626c

STOCHASTIC APPROXIMATION FOR IDENTIFICATION OF MULTIVARIABLE SYSTEMS

STOCHASTIC APPROXIMATION FOR IDENTIFICATION OF MULTIVARIABLE SYSTEMS

By

HOSSNY E. EL-SHERIEF, B.Sc.

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

March 1977

MASTER OF ENGINEERING (Electrical Engineering) — McMASTER UNIVERSITY Hamilton, Ontario

TITLE: Stochastic Approximation for Identification of Multivariable Systems

AUTHOR: Hossny E. El-sherief B.Sc. (Elect. Eng.)(Cairo University,1973)
B.Sc. (Math.)(Ain Shams University,1975)

SUPERVISOR: Professor N. K. Sinha

NUMBER OF PAGES: viii, 60

SCOPE AND CONTENTS:

In this thesis a non-parametric normalized stochastic approximation algorithm has been developed for the identification of multivariable systems from noisy data without prior knowledge of the statistics of measurement noise.

The system model is first transformed into a special canonical form, then it is formulated in a non-parametric form. The parameters of this model are estimated through a normalized stochastic approximation algorithm. Finally, the system parameters are recovered from these estimates by another transformation.

The proposed algorithm is applied to the identification of two simulated systems.

Conclusions of this work and suggestions for future work are given.

ACKNOWLEDGMENTS

The author would like to express his sincere appreciation to Dr. N.K. Sinha for his excellent guidance and encouragement throughout the preparation of this work.

The author wishes to acknowledge with thanks the useful discussions with colleagues J. Hickin and Y. Kwong.

Financial support of the work by the Department of Electrical Engineering of McMaster University is gratefully acknowledged.

TABLE OF CONTENTS

TABLE OF CONTENTS - continued

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The stochastic approximation method has been very popular for system identification. It is a sequential technique for point estimation, first introduced by Robbins and Monro in 1951 [1] and later generalized by Dvoretzky [2].

Because of the great simplicity with which it can be implemented its application has been investigated by several authors [3] - [20]. Most of this work, however, deals with single-input single-output systems, little has been done on multivariable systems [21].

In this thesis a non-parametric normalized stochastic approximation algorithm has been developed for on-line identification of linear, multivariable, discrete-time stochastic systems. This algorithm does not require the knowledge of the statistics of measurement noise and converges to the true values of the parameters in the mean-square sense.

In chapter 2, the stochastic approximation method [1],[2] and [22] - [25] is introduced in its generality.

The proposed algorithm has been developed in Chapter 3. First, the system equations are transformed into a special canonical form proposed by Luenberger [26]- [29]. Then, using the method of

Saridisand Lobbia [21], the system is formulated in a non-parametric form. The parameters of the non-parametric model are estimated by applying a set of single-output multi-input normalized stochastic approximation algorithms [4], [7], [14]. The identification interval introduced in [21] has been varied and the effect of its increase on the accuracy of the estimates is investigated. Finally, the system parameters are recovered from the estimates of the non-parametric model using a transformation given in [21].

In Chapter 4, the proposed algorithm is applied to the identification of two different multivariable systems with various noise-to-signal ratios. The effect of increasing the identification interval on the accuracy of the estimates and the increase in the computation time has been discussed. For the sake of comparison, the first system has been identified using Sen and Sinha's algorithm [31], the computation time and estimates have been compared with those of the proposed algorithm. The results of simulations are given which indicate that the proposed algorithm works quite well.

Finally, the conclusions of this work and suggestions for future work are given in Chapter 5.

CHAPTER 2

THE STOCHASTIC APPROXIMATION METHOD

2.1 Introduction

Stochastic approximation is concerned with schemes converging to some sought value when, due to the stochastic nature of the problem, the observations involve errors. The interesting schemes are those which are self-correcting, that is, in which a mistake always tends to be wiped out in the limit, and in which the convergence to the desired value is of some specified nature. An example is mean-square convergence.

Major contributions to the area of stochastic approximation have been made by Robbins and Monro [1] , Kiefer and Wolfowitz [22] and Dvoretzky [2].

2.2 Features of Stochastic Approximation:

Stochastic approximation can be applied to any problem that can be formulated as some form of regression in which repeated observations are made.

Compared to conventional methods such as maximum liklihood estimation it has the following advantages:-

(1) Only a small interval of data needs processing.

(2) Only simple computations are required, even when the actual functional dependence of the regression functions

on the parameters of interest is non-linear.

(3) 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.

2.3 Stochastic Approximation Algorithms:

The most general form of a stochastic approximation algorithm has been treated by Dvoretzky in [2]. He has proved a general theorem which deals with the convergence properties of a nonlinear measurable transformation $T(x(1),\ldots,x(n))$, of a sequence of random measurements $x(1),\ldots,x(n)$ to a point vector θ. The algorithm is of the following general form

$$x(n+1) = T\,(x(1),\ldots,x(n)) + y_n + g(x(1),\ldots,x(n)).$$

In the above y_n is a random variable and $g(\cdot)$ is a measurable function.

