181 | 0.167 | 0.171 |
| | $\hat{b}_1$ | -0.873 | -0.901 | -0.908 | -0.897 | -0.884 |
| Ordinary Stochastic Approximation | $\hat{a}_1$ | 0.107 | 0.108 | 0.108 | 0.108 | 0.108 |
| | $\hat{b}_1$ | -0.870 | -0.868 | -0.868 | -0.877 | -0.875 |
| Combined Pseudo-inverse & Stochastic Approximation Alg. | $\hat{a}_1$ | 0.172 | 0.154 | 0.168 | 0.164 | 0.164 |
| | $\hat{b}_1$ | -0.834 | -0.866 | -0.876 | -0.873 | -0.873 |
| Time Series Method | | | | | | 0.162<br>-0.873 |

Table 5.4 Estimates of the parameters of a continuous stirred tank reactor.

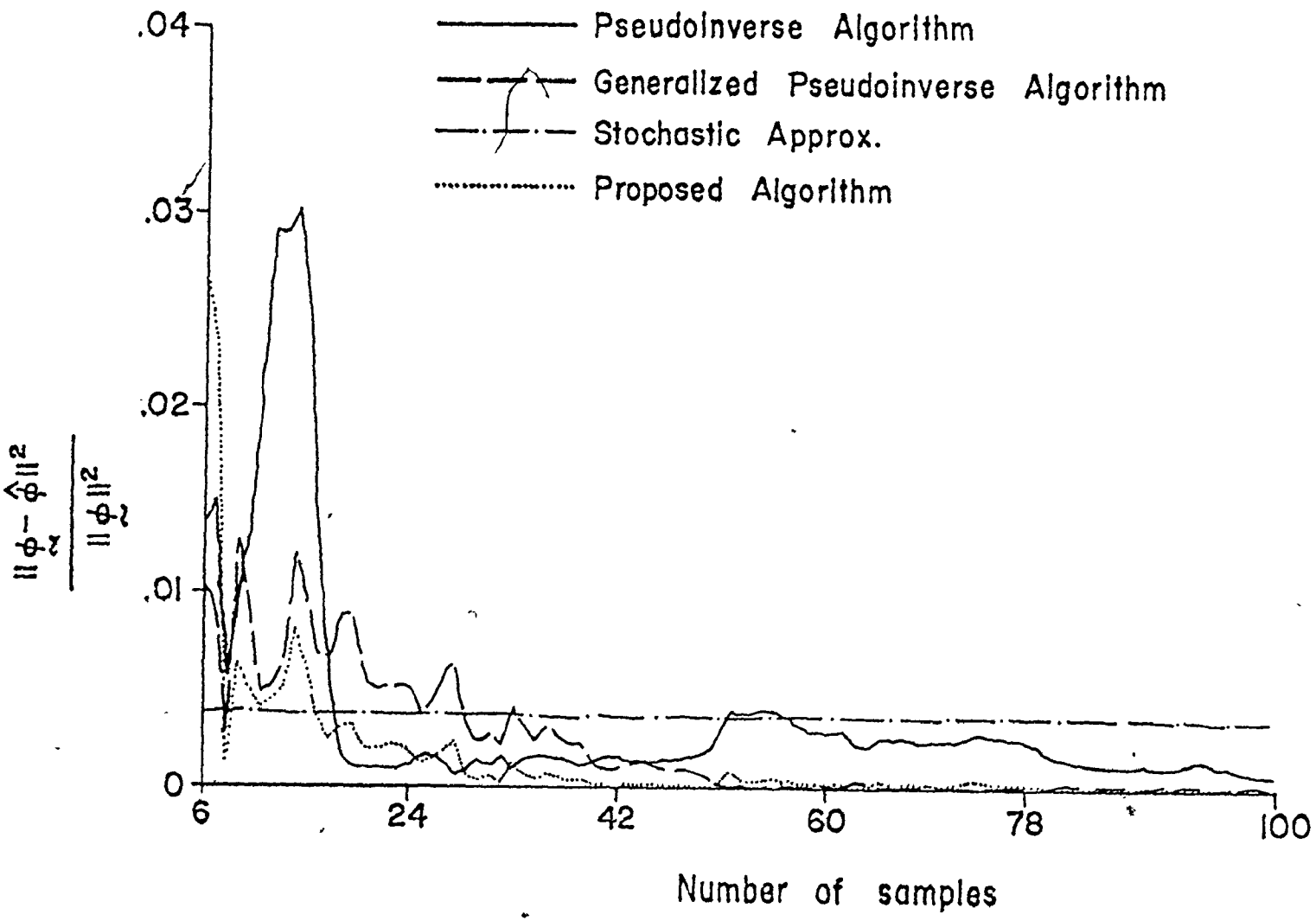



Figure 5.10. Convergence rate for continuous stirred tank reactor

| Method | $\hat{a}_1$ | $\hat{a}_2$ | $\hat{a}_3$ | $\hat{b}_1$ | $\hat{b}_2$ | $\hat{c}$ |
|---|---|---|---|---|---|---|
| Pseudoinverse | 0.009120 | 0.007396 | 0.000900 | -0.029500 | 0.012200 | -0.361000 |
| Generalized Pseudoinverse | 0.008660 | 0.006970 | 0.001336 | -0.029500 | -0.013300 | -0.323000 |
| Instrumental Variable | 0.009020 | 0.005650 | -0.001310 | -0.029650 | -0.011670 | -0.371000 |
| Pandya's Bootstrap Estimator | 0.009050 | 0.007150 | 0.000760 | -0.029650 | -0.011690 | -0.383000 |
| Ordinary Stochastic Approximation | 0.006700 | 0.007610 | -0.002420 | -0.029790 | -0.014030 | -0.259800 |
| Combined Pseudoinverse Stochastic Approximation | 0.008310 | 0.006650 | -0.001400 | -0.029590 | -0.013480 | -0.319800 |
| Values using Time Series | 0.008850 | 0.006230 | -0.002830 | -0.029500 | -0.015200 | -0.314000 |

Table 5.5: Final estimates of parameters of a dual input heat-exchanger system.

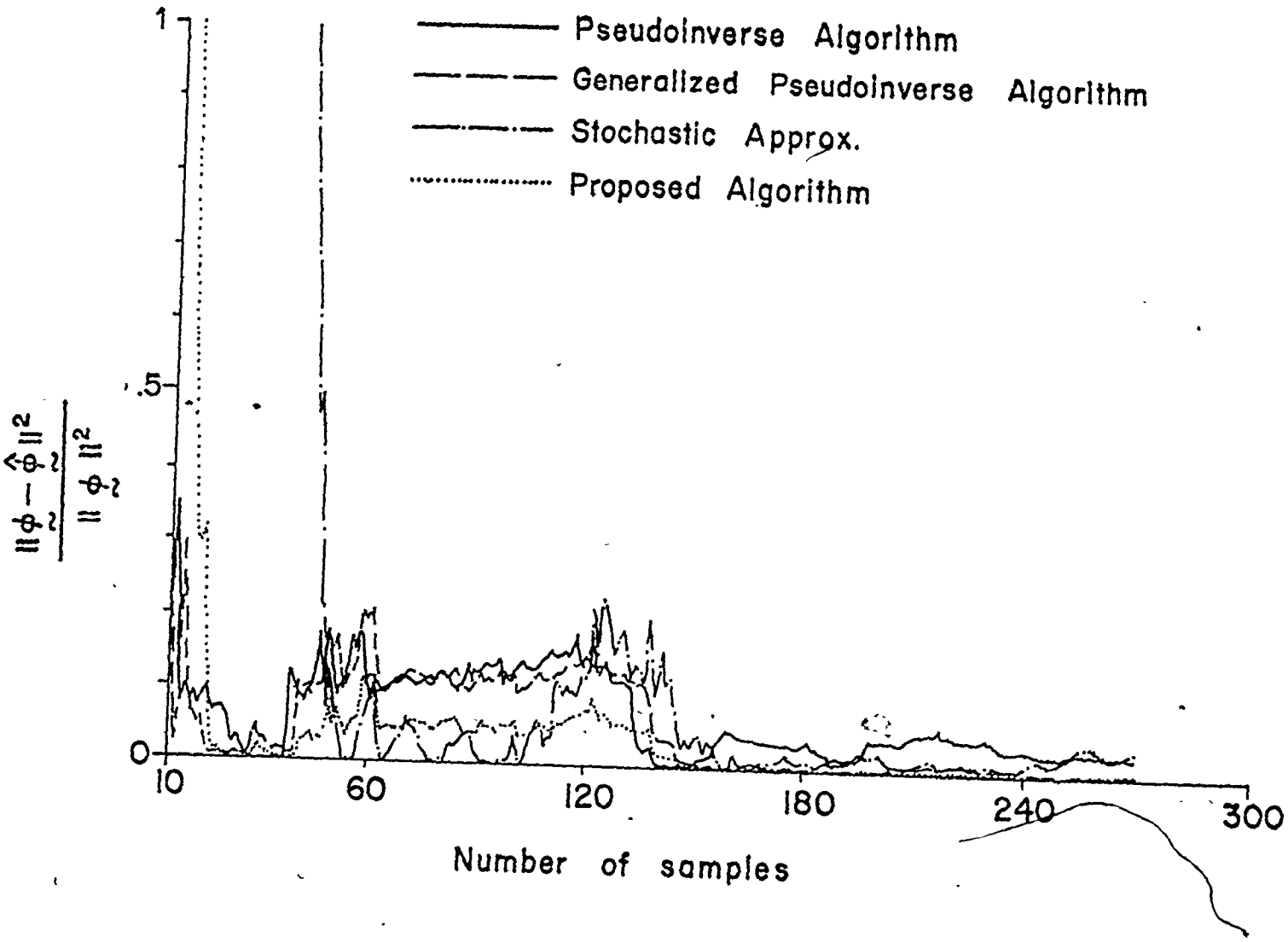



Figure 5.11. Convergence rates for dual input heat exchanger system

| Methods | Total number of additions and subtractions per iteration | Total number of multiplication and division per iteration | Total number of arithmetic operation per iteration | Total computing time for 300 iterations for a two stage heat exchanger system in CDC 6400 computer |
|---|---|---|---|---|
| Pseudoinverse | $2p^2+2p$ | $2p^2+4p$ | $4p^2+6p$ | 1.7 |
| Generalized Pseudoinverse | $2p^2+3p$ $+4s+2s^2$ | $2p^2+5p$ $+6s+2s^2$ | $4p^2+8p$ $+8p+10s$ | 7 |
| Instrumental Variables | $2p^2+4p$ | $2p^2+7p$ | $4p^2+11p$ | 2.4 |
| Pandya's Bootstrap Estimator | $3p^2+2p+1$ | $3p^2+5p+1$ | $6p^2+7p+1$ | 2.5 |
| Ordinary Stochastic approximation | $2p-1$ | $3p$ | $5p-1$ | 1.2 |
| Combined Pseudoinverse and Stochastic approximation algorithm | $2p^2+3p$ $+5s-1$ | $2p^2+5p$ $+5s+3$ | $4p^2+8p$ $+10s+2$ | 5.8 |

Table 5.6: Comparison of computation time and effortrequired for different methods

# CHAPTER 6

# ON-LINE IDENTIFICATION OF MULTIVARIABLE SYSTEM

## 6.1. Introduction.

Linear multivariable time-invariant dynamic systems may be described either by transfer function matrix H(z) or in state space by the triple {A,B,C} where:

$$S1: \quad \underset{\sim}{x}(k+1) = A\,\underset{\sim}{x}(k) + B\,\underset{\sim}{u}(k) \tag{6.1}$$

$$\underset{\sim}{y}(k) = C\,\underset{\sim}{x}(k) \tag{6.2}$$

where $\underset{\sim}{x}(k)$, $\underset{\sim}{u}(k)$ and $\underset{\sim}{y}(k)$ are n, p, and m dimensional state, control and output vectors, while A, B and C are matrices of compatible dimensions. The transfer function matrix H(z) of (S1) may be derived directly from {A,B,C} as:

$$H(z) = C(zI-A)^{-1}B \tag{6.3}$$

On the other hand, for a given H(z) the triple {A,B,C} is not defined uniquely. The problem of minimal realisation of the transfer function matrix to obtain the triple {A,B,C} for minimum n is one of the fundamental problems in linear system theory. Many canonical forms for the minimal realisation of finite di-

mensional multi-input multi-output linear systems have been developed. Several results exposing the interrelations between state variable and transfer function descriptions of multi-input multi-output system have been recently presented by Dickinson et al (1974). The problem of minimal realisation has been studied, among others by Ho and Kalman (1965), Silverman (1971), Chen and Mital (1972) and Sinha and Rózsa (1974).