For most practical applications stochastic approximation search algorithms are point estimators of the forms.

$$\alpha\,(k) = \alpha\,(k-1) + [\text{gain}]_k * [\text{error correction}]_{k-1} \qquad (2.1)$$

where the $[\text{gain}]_k = \{\nu_k\}$ is a sequence of suitable chosen smoothing values and the $[\text{error correction}]_{k-1} = \{F(k-1)\}$ sequence is generated at every time instant k by measuring the deviations from an appropriate goal [23]. In order for (2.1) to qualify as a stochastic approximation algorithm, convergence to the unbiased true parameter α must be established.

Conditions of convergence of the sequence α (k) to α in (2.1) are stated in Dvoretzky's special theorem [2]. This theorem has been modified to fit algorithm (2.1), and is presented in the sequel.

In order to formulate the theorem, the relation (2.1) can be rewritten in the following form in which the gains are presented by ν_k and the error correction sequence is partitioned into a correction term F(k) and a noise term v(k);

$$\alpha(k) = \alpha(k-1) + \nu_k[F(k) + v(k)] \tag{2.2}$$

Theorem ([2] simplified):

If the gain sequence $\{\nu_k\}$ in (2.2) satisfies

$$\lim_{k \to \infty} \nu_k = 0 \quad , \quad \sum_{k=1}^{\infty} \nu_k = \infty, \quad \sum_{k=1}^{\infty} \nu_k^2 < \infty \tag{2.3}$$

and the error correction sequence satisfies

$$E\{||\alpha(k) + \nu_{k+1}[F(k) + v(k)]||^2 / \alpha(k)\} < E\{||\alpha(k)+\nu_{k+1}F(k)||^2 / \alpha(k)\} + \nu_{k+1}^2 E\{||v(k)||^2 / \alpha(k)\} \quad ;$$

$$E\{||\alpha(0)||^2\} < \infty \quad ; \quad E\{||v(k)||^2\} \leq \sigma^2 < \infty \tag{2.4}$$

Then

$$P_r\{\lim_{k\to\infty}||\alpha(k) - \alpha|| = 0\} = 1 \quad \text{and} \quad \lim_{k\to\infty}\{||E\,\alpha(k) - \alpha||^2\} = 0$$

Proof of this theorem can be obtained from [2]. The conditions (2.3) on the gains may be interpreted heuristically as follows. The first provides the smoothing effect on the random correction term, the second provides

unlimited correction effort, and the third guarantees mutual cancellation of individual errors for a large number of iterations. The harmonic sequence $\{1/k\}$ as well as any sequence of the form $\{1/Pk + C\}$, $1/2 < P < 1$, $C > 0$ satisfies condition (2.3). Conditions (2.4) imply that there is no cross-coupling between $F(k)$ and $v(k)$ and that the search does not start with an infinite uncertainty about the parameters. The parameter α may be scalar, a vector, or a matrix. This affects only the bookkeeping of the algorithm. Historically, the first stochastic approximation algorithm was presented by Robbins and Monro in 1951 [1], to search for the zero of unknown function $f(x)$ of a random variable x, corrupted with measurement noise v.

$$z(x) = f(x) + v$$

If conditions (2.3) and (2.4) are satisfied, the sequence

$$x(k+1) = x(k) - \nu_{k+1}\, z(x(k))$$

converges with probability one and in the mean-square sense to the point Θ, $f(\Theta) = 0$. Kesten [24] proposed an acceleration scheme of the algorithm and Dupač [25] extended it to cover some nonstationary environments.

Kiefer and Wolfowitz in 1952 [22] 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.

If the new conditions (2.3),

$$\lim_{k \to \infty} \nu_k = 0 \,, \quad \lim_{k \to \infty} C_k = 0 \,, \quad \lim_{k \to \infty} \sum_{n=1}^{k} \nu_n = \infty \,,$$

$$\lim_{k \to \infty} \sum_{n=1}^{k} \left[\frac{\nu_n}{C_n} \right]^2 < \infty \qquad (2.5)$$

and the conditions (2.4) are satisfied then the sequence

$$x(k+1) = x(k) + \nu_k \left[\frac{z(x(k)+C_k) - z(x(k) - C_k)}{2C_k} \right]$$

converges with probability one and in the mean-square sense to the point Θ ; $\frac{\partial f(\Theta)}{\partial x} = 0$

Finally, in 1956 Dvoretzky presented his general and special theorems, a simplified version of which was given above.

CHAPTER 3

ALGORITHM FOR IDENTIFICATION OF MULTIVARIABLE SYSTEMS

3.1 Introduction

Consider a linear, time invariant, discrete-time system which is described by the following set of equations.

$$\begin{aligned} x(k+1) &= A\,x(k) + B\,u(k) \\ y(k) &= H\,x(k) + v(k) \end{aligned} \tag{3}$$

where,

$x(k)$ is an n-dimensional state vector.

$y(k)$ is an m-dimensional output measurement vector.

$u(k)$ is an r-dimensional vector sequence of independent random variables with zero mean and

$$E\{u(k)\,u^{T}(j)\} = \sigma^{2} I\,\delta_{kj} \quad \text{for all } k \text{ and } j.$$

$v(k)$ is an m-dimensional vector sequence of additive (not directly measured) random variables with zero mean and finite variances, uncorrelated with $u(k)$.