Much of the work on system identification reported so far concentrates on obtaining a state space model from the given input-output data. Gopinath (1969) presented the first clear picture for the realisation of multi-input multi-output discrete systems directly from input-output observations. In particular a procedure was presented to determine the triple {A,B,C}. However, this procedure requires an extensive search and it neglects the structure of resulting realisation. Subsequently Gopinath's algorithm was improved computationally by Budin (1971). Recently Passeri and Hergert (1972) also considered this problem and proposed minor modifications. Djorovic and Bingulac (1972) have considered a method of constructing a minimal realisation of general mutivariable system directly from noisy input-output observations. However, their algorithm is non-iterative in nature and the minimal realisation is constructed first by obtaining nonminimal realisation. This makes their method less attractive from computational point of view.

The problem of parameter estimation of a multivariate stochastic system described in terms of vector difference

equation has been studied by Valis (1970), Rowe (1970), Wilson (1973) and Kashyap et al (1974). For a given {A,B,C} many equivalent vector difference equations can be formed. The number of possible structures of the vector difference equation depends on the properties of the particular pair of matrices A, C. This implies that a-priori knowledge of the order of the system is inadequate to specify the form of the vector differential equation and the structure of A has to be given as well. The estimation of the parameters of the vector difference equation are rather involved.

In this chapter an identification method has been proposed from which the impulse response matrix H(z) is obtained from the observations of the system inputs and outputs over a finite interval of time. In theory the transfer function matrix H(z) can be calculated from the {A,B,C} matrices. In practice this computation can be highly inaccurate as discussed by Bosley et al (1972). The proposed algorithm determines the transfer function matrix directly from the input-output observations. This algorithm is based on the properties of matrix pseudoinverse and yields least squares estimates of the parameters of the transfer function matrix. The method is iterative in nature and rather simple to apply.

## 6.2. Pseudo-inverse Algorithm for Multivariable Systems.

### 6.2.1. Statement of the Problem.

Consider a discrete-time system with p inputs and m outputs. It can be represented by an mxp transfer function matrix, H(z), with the following input-output relationship in terms of z-transforms:

$$\underset{\sim}{Y}(z) = H(z)\ \underset{\sim}{U}(z) \tag{6.4}$$

where

$$\underset{\sim}{Y}(Z) = \left[Y_1(z)\ Y_2(z). \ . \ . \ Y_m(z)\right]^T \tag{6.5}$$

$$= Z\left[y_1(t)\ y_2(t). \ . \ . \ y_m(t)\right]^T \tag{6.6}$$

and

$$\underset{\sim}{U}(Z) = \left[U_1(z)\ U_2(z). \ . \ . \ U_p(z)\right]^T \tag{6.7}$$

$$= Z\left[u_1(t)\ u_2(t). \ . \ . \ u_2(t)\right]^T \tag{6.8}$$

The superscript T represents transposition and the symbol Z representing z-transform for a sampling period $t_s$.

The transfer function matrix H(z) may be written as:

$$H(z) = \begin{bmatrix} H_{11}(z) & H_{12}(z) & . \ . \ . & H_{1p}(z) \\ H_{21}(z) & H_{22}(z) & . \ . \ . & H_{2p}(z) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ H_{m1}(z) & H_{m2}(z) & . \ . \ . & H_{mp}(z) \end{bmatrix} \tag{6.9}$$

where each element, $H_{ij}(z)$, $i = 1,2...m$; $j = 1,2...p$, is a rational function of z. The transfer function matrix may also be expressed as:

$$H(z) = \begin{bmatrix} \frac{M_{11}(z)}{D_{11}(z)} & \frac{M_{12}(z)}{D_{12}(z)} & \frac{M_{1p}(z)}{D_{1p}(z)} \\ \frac{M_{21}(z)}{D_{21}(z)} & \frac{M_{22}(z)}{D_{22}(z)} & \frac{M_{2p}(z)}{D_{2p}(z)} \\ \vdots & \vdots & \vdots \\ \frac{M_{m1}(z)}{D_{m1}(z)} & \frac{M_{m2}(z)}{D_{mL}(z)} & \frac{M_{m2}(z)}{D_{mp}(z)} \end{bmatrix} \qquad (6.10)$$

where $H_{ij}(z)$ in equation (6.9) is replaced by $M_{ij}(z)/D_{ij}(z)$, $M_{ij}(z)$ and $D_{ij}(z)$ being polynomials in z. Though putting H(z) in this form shown in equation (6.10) results in least number of parameters, in general, this is not very suitable for parameter estimation. This is because when the system difference equation is written from (6.10) using input-output observations, some of the parameters appear jointly. This difficulty of separating the parameters makes it difficult to apply sequential linear least squares techniques for parameter estimation.

The transfer function matrix which is particularly suitable for sequential estimation may be expressed as:

$$H(Z) = \frac{1}{D(z)} \begin{bmatrix} N_{11}(z) & N_{12}(z) & \cdots & N_{1p}(z) \\ N_{21}(z) & N_{22}(z) & \cdots & N_{2p}(z) \\ \vdots & \vdots & & \\ N_{m1}(z) & N_{m2}(z) & \cdots & N_{mp}(z) \end{bmatrix} \qquad (6.10)$$

where $D(z)$ is the characteristic polynomial of the system, defined as the least common monic denominator of all minors of $H(z)$, and $N_{ij}(z)$ are polynomials in z. It will be assumed that $D(z)$ is a polynomial of order n, so that one may write:

$$D(z) = z^n + b_1 z^{n-1} + \cdots + b_{n-1} z + b_n \qquad (6.11)$$

The polynomials $N_{ij}(z)$ may, similarly, be written as:

$$N_{ij}(z) = a_{ij}(1) z^{n-1} + a_{ij}(2) z^{n-2} + \cdots + a_{ij}(n-1) z + a_{ij}(n) \qquad (6.12)$$

where

$$i = 1,2 \ldots m; \quad j = 1,2 \ldots p$$

Our problem may be then stated as the estimation of the parameters $b_1, b_2 \ldots b_n$; and $a_{ij}(1), a_{ij}(2) \ldots a_{ij}(n)$ for $i = 1,2, \ldots m$ and $j = 1,2 \ldots p$ from the measurements of the system inputs and outputs.

6.2.2. Derivation of the Identification Algorithm.

From the definition of H(z) given in equation (6.11), the ith output may be written as:

$$Y_i(z) = \frac{1}{D(z)} \sum_{j=1}^{p} N_{ij}(z)\, U_j(z) \qquad (6.14)$$

where $i = 1,2, \ldots m$.

It is easily seen that the ith output at the rth sampling instant is given by:

$$y_i(r) \triangleq y_i(rt_s)$$

$$= \sum_{j=1}^{p} \sum_{s=1}^{n} a_{ij}(s)\, u_j(r-s) - \sum_{j=1}^{n} b_j y_i(r-j) \qquad (6.15)$$

Define for $i = 1,2, \ldots m$:

$$\underset{\sim}{\phi}_i \triangleq \left[a_{i1}(1)\ a_{i1}(2) \ldots a_{i1}(n) \ldots a_{ip}(1),\ a_{ip}(2) \ldots a_{ip}(n)\right] \qquad (6.16)$$

$$\underset{\sim}{b} \triangleq \left[b_1\ b_2 \ldots b_n\right]^T, \qquad (6.17)$$

and

$$\underset{\sim}{y}_i(r) \triangleq \left[y_i(r-1)\ y_i(r-2) \ldots y_i(r-n)\right] \qquad (6.18)$$

Also define for j = 1,2 . . . p

$$\underset{\sim}{u}_j(r) \overset{\Delta}{=} \left[u_j(r-1)\ u_j(r-2)\ \ldots\ u_j(r-n)\right] \qquad (6.19)$$

Then, equation (6.15 ) may be written more compactly as:

$$y_i(r) = \left[\underset{\sim}{u}_1(r)\ \underset{\sim}{u}_2(r)\ \ldots\ \underset{\sim}{u}_p(r)\ -\underset{\sim}{y}_i(r)\right] \begin{bmatrix} \underset{\sim}{\phi}_i \\ b \end{bmatrix} \qquad (6.20)$$

Collecting the different outputs related through equation (6.20) for i = 1,2 . . . m we get the matrix equation:

$$\underset{\sim}{y}(r) = A(r)\ \underset{\sim}{\phi} \qquad (6.21)$$

where

$$\underset{\sim}{y}(r) \overset{\Delta}{=} \left[y_1(r)\ y_2(r)\ \ldots\ y_m(r)\right]^T, \qquad (6.22)$$

$$\underset{\sim}{\phi} \overset{\Delta}{=} \left[\underset{\sim}{\phi}_1^T\ \underset{\sim}{\phi}_2^T\ \ldots\ \underset{\sim}{\phi}_m^T\ b^T\right]^T \qquad (6.23)$$

and

$$A(r) = \begin{bmatrix} \underset{\sim}{u}_1(r) & \underset{\sim}{u}_2(r) & \ldots & \underset{\sim}{u}_p(r) & \underset{\sim}{0} & \underset{\sim}{0} & \ldots & \underset{\sim}{0} & \underset{\sim}{0}\ \underset{\sim}{0}\ \ldots\ \underset{\sim}{0} & -\underset{\sim}{y}_1(r) \\ \underset{\sim}{0} & \underset{\sim}{0} & \ldots & \underset{\sim}{0} & \underset{\sim}{u}_1(r) & \underset{\sim}{u}_2(r) & \ldots & \underset{\sim}{u}_p(r) & \underset{\sim}{0}\ \underset{\sim}{0}\ \ldots\ \underset{\sim}{0} & -\underset{\sim}{y}_2(r) \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \\ \cdot & & & & & & & & & \\ \underset{\sim}{0} & & & & & & & & & \underset{\sim}{u}_p(r) - \underset{\sim}{y}_m(r) \end{bmatrix} \qquad (6.24)$$

It may be noted that the dimension of the parameter vector $\underset{\sim}{\phi}$ is given by:

$$q = n(mp + 1)$$

Consequently, the matrix A(r) has m rows and q columns.