A,B, and H are system matrices of dimensions nxn, nxr and mxn respectively.

Our aim is to identify the matrices A,B,and H, of system (3),

through subsequent observations of the variables u(k) and y(k) without knowing the statistics of the measurement noise. In order to do so, first, the matrices of system (3) will be transformed into a special canonical form. Then, the system in its new canonical form will be transformed into a non-parametric form. After that, the parameters of the non-parametric model will be identified by a normalized stochastic approximation algorithm. Finally, system matrices are recovered from the parameters of the non-parametric model.

3.2 Canonical Form Representation:

Transforming the matrices of system (3) into canonical form has some advantages.

1) It reduces the problems of identifying a total of n(n+m+r) unknown parameters of A,B, and H to a problem with fewer unknown coefficients, n(m+r).

2) It permits one to obtain a set of transformations which map identification estimates into estimates of the unknown matrices of the system.

To accomplish these objectives, Luenberger's canonical form [26] - [29] is used.

The canonical form for system (3) is obtained by a linear transformation of the state vector to a new coordinate system. This transformation is defined by

$$w(k) = P\, x(k) \tag{3.2.1}$$

where,

$x(k)$ is the old state vector,

$w(k)$ is the state vector in the new coordinate system, and

P is the matrix producing linear mapping.

The fundamental assumptions imposed on system (3) are

1) The pair [A,B] is completely controllable.

2) The pair [H,A] is completely observable.

Assumption 1) implies that the nx(nr) controllability matrix defined by

$$U = [B, AB, \ldots\ldots, A^{n-1} B] \tag{3.2.2}$$

has rank n.

Assumption 2) implies that the nx(mn) observability matrix defined by

$$V = [H^T, A^T H^T, \ldots\ldots, A^{n-1^T} H^T] \tag{3.2.3}$$

has rank n.

The selection procedure for matrix P entails the choice of n independent columns of the observability matrix V which are ordered as shown below:

$$P^T = [h_1, A^T h_1, \ldots, A^{P1-1^T} h_1, h_2, A^T h_2, \ldots, A^{P2-1^T} h_2, h_3, \ldots, A^{Pm-1^T} h_m] \tag{3.2.4}$$

where H^T has been partitioned as

$$H^T = [h_1, h_2, \ldots\ldots\ldots h_m]. \tag{3.2.5}$$

The selection procedure therefore involves the following sequence of steps [26]:

1) Selecting the first column of V, h_1.

2) Selecting another column vector $A^T h_1$ form V. Retain this column vector if it is linearly independent of h_1.

3) Any selected new column vector must be of the form $A^{iT} h_j$, where all the lower powers of A^{iT}, postmultiplied by h_j, had already been previously retained. Retain the new column vector if it is linearly independent of all previously selected column vectors.

4) The selection procedure terminates when n linearly independent column vectors have been found. Arrange the n column vectors in proper order to obtain matrix P^T.

Since P^T is a square matrix, the p_i's (3.2.4) define the number of consecutive independent columns uninterrupted by a dependent one, and are referred to as observability subindices. They must satisfy the following equation.

$$\sum_{i=1}^{m} p_i = n \qquad (3.2.6)$$

The selection may result in a case in which not all the columns of H^T will appear in the expression P^T. Then the corresponding output measurement components will not provide useful information on the canonical form of the system, and may be dropped from the identification algorithm. However, it will be assumed, without any loss of generality, that all the columns of H^T appear in P^T.

Now through the change of coordinates $w(k) = P\, x(k)$, system (3) can be transformed into a canonical form which will meet the identification

objectives stated earlier:

$$\begin{aligned} \bar{w}(k+1) &= \bar{A}w(k) + \bar{B}u(k) \\ y(k) &= \bar{H}w(k) + v(k) \end{aligned} \tag{3.2.7}$$

where

$$\bar{A} = PAP^{-1} = \begin{array}{c} \begin{matrix} \quad p_1 \quad\quad p_2 \end{matrix} \\ \left[\begin{array}{cc} \begin{matrix} 0 & \cdot & \cdot \\ \cdot & I & \cdot \\ \cdot & \cdot & \cdot \\ xx & x & xx \end{matrix} & \begin{matrix} \cdot \\ 0 \\ \cdot \\ x \end{matrix} \\ \begin{matrix} 0 & 0 & I \\ xx & x & xx \end{matrix} & \begin{matrix} 0 \\ \cdot \\ x \end{matrix} \\ \begin{matrix} 0 & 0 & I \\ xx & x & xx \end{matrix} & \begin{matrix} 0 \\ \cdot \\ x \end{matrix} \end{array}\right] \end{array} \begin{matrix} p_1 \\ p_2 \\ p_m \end{matrix} \tag{3.2.8}$$

$$\bar{B} = PB = [\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_r] \tag{3.2.9}$$

$$\bar{\Pi} = \Pi P^{-1} = \begin{array}{c} \begin{matrix} \quad p_1 \quad\quad p_2 \end{matrix} \\ \left[\begin{array}{cc} \begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} \end{array}\right] \end{array} \begin{matrix} p_1 \\ p_m \end{matrix} \tag{3.2.10}$$

The matrix in (3.2.8) is seen to be composed of blocks of phase-variable submatrices, situated along the main diagonal. The m rows of x's in $\bar{A}$ correspond to unknown elements and occur in rows p_1, p_1+p_2, ..., $p_1+p_2+\ldots+p_m$. In general, matrix $\bar{B}$ assumes no special form.