Collecting vectors $\underset{\sim}{y}(r)$ for r = 1,2,...k, we have:

$$\begin{bmatrix} \underset{\sim}{y}(1) \\ \underset{\sim}{y}(2) \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{y}(k) \end{bmatrix} = \begin{bmatrix} A(1) \\ A(2) \\ \cdot \\ \cdot \\ \cdot \\ A(k) \end{bmatrix} \underset{\sim}{\phi} \qquad (6.25)$$

or,

$$Y(k) = \underset{\sim}{A}(k)\ \underset{\sim}{\phi} \qquad (6.26)$$

where,

$$Y(k) \overset{\Delta}{=} \begin{bmatrix} \underset{\sim}{y}(1) \\ \underset{\sim}{y}(2) \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{y}(k) \end{bmatrix} \qquad (6.27)$$

and

$$\underset{\sim}{A}(k) \triangleq \begin{bmatrix} A(1) \\ A(2) \\ \cdot \\ \cdot \\ \cdot \\ A(k) \end{bmatrix} \qquad (6.28)$$

Equation (6.26) provides km linear equations for the q components of the parameter vector $\underset{\sim}{\phi}$. If km > q the least-squares estimate of the parameter vector is given by:

$$\underset{\sim}{\hat{\phi}}_k = \underset{\sim}{A}^{+}(k)\, Y(k) \qquad (6.29)$$

where $\underset{\sim}{A}^{+}(k)$ is the pseudo-inverse of $\underset{\sim}{A}(k)$, defined by:

$$\underset{\sim}{A}^{+}(k) = \left[\underset{\sim}{A}^{T}(k)\, \underset{\sim}{A}(k)\right]^{-1} \underset{\sim}{A}^{T}(k) \text{ for } km > q. \qquad (6.30)$$

Although the least-squares estimate of the parameter vector may be obtained from equation (6.29), it is not convenient as the inversion of a qxq matrix is required. Moreover, this has the additional disadvantage that all the past data in the interval over which the estimates are made must be stored.

These problems suggest the necessity for a recursive scheme. In the following such a recursive algorithm is presented for estimating the parameters of the transfer function

matrix H(z).

6.2.3. Recursive Algorithm for Multivariable System Identification.

Each block of new input output data adds m rows A(k+1) to the matrix $\underset{\sim}{A}(k)$. Thus let:

$$\underset{\sim}{A}(k+1) = \begin{bmatrix} \underset{\sim}{A}(k) \\ A(k+1) \end{bmatrix} \tag{6.31}$$

and

$$Y(k+1) = \begin{bmatrix} Y(k) \\ \underset{\sim}{y}(k+1) \end{bmatrix} \tag{6.32}$$

Then the following recursive algorithm may be derived, as shown in Appendix III.

For km > q

$$\hat{\underset{\sim}{\phi}}_{k+1} = \hat{\underset{\sim}{\phi}}_k + P_k A^T(k+1) \left[I + A(k+1)P_k A^T(k+1)\right]^{-1} \left[\underset{\sim}{y}(k+1) - A(k+1)\hat{\underset{\sim}{\phi}}_k\right] \tag{6.33}$$

$$P_{k+1} = P_k - P_k A^T(k+1) \left[I + A(k+1)P_k A^T(k+1)\right]^{-1} A(k+1)P_k \tag{6.34}$$

It may be noted that unlike the recursive algorithm given in Chapter four in the case of single-input, single-output case, equations (6.33) and (6.34) do require matrix inversion. But the size of the matrix to be inverted is only mxm as compared with qxq in equation (6.29), where q = n(mp+1), and is much larger than m.

To start the algorithm, one may note that:

$$P_k = \underset{\sim}{A}^{+}(k) \left[\underset{\sim}{A}^{+}(k)\right]^{T} \tag{6.35}$$

$$= \left[\left[\underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}(k)\right]^{-1}\right]^{T} \tag{6.36}$$

Hence one may start with an $\underset{\sim}{A}(k)$ which is square for $k = q/m$ (or the next higher integer, in which case some rows may be deleted) and obtain the initial estimate $\underset{\sim}{\hat{\phi}}_o$ and $P_o$ for the case.

6.2.4. Results of Simulation.

To test the algorithm an example of a system with two inputs and two outputs was considered. As in equation (6.10), the transfer function matrix of such a system is of the following form:

$$H(z) = \frac{1}{D(z)} \begin{bmatrix} N_{11}(z) & N_{12}(z) \\ N_{21}(z) & N_{22}(z) \end{bmatrix} \tag{6.37}$$

For the given example it was assumed that:

$$D(z) = z^3 + b_1 z^2 + b_2 z + b_3 \tag{6.38}$$

$$N_{11}(Z) = a_{11}(1)\, z^2 + a_{11}(2)\, z + a_{11}(3) \tag{6.39}$$

$$N_{12}(Z) = a_{12}(1)\, z^2 + a_{12}(2)\, z + a_{12}(3) \tag{6.40}$$

$$N_{21}(Z) = a_{21}(2)\, z + a_{21}(3) + a_{21}(4)\, z^{-1} \tag{6.41}$$

$$N_{22}(Z) = a_{22}(1)\, z^2 + a_{22}(2)\, z + a_{22}(3) + a_{22}(4) z^{-1} \tag{6.42}$$

with the following values of the parameters:

$$a_{11}(1) = 3, \quad a_{11}(2) = -3.5, \quad a_{11}(3) = -1.5$$
$$a_{12}(1) = 1, \quad a_{12}(2) = -0.167, \quad a_{12}(3) = -0.167$$
$$a_{21}(2) = -4, \quad a_{21}(3) = -2, \quad a_{21}(4) = -1$$
$$a_{22}(1) = 1, \quad a_{22}(2) = -0.167, \quad a_{22}(3) = -0.083$$
$$a_{22}(4) = -0.167, \quad b_1 = 0.833, \quad b_2 = 0.417, \quad b_3 = 0.083.$$

The two inputs to the system were taken as uncorrelated unit variance white Gaussian noise sequences. To the outputs, uncorrelated white noise was added to obtain various noise to signal ratios. It was found that the proposed algorithm gave reasonably good estimates of the parameters for noise-to-signal ratios not exceeding 3%, and the results were rather poor if these ratios exceed 10%. In this case 900 samples of input-output data were used for recursively estimating the sixteen parameters of the system for noise to signal ratios ranging from 1% to 3%. The estimates are shown in Tables 6.1,6.2 and 6.3. The normalized estimation error, defined as

$$\frac{||\underset{\sim}{\phi} - \underset{\sim}{\hat{\phi}}||^2}{||\underset{\sim}{\phi}||^2}$$

has been plotted against the number of iterations k in figure 6.1.

It is seen that when the output noise level is small the estimates are fairly good, and reasonable estimates are obtained even after 200 iterations. It is seen from figure 6.1 that with higher noise-to-signal ratios the estimates are biased and inconsistent. For the case where the noise level is high the data must be prefiltered before applying the algorithm.

Although the algorithm is computationally simple and requires much less effort to obtain transfer function matrix, the number of parameters become quite large as the dimension of the system increases. This is because of the particular form of the transfer function in equation (6.10) and this may pose some convergence problems resulting in poor parameter estimates. The problem of convergence acceleration has to be further investigated.

The algorithm requires the inversion of an mxm matrix where m is the number of outputs. Since this number is often quite small, it may be possible to use this algorithm for on-line system identification.

| Parameter | Estimates after the following number of iterations. | | | | | True Value |
|---|---|---|---|---|---|---|
| | 100 | 300 | 500 | 700 | 900 | |
| $a_{11}(1)$ | 2.995 | 3.002 | 3.008 | 3.006 | 3.004 | 3 |
| $a_{11}(2)$ | -3.570 | -3.535 | -3.543 | -3.543 | -3.546 | -3.5 |
| $a_{11}(3)$ | -1.451 | -1.464 | -1.426 | -1.426 | -1.421 | -1.5 |
| $a_{12}(1)$ | 1.000 | 1.005 | 1.005 | 1.006 | 1.004 | 1 |
| $a_{12}(2)$ | -0.187 | -0.175 | -0.179 | -0.181 | -0.182 | -0.167 |
| $a_{12}(3)$ | -0.167 | -0.164 | -0.157 | -0.158 | -0.158 | -0.167 |
| $a_{21}(2)$ | -4.012 | -4.001 | -3.998 | -4.001 | -4.000 | -4 |
| $a_{21}(3)$ | -1.946 | -1.952 | -1.936 | -1.938 | -1.934 | -2 |
| $a_{21}(4)$ | -0.989 | -0.982 | -0.992 | -0.993 | -0.993 | -1 |
| $a_{22}(1)$ | 0.995 | 0.997 | 1.000 | 1.001 | 1.000 | 1 |
| $a_{22}(2)$ | -0.197 | -0.181 | -0.184 | -0.185 | -0.186 | -0.167 |
| $a_{22}(3)$ | -0.081 | -0.077 | -0.077 | -0.076 | -0.075 | -0.083 |
| $a_{22}(4)$ | -0.171 | -0.166 | -0.169 | -0.170 | -0.171 | -0.167 |
| $b_1$ | 0.814 | 0.821 | 0.817 | 0.817 | 0.817 | 0.833 |
| $b_2$ | 0.403 | 0.409 | 0.409 | 0.409 | 0.410 | 0.417 |
| $b_3$ | 0.077 | 0.080 | 0.080 | 0.080 | 0.081 | 0.083 |

Table 6.1 Estimates of different parameters with various numbers of iterations for noise-to-signal ratio of 1%.

| Parameter | Estimates after the following number of iterations. | | | | | True Value |
|---|---|---|---|---|---|---|
| | 100 | 300 | 500 | 700 | 900 | |
| $a_{11}(1)$ | 2.990 | 3.003 | 3.016 | 3.011 | 3.007 | 3 |
| $a_{11}(2)$ | -3.720 | -3.654 | -3.668 | -3.668 | -3.672 | -3.5 |
| $a_{11}(3)$ | -1.278 | -1.300 | -1.228 | -1.230 | -1.220 | -1.5 |
| $a_{12}(1)$ | 1.000 | 1.011 | 1.011 | 1.012 | 1.008 | 1 |
| $a_{12}(2)$ | -0.233 | -0.211 | -0.200 | -0.223 | -0.225 | -0.167 |
| $a_{12}(3)$ | -0.153 | -0.147 | -0.132 | -0.135 | -0.136 | -0.167 |
| $a_{21}(2)$ | -4.021 | -4.002 | -3.996 | -4.003 | -4.000 | -4 |
| $a_{21}(3)$ | -1.785 | -1.792 | -1.765 | -1.768 | -1.760 | -2 |
| $a_{21}(4)$ | -0.968 | -0.950 | -0.970 | -0.974 | -0.972 | -1 |
| $a_{22}(1)$ | 0.992 | 0.995 | 1.000 | 1.001 | 0.999 | 1 |
| $a_{22}(2)$ | -0.253 | -0.223 | -0.228 | -0.228 | -0.231 | -0.167 |
| $a_{22}(3)$ | -0.064 | -0.055 | -0.055 | -0.054 | -0.053 | -0.083 |
| $a_{22}(4)$ | -0.178 | -0.168 | -0.173 | -0.174 | -0.176 | -0.167 |
| $b_1$ | 0.769 | 0.782 | 0.774 | 0.775 | 0.774 | 0.833 |
| $b_2$ | 0.378 | 0.388 | 0.388 | 0.389 | 0.390 | 0.417 |
| $b_3$ | 0.067 | 0.072 | 0.072 | 0.073 | 0.073 | 0.083 |

Table 6.2. Estimates of different parameters with various number of iterations for noise-to-signal ratio of 2%.