Given the above canonical form, one needs to identify only mxn elements of $\bar{A}$ and nxr elements of $\bar{B}$, a total of n(m + r) unknown coefficients of system parameters.

There is another matrix, that is the dual of matrix P, which transforms system (3) into a canonical form [26]. This canonical form is derived by linear transformation of the state vector to a new coordinate system.

$$z(k) = Q\, x(k) \tag{3.2.11}$$

The matrix Q is obtained from the controllability matrix U (3.2.2) via a selection procedure similar to that used for matrix P. Thus, n independent columns of U are chosen and reordered as follows:

$$Q = [b_1,\ Ab_1,\ldots,\ A^{q_1-1}b_1, b_2,\ldots,A^{q_2-1}b_2, b_3,\ldots,A^{q_r-1}b_r] \tag{3.2.12}$$

where B has been partitioned as

$$B = [b_1, b_2, \ldots\ldots, b_r] \tag{3.2.13}$$

Since Q is a square matrix, the q's in (3.2.12) represent the number of consecutive independent columns uninterrupted by a dependent one,

and are referred to as controllability subindices. They must satisfy the following equation

$$\sum_{i=1}^{r} q_i = n \qquad (3.2.14)$$

As with matrix P^T (3.2.4), we assume, for the sake of generality that all of the columns of B appear in the expression for Q. In special situations, in which one or more columns of B do not appear in Q, the corresponding components of the vector input signal provide no useful information about the canonical form and may be dropped from the identification algorithm.

3.3 Non-Parametric Representation of the System

System (3.2.7) is in a state space parametric form, now it will be transformed into a non-parametric form.

Assuming that the initial conditions, (of equation (3.2.7)), are zeros, the state of system (3.2.7) at time k is given by:

$$w(k) = \sum_{i=0}^{\infty} \overline{A}^{i} \overline{B}\, u(k - i - 1)$$

which can be rewritten as follows:

$$w(k) = \sum_{i=0}^{L-1} \bar{A}^{i} \bar{B}\, u(k-i-1) + \sum_{i=L}^{\infty} \bar{A}^{i} \bar{B}\, u(k-i-1) \qquad (3.3.1)$$

where, L is called the identification interval, which is a function of the observability and controllability subindices defined in section 3.2, and is given by, [21],

$$L = \max[p_i] + \max[q_j] \qquad i=1,2,\ldots,m;\ j=1,2,..,r \qquad (3.3.2)$$

We now define the following (Lr) - dimensional auxiliary vectors:

$$\Theta_i^T = [\bar{h}_i^T\bar{B},\ \bar{h}_i^T\bar{A}\bar{B},\ldots,\bar{h}_i^T\bar{A}^{L-1}\bar{B}] \qquad i=1,2,\ldots,m \qquad (3.3.3)$$

$$U_{(k)}^{T(k+L-1)} = [u^T(k+L-1),\ u^T(k+L-2),\ \ldots,u^T(k)] \qquad (3.3.4)$$

where, $\bar{H}$ has been partitioned as

$$\bar{H}^T = [\bar{h}_1,\ \bar{h}_2,\ \ldots,\bar{h}_m]. \qquad (3.3.5)$$

Using equations (3.3.1), (3.3.3), and (3.3.4) with equation (3.2.7); the output of system (3.2.7) at time (k+L) can be rewritten as

$$y_i(k+L) = U_{(k)}^{T(k+L-1)}\,\Theta_i + \varepsilon_i(k-1) + v_i(k+L) \qquad (3.3.6)$$

where

$$\varepsilon_i(k-1) = \bar{h}_i^T \sum_{i=0}^{\infty} \bar{A}^{i+L} \bar{B}\, u(k-i-1) \qquad i=1,2,\ldots,m$$

or in a more compact form as

$$y_i(k+L) = U_{(k)}^{T_{(k+L-1)}} \Theta_i + e_i$$

where

$$e_i = \varepsilon_i(k-1) + v_i(k+L) \qquad i=1,2,\ldots,m \qquad (3.3.7)$$

Equation (3.3.7) represents system (3.2.7) in its non-parametric form. It also represents the impulse response of system (3.2.7) where; the elements Θ_i's represent the weighting functions of the system, and e_i may represent a noise imposed on the ith output.

It can be seen that as only L[$L = \max[p_i] + \max[q_j]$, equation (3.3.2)] values of the impulse response are used in equation (3.3.7), not all information on the process dynamics is included in estimating the parameter vectors Θ_i's. Therefore equation (3.3.7) gives an approximation to the exact impulse response of system (3.2.7); and hence the estimates of system matrices $\bar{A}$ and $\bar{B}$, recovered from the estimates of the parameter vectors [section 3.5], will always have a bias.

In order to improve the estimates of system matrices $\bar{A}$ and $\bar{B}$ the value of L, defined above, will be increased. This means that system (3.3.7) will give a better approximation for system (3.2.7) as more values of the impulse response will be used in equation (3.3.7).

The value of L will be;

$$L \geq \max[p_i] + \max[q_j] \qquad i=1,2,\ldots,m;\ j=1,2,\ldots,r$$

It is clear that increasing the value of L has certain advantages since this will increase the length of the parameter vectors Θ_i's and hence better estimates of the system matrices will be obtained. On the other hand, increasing L will increase the computation effort. Thus, a compromise has to be made between the degree of accuracy required and the computation effort. Having transformed the system into its non-parametric form (3.3.7), next a normalized stochastic approximation algorithm will be derived to identify the parameter vectors Θ_i's.

3.4 The Identification Algorithm:

The system (3.2.7) has been now transformed into a set of m non-parametric single-output multi-input subsystems (3.3.7). In this section, first a normalized stochastic approximation algorithm will be developed for the identification of the single-output multi-input systems. Then, the parameter vectors Θ_i's, i=1,2,...,m (3.3.3) of the subsystems (3.3.7) will be identified by applying a set of the single-output multi-input identification algorithms.

3.4.1 The Normalized Mean-Square Error Criterion:

Consider the following identification problem [4]. Defining the system to be identified as S, the input to the system U is a q-vector sequence and y is a scalar output sequence, both are assumed to be known precisely.

Given a model

$$y_m = U^T\Theta$$

select the vector Θ such that the model reproduces the system output as closely as possible in some specified sense.

Our objective will be to select the parameter Θ to minimize the following criterion

$$f(\Theta) = E\left[\frac{e^2}{||U||^2}\right] . \tag{3.4.1}$$

where

$$e = y_m - y = U^T\Theta - y$$

The choice of a normalized mean-square error criterion is motivated by the fact that it leads to a sequential estimation scheme with much more desirable convergence properties than does the mean-square error criterion [4], [14].

The criterion can be given a simple geometric interpretation. Let A_Θ designate the space spanned by the parameter vector Θ. Let us suppose that U and y are specified. Then the relation

$$U^T\Theta - y = e$$

defines a family of hyperplanes in A_Θ with the parameter e. These hyperplanes are parallel to each other and normal to U. The normal distance to the hyperplanes is ;

$$\frac{e}{||U||}$$

To prove this note that, if we set all elements of Θ except Θ_1 equal to zero, then the hyperplane intersects the Θ_1 axis at

$$\Theta_1 = \frac{e + y}{u_1}$$

where u_1 is the first element of U. Hence, the distance between the hyperplane of error e and the zero-error hyperplane is

$$\frac{e}{u_1}$$

Let n_1 designate a unit vector in the Θ_1 direction. Then the vector distance from the zero-error hyperplane of error e along the Θ_1 axis is

$$\frac{e}{u_1} \quad n_1$$

The normal distance measured in the direction of U is the projection of this vector on U, which is

$$\frac{e}{u_1} \quad n_1^T \frac{U}{||U||} = \frac{e}{||U||}$$

Therefore, the normalized mean-square error criterion, is that of selecting the parameter vector Θ such that the expected square distance to the hyperplane corresponding to Θ from the zero-error hyperplane is a minimum.

3.4.2 The Single-Output Multi-Input Identification Algorithm:

Using the normalized mean-square error criterion discussed in subsection 3.4.1 a single-output multi-input normalized stochastic approximation algorithm is given.

The reason for using the normalized mean-square error criterion in the identification algorithm is that stochastic approximation

identification algorithms with ordinary mean-square error criterion have the disadvantage that although asymptotic convergence is assured very little can be done to control the initial convergence properties. But those with normalized mean-square error criterion have very favourable characteristics in this regard and initial convergence is assured [4], [14].

Now, recall equation (3.3.7) and rewrite the equation of one row only;

$$y(k+L) = U_{(k)}^{T_{(k+L-1)}} \Theta + e \qquad (3.4.2)$$

where

$$e = \varepsilon(k-1) + v(k+L)$$

This equation may represent a single-output multi-input system where $y(k+L)$ is a scalar output, $U_{(k)}^{T_{(k+L-1)}}$ is the input vector of length L, defined in (3.3.4), Θ is the parameter vector of length L and e is a noise term added to the output of the system.

Our objective is to identify the parameter vector Θ, of equation (3.4.2), from the set of input output data.

Define $\hat{\Theta}(k)$ as the estimate of the parameter vector Θ at the kth iteration step which minimize the normalized mean-square error criterion, equation (3.4.1). This estimate $\hat{\Theta}(k)$ can be obtained recursively, by analogy to the standard stochastic approximation algorithm [equation (2.2), chapter 2], [4], [7], and [14], using the following single-output multi-input algorithm;

$$\hat{\Theta}(k+L) = \hat{\Theta}(k-1) + \gamma\,(k-1)\,\frac{U_{(k)}^{(k+L-1)}}{\left\| U_{(k)}^{(k+L-1)} \right\|^2}\left[y(k+L) - U_{(k)}^{T_{(k+L-1)}}\hat{\Theta}(k-1)\right]$$

$$k=1,\ L+2,\ldots\ldots \qquad (3.4.3)$$

Fig. (3.1) shows the scheme which illustratesthe identification algorithm given above.