| Parameter | Estimates after the following number of iterations | | | | | True Value |
|---|---|---|---|---|---|---|
| | 100 | 300 | 500 | 700 | 900 | |
| $a_{11}(1)$ | 2.986 | 3.004 | 3.023 | 3.015 | 3.010 | 3 |
| $a_{11}(2)$ | -3.921 | -3.835 | -3.851 | -3.848 | -3.855 | -3.5 |
| $a_{11}(3)$ | -1.024 | -1.040 | -0.944 | -0.948 | -0.933 | -1.5 |
| $a_{12}(1)$ | 1.000 | 1.016 | 1.016 | 1.017 | 1.012 | 1 |
| $a_{12}(2)$ | -0.298 | -0.269 | -0.280 | -0.285 | -0.287 | -0.167 |
| $a_{12}(3)$ | -0.129 | -0.120 | -0.098 | -0.102 | -0.104 | -0.167 |
| $a_{21}(2)$ | -4.029 | -4.002 | -3.994 | -4.004 | -4.000 | -4 |
| $a_{21}(3)$ | -1.553 | -1.550 | -1.518 | -1.523 | -1.511 | -2 |
| $a_{21}(4)$ | -0.941 | -0.909 | -0.939 | -0.944 | -0.942 | -1 |
| $a_{22}(1)$ | 0.990 | 0.992 | 0.999 | 1.000 | 0.999 | 1 |
| $a_{22}(2)$ | -0.326 | -0.284 | -0.291 | -0.290 | -0.295 | -0.167 |
| $a_{22}(3)$ | -0.036 | -0.021 | -0.024 | -0.021 | -0.019 | -0.083 |
| $a_{22}(4)$ | -0.186 | -0.172 | -0.178 | -0.179 | -0.182 | -0.167 |
| $b_1$ | 0.706 | 0.721 | 0.712 | 0.713 | 0.711 | 0.833 |
| $b_2$ | 0.345 | 0.357 | 0.358 | 0.360 | 0.361 | 0.417 |
| $b_3$ | 0.053 | 0.061 | 0.061 | 0.061 | 0.062 | 0.083 |

Table 6.3 Estimates of different parameters with various number of iterations for noise-to-signal ratio of 3%.

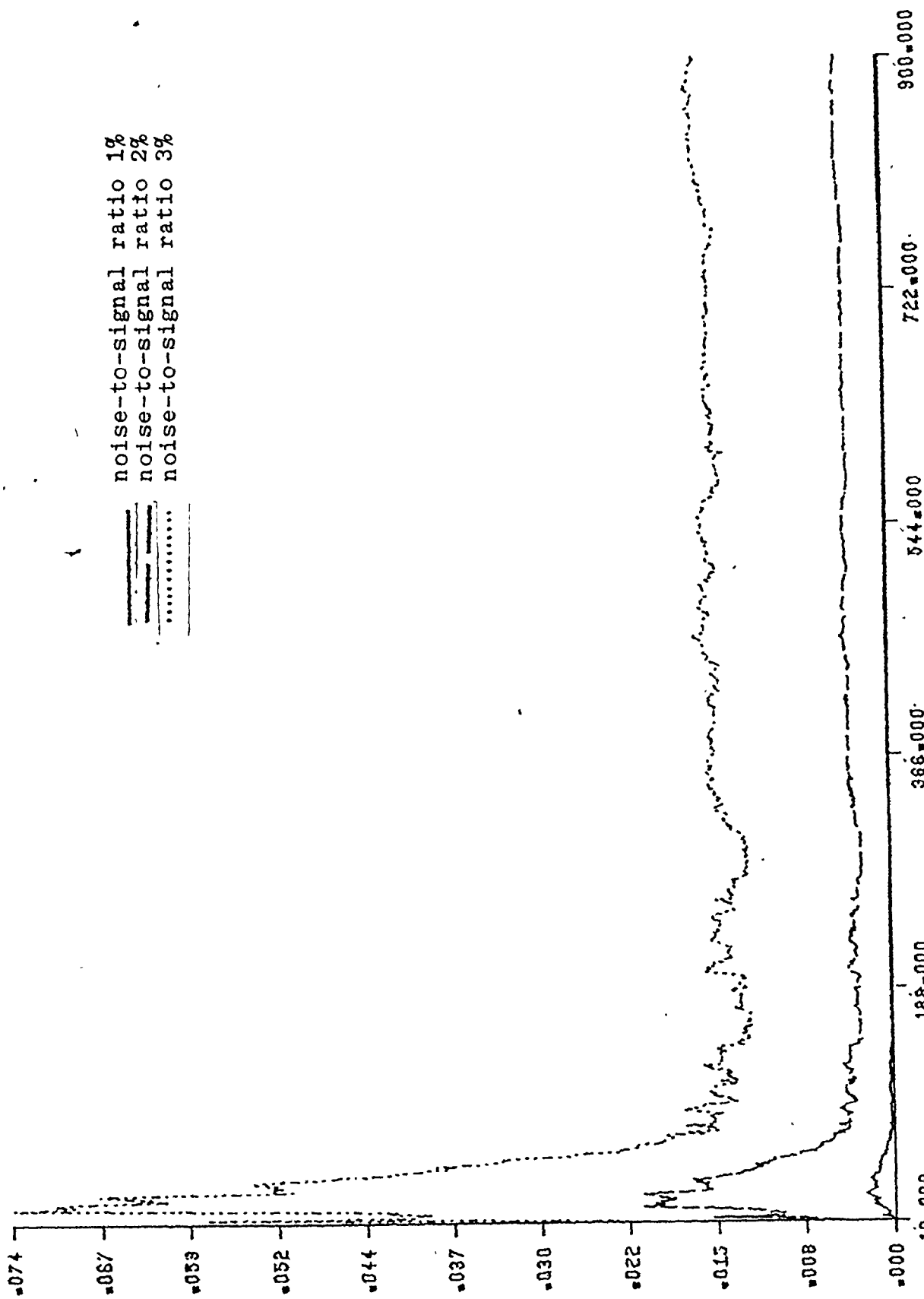



Figure 6.1 Convergence rates for two-input, two-output system for different noise to signal ratio

# CHAPTER 7

# CONCLUSIONS

The major effort in this thesis is directed towards the problem of System Identification in the presence of disturbances. Attention is directed towards the identification of both single-input, single-output and multi-input, multi-output stable linear discrete-time models from the input-output measurements which are normally contaminated with noise in realistic problems. The implementation of the parameter estimation algorithms requires only samples of the normal input and output of the system, and no description of the noise statistics is involved. This is the situation which often arises in system analysis, where the noise statistics are not known and only the input-output record is available.

Chapter 1 provides an introduction to this work through a brief literature survey which shows the need for developing better schemes to identify the parameters of the model.

Chapter 2 is an exposition on the problems associated with the modelling of an unknown process operating in a stochastic environment, under open-loop conditions.

Chapter 3 reviews the different methods for off-line identification.

A more practical problem is that of identifying a system while it is in normal operation. In Chapter 4 different schemes for on-line identification are discussed. Some of the disadvantages of these schemes are demonstrated. Two new algorithms are proposed to overcome some of the problems encountered in previous known schemes.

Results of simulation with tables and plots are presented in Chapter five.

Chapter Six considers the identification of a multi-input, multi-output system. A new algorithm is developed which directly determines the parameters of the multivariable system described by the transfer function matrix. The algorithm is applied to estimate the parameters of a dual-input, dual-output system.

The identification algorithms discussed in this thesis can be put into two main groups: (i) those which do not use any statistical models for the noise, and, (ii) those which attempt to filter the data using some kind of noise model so that the residual errors are uncorrelated. All the methods except the generalized pseudoinverse algorithm and the algorithm combining stochastic approximation and pseudo-inverse belong to the first category, whereas the latter algorithms belong to the second category.

Under conditions of low input-output noise levels the pseudoinverse algorithm provides good estimates of the system parameters. Not only does it require the least time for computation (with the exception of the stochastic approximation method), but also the estimates converge to correct values very fast. However, for higher noise-to-signal ratio the bias and inconsistency in estimates is quite pronounced. The stochastic approximation method, on the other hand, requires the least amount of computation per iteration but it has a very slow rate of convergence. All the other methods considered in this dissertation try to develop prefiltering techniques to reduce the effect of noise. The generalized least squares method proposed by Clarke (1967) was an important contribution in this direction. The generalized pseudoinverse algorithm proposed in this thesis is an attempt to obtain a recursive version of Clarke's method using the pseudoinverse for estimating the system parameters as well as the parameters of the auxiliary noise model. The algorithm tends to remove the bias of the estimates generated by the pseudionverse algorithm. This is accomplished through whitening the correlated residuals by filtering input-output data. However, this requires considerably more computing time compared to other algorithms.

The second algorithm which is proposed in the thesis uses pseudoinverse algorithm to determine the process model parameters and stochastic approximation algorithm for the auxiliary noise

parameters. This algorithm is found to be computationally more efficient relative to recursive version of generalised least squares. The main reason for this is the inclusion of a much simpler stochastic approximation algorithm to estimate the parameters of the noise model. The inclusion of the noise model, as in the case of the generalised pseudoinverse algorithm tends to remove the bias in estimates generated by ordinary pseudoinverse algorithm. In all the examples considered the algorithm combining stochastic approximation exhibits faster convergence and better accuracy with some increase in computational effort.

The major appeal of the proposed on-line parameter estimation algorithms is their simplicity. Since the storage requirements for data are small and the identification algorithms require few arithmetic operations, they may be implemented on a minicomputer in real time as reported by Tang (1975).

However, no theoretical justification has yet been offered for the observed improvement in convergence obtained with combined pseudoinverse and stochastic approximation algorithm. This theoretical aspect of convergence of the 2nd algorithm needs to be further explored.

For multivariable system a new recursive algorithm is proposed. This extends the recursive pseudoinverse algorithm to estimate parameters of single input single output case. Although the algorithm is computationally simple and easy to implement, the number of parameters may become quite large as the dimension of the system increases. This may cause some convergence problems

resulting in poor parameter estimates. It may be added that although a state-space formulation requires fewer parameters, no on-line algorithms have been developed so far. Even the off-line algorithms for this formulation are rather complicated and require the knowledge of the structure of the A-matrix as well as the order.

## 7.1. Suggestions for Future Research.

1. At present there is no rationale for the choice of the gain factor $\gamma(k)$ in the stochastic approximation algorithm which would best suit a particular situation. It would be a worthwhile research area to devise an iterative scheme by which an optimal prediction of this gain factor can be done on-line. However, a time consuming iterative scheme to predict the optimal gain would destroy the major appeal of our present approach. This computational aspect could be further investigated.

2. The generalised pseudoinverse algorithm, and the combined pseudoinverse and stochastic approximation algorithm have been shown to work well in specific examples, with reasonable ad hoc choices for the order of the noise filter. There are no systematic rules for the choice of the order of autoregression. The problem of predicting the optimal filter order needs to be further explored.

3. The convergence of the algorithm combining pseudoinverse and stochastic approximation has not been theoretically justified. When these two estimators are treated seperately, the convergence of each can be argued if the other one satisfies certain properties. However, the convergence of overall estimator is not obvious. This aspect has to be further studied.

4. Throughout our discussion it was assumed that the system to be identified had an open loop structure. However, despite the fundamental importance of closed-loop identification, most parameter estimation algorithms suffer severe difficulties in the presence of feedback. There appears to be little work done in the areas of on-line identification of a closed-loop system. This is an area where there is much scope for further work.

5. The problem of identifying multivariable systems has not been solved completely. The basic difficulty with these systems is in finding a suitable representation. Unlike the scalar output case, a multivariable system does not admit a unique canonical representation. Further work is needed for choosing the structure of the multivariable systems which is best suited for identification purposes under realistic conditions, i.e., when the data are contaminated with noise.