The above algorithm uses only measurements of the inputs and output and does not require knowledge of the statistics of measurement noise. The resulting estimate $\hat{\Theta}(i(L+1))$, $i=1,2,\ldots$ converges in the mean-square sense to the true value of the parameter vector Θ if the sequence $\gamma(k)$ satisfies the conditions (2.3) chapter 2 and the initial estimate $\hat{\Theta}(0)$ satisfies

$$E\{\left\| \hat{\Theta}(0) \right\|^2\} < \infty .$$

3.4.3 The Identification Algorithm For Multivariable Systems:

Now, starting with equation (3.2.7), we make the following assumptions:

(1) System (3.2.7) is completely controllable. Complete observability is implicit in (3.2.8) and (3.2.10).

(2) The vector random sequences $v(k)$ and $u(k)$ all have finite moments up to the fourth order.

Defining $\hat{\Theta}_i$ as the identification estimates of the unknown

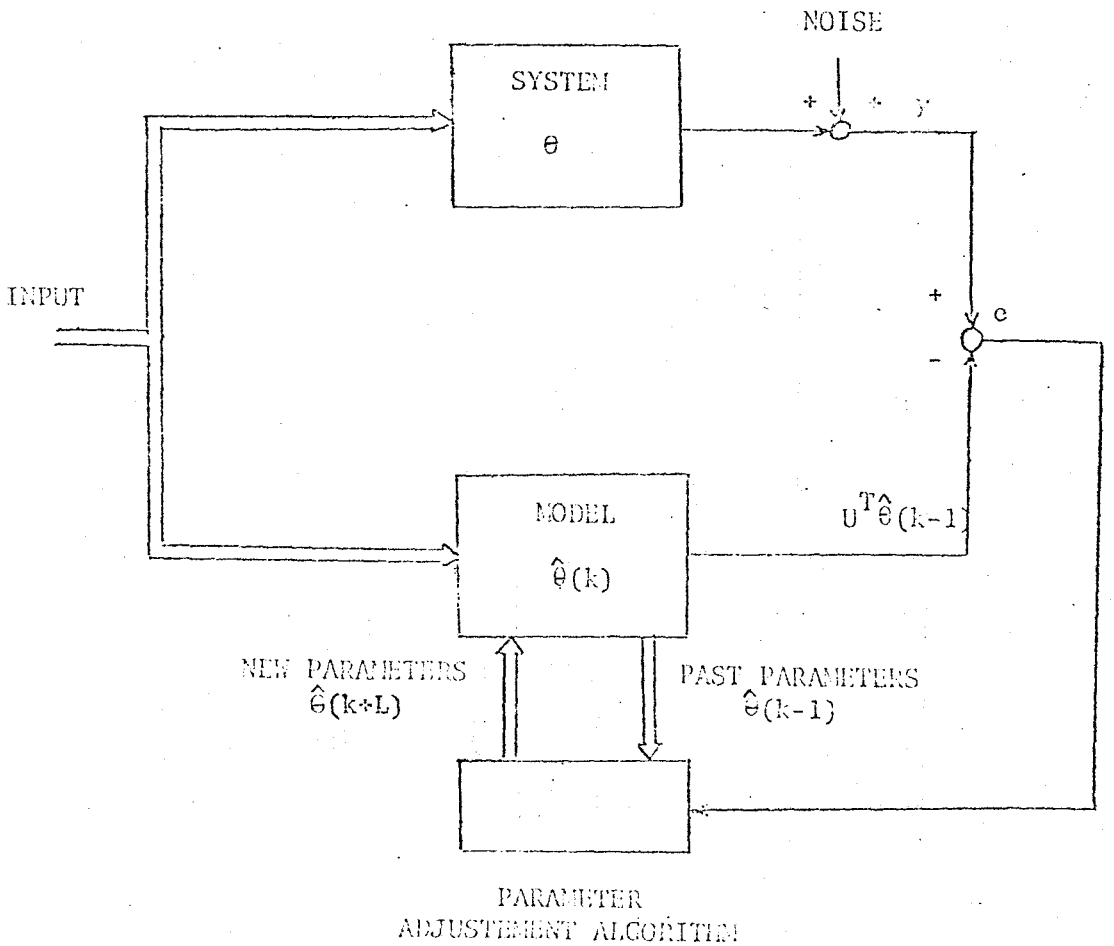

Fig. 3.1 Scheme for System Identification
for Multi-Input Single-Output Systems

parameter vectors Θ_i, i=1,2,...., m of equation (3.3.7), which minimize the following set of normalized mean-square error criterion.

$$f(\hat{\Theta}_i) = E\left[\frac{e_i^2}{U_{(k)}^{T(k+L-1)}\ U_{(k)}^{(k+L-1)}}\right] \qquad (3.4.4)$$

where

$$e_i = y_i\ (k+L) - U_{(k)}^{T(k+L-1)}\ \Theta_i \qquad i = 1,2,....,m$$

Then, using m single-output multi-input identification algorithms, (equation (3.4.3) subsection 3.4.2), the estimates $\hat{\Theta}_i$, i=1,2,...,m are computed recursively by means of the following set of normalized stochastic approximation algorithms;

$$\hat{\Theta}_i(k+L) = \hat{\Theta}_i(k-1) + \nu\frac{(k-1)}{(L+1)}\ \frac{U_{(k)}^{(k+L-1)}}{||U_{(k)}^{(k+L-1)}||^2}\ [y_i(k+L) - U_k^{T(k+L-1)}\ \hat{\Theta}_i(k-1)]$$

$$k=1, L+2,\ 2L+3,...,;\ i=1,2,...,m \qquad (3.4.5)$$

A quick glance shows that these algorithms are not dependent upon knowledge of the statistics of the noise vector v(k). If the following conditions:

$$\lim_{i\to\infty}\ \nu(i) = \infty\ , \qquad \sum_{i=1}^{\infty}\ \nu(i) = \infty\ , \qquad \sum_{i=1}^{\infty}\ \nu^2(i) < \infty$$

and $\quad E\{|| \hat{\theta}_i(0)||^2\} < \infty \qquad i=1,2,.....,m$

are satisfied, estimates (3.4.5) converge in the mean-square sense to the true parameter vectors θ_i, i=1,2,....m.

The main advantage of the proposed algorithm is that it requires simple computation, that is because it uses stochastic approximation method. To show this, the number of arithmetic operations per iteration needed in the proposed algorithm and in Sen and Sinha's algorithm [31] are given in table [3.1], where L is given the value of 2n in order to make the comparison.

3.5 Recovering the Estimates of System Parameters

After identifying the unknown parameter vectors θ_i, i=1,2,.., m in equation (3.3.3), we need a set of transformations that map the components of θ_i into the matrices $\bar{A}$ and $\bar{B}$. These are found in two steps [21]. First, we find the transformation which yields $\bar{B}$ and then the transformation that maps θ_i into $\bar{A}$, whereby we utilize the matrices P and Q, defined in section 3.2. The final result is a matrix transformation which, in addition to having a tractable form, also provides a valuable test for determining the observability and controllability subindices, as well as the order of the system, should they be unknown.

TABLE 3.1

COMPARISON OF THE ARITHMETIC OPERATIONS

Algorithm †	No. of additions and subtractions per iteration	No. of multiplications and divisions per iteration	Matrix inversion per iteration
Proposed Alg. ††	$(2nr+1)m+1$	$4nmr+1$	None
Sen and Sinha's Alg.	$m[n^2(mr+1)^2+2mn(mr+1)-m+1-n(mr+1)]$	$mn(m+1)[3n(mr+1)+2m+2]$	One matrix inversion (mxm)

† n is the order of the system
r is the number of inputs
m is the number of outputs.

†† L is given the value $L = 2n$

3.5.1 Recovering $\bar{B}$ from Θ_i, i=1,2,....,m

We partition the (nxr)- dimensional matrix $\bar{B}$ in equation (3.2.9) as follows:

$$\bar{B} = \begin{bmatrix} \bar{B}^1 \\ \cdots\cdots \\ \bar{B}^2 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ \bar{B}^m \end{bmatrix} \begin{matrix} p_1 \\ \\ p_2 \\ \\ \\ \\ p_m \end{matrix} \qquad (3.5.1)$$

By direct substitution of the canonical form, equations (3.2.8) to (3.2.10), and using definition of Θ_i, equation (3.3.3), we get;

$$\bar{B} = \begin{bmatrix} \Theta_1^1 & \Theta_1^2 & \cdots & \Theta_1^r \\ \Theta_1^{(p_1-1)r+1} & \cdots & & \Theta_1^{p_1 r} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \Theta_2^1 & \cdots & & \Theta_2^r \\ \Theta_2^{(p_2-1)r+1} & \cdots & & \Theta_2^{p_2 r} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ & \vdots & & \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \Theta_m^1 & \Theta_m^2 & \cdots & \Theta_m^r \\ \Theta_m^{(p_m-1)r+1} & \cdots & & \Theta_m^{p_m r} \end{bmatrix} \begin{matrix} \\ p_1 \\ \\ \\ p_2 \\ \\ \\ \\ p_m \end{matrix} \qquad (3.5.2)$$

where the superscript j on θ_i^j corresponds to the position of an element in the (Lr)- dimensional vector θ_i.

3.5.2 Recovering $\bar{A}$ from θ_i, i=1,2,...,m

In equation (3.2.8), matrix $\bar{A}$ was obtained from A by similarity transformation P, i.e.,

$$\bar{A} = P A P^{-1} \tag{3.5.3}$$

and

$$\bar{A} = P A Q [PQ]^{-1} \tag{3.5.4}$$

By substituting the elements of P and Q from equations (3.2.4) and (3.2.12) into equation (3.5.4) and performing the multiplication, one obtains: (For convenience equations (3.5.5) and (3.5.7) are photocopied from [21].)