## APPENDIX 1

## DEFINITION OF PSEUDOINVERSE AND ITS PROPERTIES

When the matrix A is invertible it is possible to solve the equation:

$$Ax = y \tag{A1.1}$$

by operating on both sides with the inverse of A. The concept of pseudoinverse of a matrix is introduced to extend the above technique to situations in which A has no inverse but the equation (A1.1) has a solution. One way of defining the pseudoinverse is as follows:

Let A be a matrix of dimension mxn with rank equal to r. Let A be factorized into two matrices B and C such that:

$$A = BC \tag{A1.2}$$

where B is a mxr matrix of rank r, and C is a rxn matrix of rank r. The above factorization can be obtained first by selecting B such that its columns are the linearly independent columns of A. Since A is of rank r, the dimension of B is mxr. C is chosen, such that it satisfies equation (A1.2).

The pseudoinverse of A is defined as:

$$\begin{aligned} A^{+} &= C^{T}\left[CC^{T}\right]^{-1}\left[B^{T}B\right]^{-1}B^{T} \quad \text{if } A \neq 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } A = 0 \end{aligned} \tag{A1.3}$$

It can be proved (Greville, 1959), that there always exists a unique pseudoinverse, $A^+$ defined as in (A1.3) for any matrix A. Two cases can be distinguished:

(1) Let the rank of A be n, i.e. n = r.

In this case (A1.2) reduces to:

$$A = BI \tag{A1.4}$$

and (A1.3) reduces to

$$A_L^+ = \left[B^T B\right]^{-1} B^T = A^T A^{-1} AT \tag{A1.5}$$

$A_L^+$ is called left pseudoinverse of A.

(2) Let the rank of A be m i.e. m = r.

In this case (A1.2) reduces to:

$$A = IC \tag{A1.6}$$

and (A1.3)reduces to:

$$\begin{aligned} A_R^+ &= C^T\left[CC^T\right]^{-1} \\ &= A^T\left[AA^T\right]^{-1} \end{aligned} \tag{A1.7}$$

$A_R^+$ is called the right pseudoinverse of A.

## A1.1. Properties of the Pseudoinverse

For every real mxn matrix A, there exists a unique real pseudoinverse $A^+$, defined as in (A1.3), which satisfies the following identities:

$$\begin{aligned} A^+AA^+ &= A^+ \\ AA^+A &= A \\ \left[AA^+\right]^T &= AA^+ \\ \left[A^+A\right]^T &= A^+A \end{aligned} \tag{A1.8}$$

where the superscript T denotes transposition.

Theorem 1: Let A be an (mxn) matrix of rank r = m. Let $A^T\left[AA^T\right]^{-1}$ be the right pseudoinverse of A as defined in equation (A1.7). Then the minimum norm solution $\hat{x}$ for the equation:

$$Ax = y$$

with

$$||\hat{x}||^2 \leq ||x||^2$$

is given by

$$\hat{x} = A_R^+ y \qquad \text{(A1.9)}$$

Proof: Let $J = \frac{1}{2}||x||^2 + \lambda^T (y - Ax)$

$$\frac{\partial J}{\partial x}\Big|_{\hat{x}} = \hat{x}^T - \lambda^T A = 0$$

therefore

$$\hat{x}^T = \lambda^T A \qquad \text{(A1.10)}$$

i.e. $$\hat{x}^T A_R^+ = \lambda^T A A_R^+$$
$$= \lambda^T$$

i.e. $$\lambda = A_R^{+T} \hat{x} \qquad \text{(A1.11)}$$

Thus:

$$\lambda = A_R^{+T} A_R^+ y$$
$$= \left[AA^T\right]^{-1} y \qquad \text{(A1.12)}$$

From (A1.10) and (A1.12) we have:

$$\hat{x} = A^T\left[AA^T\right]^{-1} y$$
$$= A_R^+ y \qquad \text{Q.E.D.}$$

Theorem 2: Let A be an (mxn) matrix of rank = n. Let $A_L^+ = \left[A^T A\right]^{-1} A^T$ be the left pseudoinverse of A as defined in equation (A1.5). Then the solution $\hat{x}$ minimising the norm of the residual error:

$$e = y - Ax$$

of the vector equation

$$Ax = y$$

is given by

$$\hat{x} = A_L^+ y \qquad \text{(A1.13)}$$

Proof: Let $J = ||y - A\phi||^2$

$$\frac{\partial J}{\partial x}_{|\hat{x}} = 0 = -2\left[y - A\hat{x}\right]^T A^T$$

$$A^T A\hat{x} = A^T y$$

$$\hat{x} = \left[A^T A\right]^{-1} A^T y$$

$$= A_L^+ y \qquad \text{Q.E.D.}$$

Thus the left pseudoinverse yields the least squares approximate solution in the sense that it minimises the square of residual error.

# APPENDIX II

## DERIVATION OF RECURSIVE GENERALIZED PSEUDOINVERSE ALGORITHM

It is shown in section 4.4.1 of chapter four that in order to arrive at an unbiased least squares estimate of the process parameters $\underset{\sim}{\phi}$ the incoming data has to be filtered. Filtering of the inputs and outputs is accomplished via equation (4.42). Let $\hat{F}_k(z^{-1})$ be the present estimate of the autoregressive noise process.

$$r^*_{k+1} = \hat{F}_k(z^{-1})\, r_{k+1} \qquad \text{(A2.1)}$$

$$y^*_{k+1} = \hat{F}_k(z^{-1})\, y_{k+1} \qquad \text{(A2.2)}$$

From equation (4.41) we have:

$$\underset{\sim}{y}^*_k = A^*_k \underset{\sim}{\phi} + \underset{\sim}{w}_k \qquad \text{(A2.3)}$$

where

$$\underset{\sim}{y}^*_k = \left[y^*_1 y_2^* \cdots y^*_k\right]^T \qquad \text{(A2.4)}$$

$$\underset{\sim}{w}_k = \left[w_1 w_2 \cdots w_k\right]^T \qquad \text{(A2.5)}$$

$$A_k^* = \begin{bmatrix} r_1^* & r_o^* & \cdots & r_{1-m}^* & -y_o^* & -y_{-1}^* & \cdots & -y_{1-n}^* \\ r_2^* & r_1^* & \cdots & r_{2-m}^* & -y_1^* & -y_o^* & \cdots & -y_{2-n}^* \\ \cdot & & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & & \cdot \\ r_k^* & r_{k-1}^* & \cdots & r_{k-m}^* & -y_{k-1}^* & -y_{k-2}^* & \cdots & -y_{k-n}^* \end{bmatrix} \quad \text{(A2.6)}$$

The effect of the new filtered data on the present process parameter $\hat{\underset{\sim}{\phi}}_k$ adds a new row to the matrix $A_k^*$. Thus let:

$$A_{k+1}^* = \begin{bmatrix} A_k^* \\ \underset{\sim}{a}_{k+1}^{*T} \end{bmatrix} \quad \text{(A2.7)}$$

where

$$\underset{\sim}{a}_{k+1}^{*T} = \begin{bmatrix} r_{k+1}^* & r_k^* & \cdots & r_{k+1-m}^* & -y_k^* & -y_{k-1}^* & \cdots & -y_{k+1-n}^* \end{bmatrix} \quad \text{(A2.8)}$$

and let

$$\underset{\sim}{y}_{k+1}^* = \begin{bmatrix} \underset{\sim}{y}_k^* \\ y_{k+1}^* \end{bmatrix} \quad \text{(A2.9)}$$

Then from equation (A2.3) we have

$$\underset{\sim}{y}_{k+1}^* = A_{k+1}^* \underset{\sim}{\phi} + \underset{\sim}{w}_{k+1} \quad \text{(A2.10)}$$

$$\begin{bmatrix} \underset{\sim}{y}_k^* \\ y_{k+1}^* \end{bmatrix} = \begin{bmatrix} A_k^* \\ \underset{\sim}{a}_{k+1}^{*T} \end{bmatrix} \underset{\sim}{\phi} + \begin{bmatrix} \underset{\sim}{w}_k \\ w_{k+1} \end{bmatrix} \quad \text{(A2.11)}$$

The least squares estimates $\hat{\phi}_{k+1}$ is given by:

$$\underset{\sim}{\hat{\phi}}_{k+1} = A^{*+}_{k+1} \underset{\sim}{y}^{*}_{k+1} \qquad \text{(A2.12)}$$

where

$$A^{*+}_{k+1} = \text{Pseudoinverse of } A^{*}_{k+1}$$

$$= \left[A^{*T}_{k+1} A^{*}_{k+1}\right]^{-1} A^{*T}_{k+1} \text{ for } k > p-1 \qquad \text{(A2.13)}$$

where $p = m+n+1$

Assuming $A^{*+}_{k}$ is of full rank, the pseudoinverse of $A_{k+1}$ may be written as:

$$A^{*+}_{k+1} = \left[C^{T}_{k} \;:\; c_{k+1}\right] \qquad \text{(2.14)}$$

From: (A2.7) and (A2.14):

$$A^{*+}_{k+1} A^{*}_{k+1} = C^{T}_{k} A^{*}_{k} + c_{k+1} \underset{\sim}{a}^{*T}_{k+1} \qquad \text{(A2.15)}$$

Postmultiplying equation (A2.15) by $A^{*+}_{k}$, we have:

$$A^{*+}_{k+1} A^{*}_{k+1} A^{*+}_{k} - C^{T}_{k} A^{*}_{k} A^{*+}_{k} + c_{k+1} a^{*T}_{k+1} A^{*+}_{k} \qquad \text{(A2.16)}$$

Now $$A^{*+}_{k+1} A^{*}_{k+1} = I \qquad \text{(A2.17)}$$

and $$C^{T}_{k} A^{*}_{k} A^{*+}_{k} = C^{T}_{k} \qquad \text{(A2.18)}$$

Employing (A2.17) and (A2.18) from (A2.16) we have:

$$C_k^T = A_k^{*+} - c_{k+1} \underset{\sim}{a}_{k+1}^{*T} A_k^{*+} \qquad \text{(A2.19)}$$

Therefore:

$$A_{k+1}^{*+} = \left[A_k^{*+} - c_{k+1} \underset{\sim}{a}_{k+1}^{*T} A_k^{*+} : c_{k+1}\right] \qquad \text{(A2.20)}$$

From the properties of pseudoinverse in appendix I equation (A1.8) we have:

$$A_{k+1}^{*+} A_{k+1}^{*+T} A_{k+1}^{*T} = A_{k+1}^{*+} \qquad \text{(A2.21)}$$

Substituting equations (A2.7) and (A2.20) into equation (A2.21) and postmultiplying the first resulting equations by $A_k^{*+T} \underset{\sim}{a}_{k+1}^{*}$ and substituting the result into second equation yields:

$$c_{k+1} = \left[1 + \underset{\sim}{a}_{k+1}^{*T} A_k^{*+} A_k^{*+T} \underset{\sim}{a}_{k+1}^{*}\right]^{-1} A_k^{*+} A_k^{+T} \underset{\sim}{a}_{k+1}^{*} \qquad \text{(A2.22)}$$

Define:

$$P_k = A_k^{*+} A_k^{*+T} \qquad \text{(A2.23)}$$

Then

$$c_{k+1} = \left[1 + \underset{\sim}{a}_{k+1}^{*T} P_k \underset{\sim}{a}_{k+1}^{*}\right]^{-1} P_k \underset{\sim}{a}_{k+1} \qquad \text{(A2.24)}$$

Also

$$P_{k+1} = A_{k+1}^{*+} A_{k+1}^{*+T} \qquad \text{(A2.25)}$$

Substituting (A2.7), (A2.20) and (A2.24) in equation (A2.25) yields:

$$P_{k+1} = P_k - \left[P_k \underset{\sim}{a}^*_{k+1}\right] \left[P_k \underset{\sim}{a}^*_{k+1}\right]^T \left[1 + \underset{\sim}{a}^{*T}_{k+1} P_k \underset{\sim}{a}^*_{k+1}\right]^{-1} \tag{A2.26}$$

Now from (A2.12)

$$\begin{aligned} \hat{\underset{\sim}{\phi}}_{k+1} &= A^{*+}_{k+1} \underset{\sim}{y}^*_{k+1} \\ &= \left[C^T_k \underset{\sim}{y}^*_k + c_{k+1} y^*_{k+1}\right] \end{aligned} \tag{A2.27}$$

Substituting (A2.19), (A2.24) in (A2.27) and using the fact $\hat{\underset{\sim}{\phi}}_k = A^{*+}_k \underset{\sim}{y}^*_k$ we have for $k > p$:

$$\hat{\underset{\sim}{\phi}}_{k+1} = \hat{\underset{\sim}{\phi}}_k + P_k \underset{\sim}{a}^*_{k+1} (y^*_{k+1} - \underset{\sim}{a}^{*T}_{k+1} \hat{\underset{\sim}{\phi}}_k) \left[1 + \underset{\sim}{a}^*_{k+1} P_k \underset{\sim}{a}^*_{k+1}\right]^{-1} \tag{A2.28}$$

Thus for $k > p$ the recursive algorithm is given by:

$$\hat{\underset{\sim}{\phi}}_{k+1} = \hat{\underset{\sim}{\phi}}_k + \frac{P_k \underset{\sim}{a}^*_{k+1} (y^*_{k+1} - \underset{\sim}{a}^{*T}_{k+1} \hat{\underset{\sim}{\phi}}_k)}{1 + \underset{\sim}{a}^{*T}_{k+1} P_k \underset{\sim}{a}^*_{k+1}} \tag{A2.29}$$

$$P_{k+1} = P_k - \frac{P_k \underset{\sim}{a}^*_{k+1} \left[P_k \underset{\sim}{a}^*_{k+1}\right]^T}{1 + \underset{\sim}{a}^{*T}_{k+1} P_k \underset{\sim}{a}^*_{k+1}} \tag{A2.30}$$

The values of $P_p$ and $\hat{\underset{\sim}{\phi}}_p$ are given by:

$$P_p = \left[A_p^{*-1}\right]\left[A_p^{*-1}\right]^T \qquad \text{(A2.31)}$$

$$\hat{\underset{\sim}{\phi}}_p = \left[A_p^*\right]^{-1} \underset{\sim}{y}^*_p \qquad \text{(A2.32)}$$

It remains to update the noise model parameters $\hat{F}_k$ on the basis of new observations, $y_{k+1}$. The noise present in (k+1)th. output observation $y_{k+1}$.

$$\hat{e}_{k+1} = y_{k+1} - \underset{\sim}{a}^{*T}_{k+1}\ \hat{\underset{\sim}{\phi}}_{k+1}$$

From equation (4.39) of Chapter 4 we have:

$$\hat{\underset{\sim}{e}}_k = G_k\ \underset{\sim}{\psi} + \underset{\sim}{w}_k \qquad \text{(A2.33)}$$

where $\hat{e}_k$ is the estimate of $e_k$

$$G_k = \begin{bmatrix} -\hat{e}_o & -\hat{e}_{-1} & \cdots & -\hat{e}_{1-s} \\ -\hat{e}_1 & -\hat{e}_o & \cdots & -\hat{e}_{2-s} \\ \cdot & & & \\ \cdot & & & \\ -\hat{e}_{k-1} & -\hat{e}_{k-2} & \cdots & -\hat{e}_{k-s} \end{bmatrix} \qquad \text{(A2.34)}$$

and $$\hat{\underset{\sim}{e}}_k = \left[\hat{e}_1\ \hat{e}_2 \cdots \hat{e}_k\right]^T$$

Writing $G_{k+1}$ as:

$$G_{k+1} = \begin{bmatrix} G_k \\ \underset{\sim}{g}^T_{k+1} \end{bmatrix} \qquad \text{(A2.35)}$$

where

$$\underset{\sim}{g}_{k+1}^{T} = \left[-\hat{e}_{k} \; -\hat{e}_{k-1} \; \cdots \; -\hat{e}_{k+1-s}\right] \qquad \text{(A2.36)}$$

we have

$$\underset{\sim}{\hat{e}}_{k+1} = G_{k+1}\,\psi + \underset{\sim}{w}_{k+1} \qquad \text{(A2.37)}$$

i.e.

$$\begin{bmatrix} \underset{\sim}{\hat{e}}_{k} \\ \hat{e}_{k+1} \end{bmatrix} = \begin{bmatrix} G_{k} \\ g_{k+1}^{T} \end{bmatrix} \psi + \begin{bmatrix} \underset{\sim}{w}_{k} \\ w_{k} \end{bmatrix} . \qquad \text{(A2.38)}$$

Thus least square estimates of $\hat{\psi}$ at (k+1) instant:

$$\hat{\psi}_{k+1} = \hat{G}_{k+1}^{+}\,\underset{\sim}{\hat{e}}_{k+1} \qquad \text{(A2.39)}$$

where $\hat{G}_{k+1}^{+}$ is the pseudoinverse of $G_{k+1}$. This equation (A2.39) is similar in form to the equation (A2.12). Thus the recursive equations for estimations $\psi$ can similarly be evaluated. This is given by:

$$\underset{\sim}{\hat{\psi}}_{k+1} = \underset{\sim}{\hat{\psi}}_{k} + \frac{R_{k}\,\underset{\sim}{g}_{k+1}\,(\underset{\sim}{\hat{e}}_{k+1} - \underset{\sim}{g}_{k+1}^{T}\,\underset{\sim}{\hat{\psi}}_{k})}{1 + \underset{\sim}{g}_{k+1}^{T}\,R_{k}\,\underset{\sim}{g}_{k+1}} \qquad \text{(A2.40)}$$

and

$$R_{k+1} = R_{k} - \frac{R_{k}\,\underset{\sim}{g}_{k+1}\,(R_{k}\,\underset{\sim}{g}_{k+1})^{T}}{1 + \underset{\sim}{g}_{k+1}^{T}\,R_{k}\,\underset{\sim}{g}_{k+1}} \qquad \text{(A2.41)}$$

# APPENDIX III

## DERIVATION OF EQUATIONS FOR RECURSIVE ESTIMATION FOR MULTIVARIABLE SYSTEM

Let

$$\underset{\sim}{A}(k+1) = \begin{bmatrix} A(1) \\ A(2) \\ \cdot \\ \cdot \\ \cdot \\ A(k) \\ A(k+1) \end{bmatrix} = \begin{bmatrix} \underset{\sim}{A}(k) \\ A(k+1) \end{bmatrix} \qquad \text{(A3.1)}$$

$$Y(k+1) = \begin{bmatrix} \underset{\sim}{y}(1) \\ \underset{\sim}{y}(2) \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{y}(k) \\ \underset{\sim}{y}(k+1) \end{bmatrix} = \begin{bmatrix} Y(k) \\ \underset{\sim}{y}(k+1) \end{bmatrix} \qquad \text{(A3.2)}$$

We will first derive $\underset{\sim}{A}(k+1)$, the pseudoinverse of $\underset{\sim}{A}(k+1)$.

Let $\underset{\sim}{A}^{+}(k+1) = \begin{bmatrix} C_k^T & D_{k+1} \end{bmatrix}$

where $C_k^T$ : q x km matrix

$D_{k+1}$ : q x m matrix

For $(k+1)m > q$

$$\underset{\sim}{A}^{+}(k+1) = \left[\underset{\sim}{A}^{T}(k+1) \quad \underset{\sim}{A}(k+1)\right]^{-1} \underset{\sim}{A}^{T}(k+1) \qquad \text{(A3.4)}$$

$$\underset{\sim}{A}^{+}(k+1)\,\underset{\sim}{A}(k+1) = \left[\underset{\sim}{A}^{T}(k+1)\,\underset{\sim}{A}(k+1)\right]^{-1} \underset{\sim}{A}^{T}(k+1)\,\underset{\sim}{A}(k+1)$$

$$= I$$

$$\underset{\sim}{A}^{+}(k+1)\,\underset{\sim}{A}(k+1)\,\underset{\sim}{A}^{+}(k) = \underset{\sim}{A}^{+}(k) \qquad \text{(A3.5)}$$

From (A3.1), (A3.3), and (A3.5)

$$\left[C_k^T \; D_{k+1}\right]\begin{bmatrix}\underset{\sim}{A}(k) \\ \underset{\sim}{A}(k+1)\end{bmatrix}\underset{\sim}{A}^{+}(k) = \underset{\sim}{A}^{+}(k)$$

$$\left[C_k^T\,\underset{\sim}{A}(k) + D_{k+1}A(k+1)\right]\underset{\sim}{A}^{+}(k) = \underset{\sim}{A}^{+}(k)$$

$$C_k^T\,\underset{\sim}{A}(k)\,\underset{\sim}{A}^{+}(k) + D_{k+1}A(k+1)\,\underset{\sim}{A}^{+}(k) = \underset{\sim}{A}^{+}(k) \qquad \text{(A3.6)}$$

Lemma: 1

$$C_k^T\,\underset{\sim}{A}(k)\,\underset{\sim}{A}^{+}(k) = C_k^T \text{ for } km > q \qquad \text{(A3.7)}$$

Proof: Since from the properties of pseudoinverse

$$\underset{\sim}{A}(k) = \underset{\sim}{A}(k)\,\underset{\sim}{A}^{+}(k)\,\underset{\sim}{A}(k)$$

$$\underset{\sim}{A}(k)\,\underset{\sim}{A}^{+}(k) = \underset{\sim}{A}(k)\,\underset{\sim}{A}^{+}(k)\,\underset{\sim}{A}(k)\,\underset{\sim}{A}^{+}(k)$$

since $\underset{\sim}{A}(k)\,\underset{\sim}{A}^{+}(k)$ is a projector matrix

$C_k^T$ has columns in the column space of $\underset{\sim}{A}_k^T$

$$C_k^T = C_k^T \underset{\sim}{A}(k) \underset{\sim}{A}^+(k)$$

From using (A3.6) Lemma 1

$$C_k^T + D_{k+1}A(k+1) \underset{\sim}{A}^+(k) = \underset{\sim}{A}^+(k) \qquad \text{(A3.8)}$$

or

$$C_k^T = \underset{\sim}{A}^+(k) - D_{k+1} A(k+1) \underset{\sim}{A}^+(k) \qquad \text{(A3.9)}$$

From(A3.3)and (A3.9)

$$\underset{\sim}{A}^+(k+1) = \left[A^+(k)-D_{k+1}A(k+1) \underset{\sim}{A}^+(k) D_{k+1}\right] \qquad \text{(A3.10)}$$

Lemma 2

$$\underset{\sim}{A}^+(k+1) \underset{\sim}{A}^{+T}(k+1) \underset{\sim}{A}^T(k+1) = \underset{\sim}{A}^+(k+1) \qquad \text{(A3.11)}$$

Proof: From the properties of pseudoinverse

$$\underset{\sim}{A}^+(k+1) \underset{\sim}{A}(k+1) \underset{\sim}{A}^+(k+1) = \underset{\sim}{A}^+(k+1)$$

$$\underset{\sim}{A}^+(k+1)\left[\underset{\sim}{A}(k+1) \underset{\sim}{A}^+(k+1)\right]^T = \underset{\sim}{A}^+(k+1)$$

$$\underset{\sim}{A}^+(k+1) \underset{\sim}{A}^{+T}(k+1) \underset{\sim}{A}^T(k+1) = \underset{\sim}{A}^+(k+1) \qquad \text{Q.E.D.}$$

Since $\underset{\sim}{A}^+(k+1) \underset{\sim}{A}^{+T}(k+1) \underset{\sim}{A}^T(k+1) = \underset{\sim}{A}^+(k+1)$

$$\left[C_k^T D_{k+1}\right] \begin{bmatrix} C_k \\ D_{k+1}^T \end{bmatrix} \left[\underset{\sim}{A}^T(k) A^T(k+1)\right] = \left[C_k^T D_{k+1}\right]$$

$$\left[C_k^T C_k + D_{k+1} D_{k+1}^T\right] \left[\underset{\sim}{A}^T(k) A^T(k+1)\right] = \left[C_k^T D_{k+1}\right] \qquad \text{(A3.12)}$$

From (A3.12)

$$\left[C_k^T C_k + D_{k+1} D_{k+1}^T\right] \underset{\sim}{A}^T(k) = C_k^T \tag{A3.13}$$

and

$$\left[C_k^T C_k + D_{k+1} D_{k+1}^T\right] A^T(k+1) = D_{k+1} \tag{A3.14}$$

Multiplying (A3.13) by $\underset{\sim}{A}^{+T}(k)\ A^T(k+1)$, we have

$$\left[C_k^T C_k + D_{k+1}\ D_{k+1}^T\right] \underset{\sim}{A}^T(k)\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1) = C_k^T\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1) \tag{A3.15}$$

Now $C_k^T C_k$ =

$$\left[\underset{\sim}{A}^{+}(k) - D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\right]\left[\underset{\sim}{A}^{+T}(k) - \underset{\sim}{A}^{+T}(k)\ A^T(k+1)\ D_{k+1}^T\right]$$

$$= \Big[\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k) - \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1)\ D_{k+1}^T$$

$$- D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)$$

$$+ D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1)\ D_{k+1}^T\Big] \tag{A3.16}$$

From (A3.15) and (A3.16)

$$\Big[\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k) - \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1)\ D_{k+1}^T$$

$$- D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)$$

$$+ D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1)\ D_{k+1}^T$$

$$+ D_{k+1} D_{k+1}^T\Big]\ \underset{\sim}{A}^T(k)\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1)$$

$$= C_k^T\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1)$$

or

$$\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ \underset{\sim}{A}^T(k)\ \underset{\sim}{A}^{+T}(k)\ A^T(k+1)$$

$$- \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$- D_{k+1}A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$+ D_{k+1}A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$+ D_{k+1}D_{k+1}^{T}\underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$= C_{k}^{T}\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1) \qquad \text{(A3.17)}$$

Since

$$\underset{\sim}{A}^{+T}(k)\ \underset{\sim}{A}^{T}(k)\ A^{+T}(k) = \underset{\sim}{A}^{+T}(k)$$

from (A3.17)

$$\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$-\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$- D_{k+1}\ A_{k+1}\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$+ D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$+ D_{k+1}\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$= C_{k}^{T}\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$= \left[\underset{\sim}{A}^{+}(k) - D_{k+1}\ A_{k+1}\ \underset{\sim}{A}^{+}(k)\right]\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

Hence

$$D_{k+1}\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$= \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$- D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)A^{T}(k+1)$$

(A3.18)

From (A3.14) and (A3.16)

$$\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1) - \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ A^{T}(k+1)$$

$$- D_{k+1}A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$+ D_{k+1}A_{k+1}\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ A^{T}(k+1)$$

$$+ D_{k+1}\ D_{k+1}^{T}\ A^{T}(k+1) = D_{k+1} \qquad \text{(A3.19)}$$

Postmultiply (A3.19) by $D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$

$$\underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$-\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)D_{k+1}^{T}A^{T}(k+1)D_{k+1}^{T}\underset{\sim}{A}^{T}(k)D_{k+1}^{T}\underset{\sim}{A}^{T}(k)\underset{\sim}{A}^{+T}(k)\underset{\sim}{A}^{T}(k+1)$$

$$- D_{k+1}A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)D_{k+1}^{T}\underset{\sim}{A}^{T}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)$$

$$+ D_{k+1}A_{(k+1)}\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)D_{k+1}^{T}A^{T}(k+1)D_{k+1}^{T}\underset{\sim}{A}^{T}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)$$

$$+ D_{k+1}\ D_{k+1}^{T}\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)$$

$$= D_{k+1}\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1) \qquad \text{(A3.20)}$$

Equating (A3.20) and (A3.18)

$$- \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ A^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k+1)\ D_{k+1}^{T}\ \underset{\sim}{A}^{T}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)$$

$$+ D_{k+1}A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)D_{k+1}^{T}A^{T}(k+1)D_{k+1}^{T}\underset{\sim}{A}^{T}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)$$

$$+ D_{k+1}D_{k+1}^{T}A^{T}(k+1)D_{k+1}^{T}\underset{\sim}{A}^{T}(k)\underset{\sim}{A}^{+T}(k)A^{T}(K+1) = 0$$

$$\left[-\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)+D_{k+1}A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)\right.$$

$$\left.+D_{k+1}\right]\ D_{k+1}^{T}A^{T}(k+1)D_{k+1}^{T}\underset{\sim}{A}^{T}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1) = 0$$

$$-\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)+D_{k+1}A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)+D_{k+1} = 0$$

or

$$D_{k+1}\left[I+A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)\right] = \underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)$$

$$D_{k+1} = \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)A^{T}(k+1)\ \left[I+A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)\right]^{-1}$$

(A3.21)

So

$$\underset{\sim}{A}^{+}(k+1) = \left[C_k^T\ D_{k+1}\right]$$

where

$$C_k^T = \underset{\sim}{A}^{+}(k) - D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)$$

$$D_{k+1} = \underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)\ \left[I+A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)\right]^{-1}$$

Define

$$P_k = \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)\ \{\text{Dimension of } p_k; q \times q\} \qquad \text{(A3.22)}$$

$$P_{k+1} = \underset{\sim}{A}^{+}(k+1)\ \underset{\sim}{A}^{+T}(k+1)$$

$$= \left[C_k^T\ D_{k+1}\right]\begin{bmatrix} C_k \\ D_{k+1}^T \end{bmatrix}$$

$$P_{k+1} = C_k^T C_k + D_{k+1}\ D_{k+1}^T$$

$$= \left[\underset{\sim}{A}^{+}(k)-D_{k+1}A(k+1)\underset{\sim}{A}^{+}(k)\right]\left[\underset{\sim}{A}^{+T}(k)-\underset{\sim}{A}^{+T}(k)A^{T}(k+1)D_{k+1}^T\right]$$

$$+ D_{k+1}D_{k+1}^T$$

$$P_{k+1} = \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k) - \underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)D_{k+1}^T$$

$$- D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\ \underset{\sim}{A}^{+T}(k)$$

$$+ D_{k+1}A(k+1)\underset{\sim}{A}^{+}(k)\underset{\sim}{A}^{+T}(k)A^{T}(k+1)D_{k+1}^{T}+D_{k+1}D_{k+1}^{T}$$

$$P_{k+1} = P_k - P_k A^{T}(k+1) D_{k+1}^{T}-D_{k+1}A_{k+1}P_k$$

$$+ D_{k+1} A(k+1) P_k A^{T}(k+1) D_{k+1}^{T} + D_{k+1} D_{k+1}^{T}$$

$$= P_k + \left[-P_k A^{T}(k+1) + D_{k+1} A(k+1) P_k A^{T}(k+1) + D_{k+1}\right] D_{k+1}^{T}$$

$$- D_{k+1} A_{k+1} P_k \qquad \text{(A3.23)}$$

Since by (A3.21)

$$D_{k+1}\left[I + A(k+1) P_k A^{T}(k+1)\right] = P_k A^{T}(k+1)$$

$$D_{k+1} + D_{k+1} A(k+1) P_k A^{T}(k+1)= P_k A^{T}(k+1)$$

$$-P_k A^{T}(k+1) + D_{k+1}A(k+1) P_k A^{T}(k+1) + D_{k+1} = 0 \qquad \text{(A3.24)}$$

By (A3.22) and (A3.23)

$$P_{k+1} = P_k - D_{k+1} A_{k+1} P_k$$

$$= P_k - P_k A^{T}(k+1) \left[I+A(k+1)P_k A^{T}(k+1)\right]^{-1} A(k+1)P_k$$

$$P_{k+1} = P_k - P_k A^{T}(k+1) \left[I+A(k+1)P_k A^{T}(k+1)\right]^{-1} A(k+1)P_k \qquad \text{(A3.25)}$$

For mk > q

$$\underset{\sim}{\hat{\phi}}_k = \underset{\sim}{A}^{+}(k)\ Y(k) \qquad \text{(A3.26)}$$

Consider one more block of data

$$\underset{\sim}{\hat{\phi}}_{k+1} = \underset{\sim}{A}^{+}(k+1)\ Y(k+1)$$

$$= \left[C_k^{T}\ D_{k+1}\right] \begin{bmatrix} Y(k) \\ \underset{\sim}{y}(k+1) \end{bmatrix}$$

$$= C_k^{T}\ Y(k) + D_{k+1}\ \underset{\sim}{y}(k+1)$$

$$= \left[\underset{\sim}{A}^{+}(k) - D_{k+1}\ A(k+1)\ \underset{\sim}{A}^{+}(k)\right] Y(k) + D_{k+1}\ \underset{\sim}{y}(k+1)$$

$$\qquad \text{(A3.27)}$$

$$\hat{\underset{\sim}{\phi}}_{k+1} = \underset{\sim}{A}^{+}(k)\ Y(k) + D_{k+1}\left[\underset{\sim}{y}(k+1) - A(k+1)\ \underset{\sim}{A}^{+}(k)\ Y(k)\right]$$

$$= \hat{\underset{\sim}{\phi}}_{k} + D_{k+1}\left[\underset{\sim}{y}(k+1) - A(k+1)\ \hat{\underset{\sim}{\phi}}_{k}\right] \qquad \text{(A3.28)}$$

$$\hat{\underset{\sim}{\phi}}_{k+1} = \hat{\underset{\sim}{\phi}}_{k} + P_k A^T(k+1)\left[I+A(k+1)P_k A^T(k+1)\right]^{-1}\left[\underset{\sim}{y}(k+1)-A(k+1)\hat{\underset{\sim}{\phi}}_{k}\right] \qquad \text{(A3.29)}$$

So for mk > q

$$\hat{\underset{\sim}{\phi}}_{k+1} = \hat{\underset{\sim}{\phi}}_{k} + P_k A^T(k+1)\left[I+A(k+1)P_k A^T(k+1)\right]^{-1}\left[\underset{\sim}{y}(k+1)-A(k+1)\hat{\underset{\sim}{\phi}}_{k}\right]$$

$$P_{k+1} = P_k - P_k A^T(k+1)\left[I+A(k+1)P_k A^T(k+1)\right]^{-1} A(k+1)P_k$$

# BIBLIOGRAPHY


A. Albert &. L. Gardner (1967): "Stochastic Approximation and Non-linear Regression"; Cambridge, Mass. M.I.T. press.

A. Albert & R.W. Sittler (1966): "A Method for the Computing Least Squares Estimators that keep up with the Data"; J. SIAM Control Ser.A, vol. 3, pp. 384-417.

K.J. Astrom (1967): "On the Achievable Accuracy in Identification Problems"; Proc. IFAC Symposium on Identification in Automatic Control Systems, Prague, Paper 1-8.

K.J. Astrom & P. Eykhoff (1971): "System Identification - a survey"; Automatica, vol. 7, pp. 123-162.

K.J. Astrom & T. Bohlin (1965): "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", Proc. IFAC Symposium on Self-Adaptive Control Systems. Teddington, pp. 96-111.

T. Bohlin (1971): "On the Problem of Ambiguities in Maximum Likelihood Identification"; Automatica, vol. 7 no. 2, pp. 199-210.

M.J. Bosley, H.W. Kropholler, F.P. Lees, & R.N. Neale (1972): "The Determination of Transfer Functions from State Variable Modles"; Automatica, vol. 8, pp. 213-218.

G.E.P. Box & G.M. Jenkins (1970): "Time Series Analysis, Forecasting and Control"; Holden-Day, San Francisco.

G.E.P. Box, J.F. MacGregor (1974): "The Analysis of Closed-loop Dynamic-stochastic System", Technometrics, vol. 16, no. 3, pp. 391-398.

G.E.P. Box, J.F. MacGregor (1975): "Parameter Estimation with Closed-loop Operating Data"; MacMaster University, Faculty of Engineering, Report No. SOC-99.

M.A. Budin (1971): "Minimal Realisations of Descrete Linear Systems from Input-output observations"; IEEE Trans. Automatic Control, AC-16, pp. 395-401.

C.T. Chen, D.P. Mital (1972): A simplified Irreduciable Realization Algorithm"; IEEE Trans. on Automatic Control, AC-17, pp. 535-537.

J.C. Chow (1972): "On Estimating the Orders of an Autoregressive Moving Average process with Uncertain Observations"; IEEE Trans. on Automatic Control, AC-17, pp. 268-269.

D.W. Clarke (1967): "Generalized Least Squares Estimation of the Parameters of a Dynamic Model"; IFAC Symposium on Identification in Automatic Control Systems, Prague, Paper - 3.17.

M. Cuenod, A.P. Sage (1968): "Comparison of Some Methods Used for Process Identification"; Automatica, vol. 4, pp. 235-269.

D.B. Cooper (1964): "Adaptive Pattern Recognition and Signal Detection Using Stochastic Approximation"; IEEE Trans. Electron. Comput. (Corresp.), vol. EC-13, pp. 306-307.

M. Djorovic and S.P. Bingulac (1972): "Order Determination and Parameter Identification of Time-invariant State Variable Models"; Proc. 5th IFAC Congress, Paris, paper no. 38.5.

A. Dvoretzky (1956): "On Stochastic Approximation"; Proc 3d Berkeley Symposium Math. Stat. and Probability, vol. 1, pp. 39-55.

P. Eykhoff (1968): "Process Parameter Estimation"; Automatica, vol. 4, pp. 205-233.

E.E. Fisher (1965): "Identification of Linear Systems"; J.A.C.C. pp. 473-475.

G.C. Goodwin, M.B. Zarrop, R.L. Payne (1974): "Coupled Design of Test Signals, Sampling Intervals, and Filters for System Identification"; IEEE Trans. Automatic Control, vol. AC-19, pp. 748-752.

B. Gopinath (1969): "On the Identification of Linear Time-Invariant Systems from Input-output Data"; Bell Syst. Tech. J., vol. 48, pp. 1101-1113.

T.N.E. Greville (1960): "Some Applications of the Pseudoinverse of a Matrix"; SIAM Review, vol. 2, no. 1, pp. 17-22.

T.N.E. Greville (1959): "The Pseudoinverse of a Rectangular or Singular Matrix and Its Applications to the Solution of Systems of Linear Equations"; SIAM Review, vol. 1, no. 1, pp. 38-43.

I. Gustavsson (1971): "Choice of Sampling Interval for Parametric Identification"; Report 7103, Div. of Automatic Control, Lund Institute of Technology, Lund, Sweden.

R. Hasting-James & M.W. Sage (1969): "Recursive Generalised Least Squares Procedure for On-Line Identification of Process Parameters"; Proc. IEE, vol. 116, no. 2, pp. 2057-2062.

V. Herles (1970): "Limits of Precision in Identification Procedures"; Second IFAC Symposium on Identification and Process Parameter Estimation, Prague, Paper 9.8.

B.L. Ho, R.E. Kalman (1965): "Effective construction of linear state-variable models from input-output functions"; Proc. 3rd Allerton Conf. on Circuit and System Theory, pp. 449-459.

J.K. Holmes (1969): "Two Stochastic approximation procedures for identifying linear systems"; IEEE Trans. Automatic Control, vol. AC-14, pp. 292-295.

H.M. Hong (1974): "Stochastic Digital Control of a Continuous Stirred Tank Reactor"; M.Eng. Thesis, Dept. of Chemical Engineering, McMaster University, Hamilton, Ontario.

H.M. Hong & J.F. MacGregor (1974): "Identification and Direct Digital Stochastic Control of a Continuous Stirred Tank Process"; McMaster University, Faculty of Engineering, Report no. SOC-63,

R. Iserman, U. Baur, W. Bamberger, P. Kneppo, and H. Siebert (1974): "Comparison of Six On-line Identification and Parameter Estimation Methods"; Automatica, vol. 10, pp. 81-103.

P. Joseph, J. Lewis, & J. Tow (1961): "Plant Identification in the Presence of Disturbances and Application to Digital Adaptive Systems"; Trans. AIEE (Applications & Industry), vol. 80, pp. 18-24.

R.L. Kashyap and R.E. Nasburg (1974): "Parameter Estimation in Multivariable Stochastic Difference Equations"; IEEE Trans. on Automatic Control, vol. AC-19, pp. 784-797.

J. Kiefer, & J. Wolfowitz (1952): "Statistical Estimation of the Maximum of a Regression Function"; Ann Math, Statistics, vol. 23, pp. 462-466.

H.G. Kwatny (1972): "A Note on Stochastic Approximation Algorithms in System Identification"; IEEE Trans. on Automatic Control, vol. AC-17, pp. 571-572.

R.C.K. Lee (1964): "Optimal Identification, Estimation and Control"; M.I.T. Press, Cambridge, Massa.

A.I. Liff (1966): "System Identification in the Presence of Noise by Digital Techniques"; IEEE International Convention, vol. 14, part 6, pp. 152-156.

L. Ljung (1974): "Convergence of Recursive Stochastic Algorithms"; IFAC Symposium on Stochastic Control, Budapest, pp. 215-223.

L. Ljung, I. Gustowsson, T. Soderstrom (1974): "Identification of Linear Multivariable Systems Operating under Linear Feedback Control"; IEEE Trans. Automatic Control, vol. AC-19, pp. 836-840.

J.F. MacGregor & H.M. Huynh (1974): "The Effect of Sampling Frequency on Discrete Time Series Models"; McMaster University, Faculty of Engineering, Report No. Soc-30.

J.M. Mendel & K.S. Fu, Eds. (1970): "Adaptive Learning and Pattern Recognition Systems"; New York: Academic 1970.

R.N. Pandya (1972): "A Class of Bootstrap Estimators for Linear System Identification"; International Journal of Control, vol. 15, no. 6, pp. 1091-1104.

R.N. Pandya (1974): "A Class of Bootstrap Estimatorw and Their Relationship to the Generalized Two Stage Least Squares Estimators"; IEEE Trnas. on Automatic Control, AC-19, pp. 836-840.

D.P. Passeri & C.J. Hergert (1972): "Parameter Identification of a class of Multiple Input/Multiple Output Linear Discrete-Time Systems"; J.A.C.C., pp. 786-792.

H. Robbins & S. Monro (1951): "A Stochastic Approximation Method"; Ann. Math. Statistics, vol. 22, pp. 400-407.

I.H. Rowe (1970): "A Bootstrap Method for the Statistical Estimation of Model Parameters"; International Journal of Control, vol. 12, no. 4, pp. 721-738.

D.J. Sakrison (1966): "Stochastic Approximation: A Recursive Method for Solving Regression Problems"; Advances in Communication Systems, vol. 2, pp. 51-106.

G.N. Saridis (1974): "Stochastic Approximation Methods for Identification and Control - A survey"; IEEE Trans. on Automatic Control, vol. AC-19, no. 6, pp. 798-809.

G.N. Saridis, Z.Z. Nikolic, & K.S. Fu (1969): "Stochastic Approximation Algorithms for System Identification, Estimation and Decomposition of Mixtures"; IEEE Trans. System Science & Cybernetics, vol. SSC-5, pp. 8-15.

G.N. Saridis & G. Stein (1968): "Stochastic Approximation Algorithms for Linear Discrete-Time System Identification"; IEEE Trans. on Automatic Control, vol. AC-13, no. 5, pp. 515-523.

A. Sen, N.K. Sinha and J.D. Wright (1974): "On-Line Identification of the Parameters of a Dual Input Heat-exchanger System"; The Canadian J. Chemical Engineer., vol. 52, no. 5, pp. 682-684.

L.M. Silverman (1971): "Realization of Linear Dynamical Systems"; IEEE Trans. on Automatic Control, AC-16, pp. 554-567.

N.K. Sinha & W. Pille (1971): "On-Line Parameter Estimation Using Matrix Pseudoinverse"; Proc. IEE, vol. 118, no. 8.

N.K. Sinha and P. Rozsa (1974): "Minimal Realisation of Transfer Function Matrices"; Proc. of Milwaukee Symposium on Automatic Control, Milwaukee, Wisconsin, pp. 345-348.

N.K. Sinha & A. Sen (1974): "Comparison of Some On-Line Identification Methods for a Simulated First Order Process"; Automatic Control Theory and Applications, vol. 2, no. 2, pp. 37-43.

J. Sklansky (1966): "Learning Systems for Automatic Control"; IEEE Trans. on Automatic Control, vol. AC-11, pp. 6-19.

J. Stepan (1970): "The Uncertainty Problem in System Identification"; Proc. 2nd IFAC Symposium on Identification and Process Parameter Estimation, Prague, Paper 1.8.

M.Y. Tang (1975): "On-Line System Identification in Real Time Using a Mini-computer"; M.Eng.Thesis, Dept. of Electrical Engineering, McMaster University, Hamilton, Ontario.

J. Valis (1970): "On-Line Identification of Multivariable Linear Systems of Unknown Structure from Input-output Data"; Proc. Second IFAC Symposium on "Identification and Process Parameter Estimation", Prague, paper 1.5.

N.T. Wasan (1969): "Stochastic Approximation"; Cambridge University Press, Cambridge, England.

C.H. Wells (1967): "Minimum Norm Control of Discrete Systems"; IEEE International Convention Record, vol. 15, part 3, pp. 55-64.

G.T. Wilson (1973): "The Estimation of Parameters in Multivariable Time Series Models"; J. Royal Statistical Society, vol. 35, pp. 76-85.

K.Y. Wong & E. Polak (1967): "Identification of Linear Discrete Time Systems Using the Instrumental Variable Method"; IEEE Trans. on Automatic Control, vol. AC-12, no. 6, pp. 707-718.

C.M. Woodside (1971): "Estimation of the Order of the Linear Systems"; Automatica, vol. 7, pp. 727-733.

J.D. Wright, & D.W. Bacon (1975): "Analysis of a Heat Exchanger Network using Statistical Time Series Analysis Methods"; Ind. Eng. Chem. Process Des. Dev., vol. 14, no. 4 pp. 453-459.

J.D. Wright & D.W. Bacon (1974): "Statistical Identification of the Dynamic Behaviour of a Dual Input Heat Exchanger Network"; Can. J. Chem. Eng., no. 52, pp. 405-414.

P.C. Young (1970): "An Instrumental Variable Method for Real Time Identification of a Noisy Process"; Automatica, vol. 6, pp. 271-287.