$$\bar{A} = \begin{bmatrix} h_1^T A b_1 & \cdots & h_1^T A^{q_1} b_1 & h_1^T A b_2 & \cdots & h_1^T A^{q_r} b_r \\ h_1^T A^2 b_1 & \cdots & h_1^T A^{q_1+1} b_1 & h_1^T A^2 b_2 & \cdots & h_1^T A^{q_r+1} b_r \\ \vdots & & \vdots & \vdots & & \vdots \\ h_1^T A^{p_1} b_1 & \cdots & h_1^T A^{p_1+q_1-1} b_1 & h_1^T A^{p_1} b_2 & \cdots & h_1^T A^{p_1+q_r-1} b_r \\ h_2^T A b_1 & \cdots & h_2^T A^{q_1} b_1 & \cdots & & h_2^T A^{q_r} b_r \\ \vdots & & \vdots & & & \vdots \\ h_m^T A^{p_m} b_1 & \cdots & h_m^T A^{p_m+q_1-1} b_1 & \cdots & & h_m^T A^{p_m+q_r-1} b_r \end{bmatrix} \times$$

$$\begin{bmatrix} h_1^T b_1 & \cdots & h_1^T A^{q_1-1} b_1 & h_1^T b_2 & \cdots & h_1^T A^{q_r-1} b_r \\ h_1^T A b_1 & \cdots & h_1^T A^{q_1} b_1 & h_1^T A b_2 & \cdots & h_1^T A^{q_r} b_r \\ \vdots & & \vdots & \vdots & & \vdots \\ h_1^T A^{p_1-1} b_1 & \cdots & h_1^T A^{p_1+q_1-2} b_1 & h_1^T A^{p_1-1} b_2 & \cdots & h_1^T A^{p_1+q_r-2} b_r \\ h_2^T b_1 & \cdots & h_2^T A^{q_1-1} b_1 & \cdots & & h_2^T A^{q_r-1} b_r \\ \vdots & & \vdots & & & \vdots \\ h_m^T A^{p_m-1} b_1 & \cdots & h_m^T A^{p_m+q_1-2} b_1 & \cdots & & h_m^T A^{p_m+q_r-2} b_r \end{bmatrix}^{-1} \qquad (3.5.5)$$

Since the input-output relationship of a linear system remains invariant under a similarity transformation, we have

$$h_i^T A^j b_k = \bar{h}_i^T \bar{A}^j \bar{b}_k , \quad i = 1,2,\ldots\ldots,m;\ k=1,2,\ldots,r;j = 0,1,2,.. \qquad (3.5.6)$$

The elements of the matrices in equation (3.5.5) can therefore be replaced by the elements of vectors θ_i from eq. (3.3.3) i.e.,

$$\bar{A} = \begin{bmatrix} \theta_1^{r+1} & \cdots & \theta_1^{q_1 r+1} & \theta_1^{r+2} & \cdots & \theta_1^{q_r r+r} \\ \theta_1^{2r+1} & \cdots & \theta_1^{(q_1+1)r+1} & \theta_1^{2r+2} & \cdots & \theta_1^{q_r r+2r} \\ \vdots & & \vdots & \vdots & & \vdots \\ \theta_1^{p_1 r+1} & \cdots & \theta_1^{(q_1+p_1-1)r+1} & \theta_1^{p_1 r+2} & \cdots & \theta_1^{(q_r+p_1-1)r+r} \\ \theta_2^{r+1} & \cdots & \theta_2^{q_1 r+1} & \theta_2^{r+2} & \cdots & \theta_2^{q_r r+r} \\ \vdots & & \vdots & \vdots & & \vdots \\ \theta_m^{p_m r+1} & \cdots & \theta_m^{(q_1+p_m-1)r+r} & \theta_m^{p_m r+2} & \cdots & \theta_m^{(q_r+p_m-1)r+r} \end{bmatrix} \times \tag{3.5.7}$$

$$\begin{bmatrix} \theta_1^{1} & \cdots & \theta_1^{(q_1-1)r+1} & \theta_1^{2} & \cdots & \theta_1^{q_r r} \\ \theta_1^{r+1} & \cdots & \theta_1^{q_1 r+1} & \theta_1^{r+2} & \cdots & \theta_1^{q_r r+r} \\ \vdots & & \vdots & \vdots & & \vdots \\ \theta_1^{(p_1-1)r+1} & \cdots & \theta_1^{(q_1+p_1-2)r+1} & \theta_1^{(p_1-1)r+2} & \cdots & \theta_1^{(q_r+p_1-1)r} \\ \theta_2^{1} & \cdots & \theta_2^{(q_1-1)r+1} & \theta_2^{2} & \cdots & \theta_2^{q_r r} \\ \vdots & & \vdots & \vdots & & \vdots \\ \theta_m^{(p_m-1)r+1} & \cdots & \theta_m^{(q_1+p_m-2)r+1} & \theta_m^{(p_m-1)r+2} & \cdots & \theta_m^{(q_r+p_m-1)r} \end{bmatrix}^{-1}$$

Equations (3.5.2) and (3.5.7) thus provide the set of transformations which map the Θ_i's into $\bar{B}$ and $\bar{A}$. So estimates of $\bar{A}$ and $\bar{B}$ are found by substituting estimates (3.4.5) into equations (3.5.7) and (3.5.2), respectively.

3.5.3 Determination of Unknown Controllability and Observability Subindices of the System

The identification interval L in equation (3.3.2) is determined by the maximum values of the controllability and observability subindices, given in equations (3.2.12) and (3.2.4), respectively. In addition, placement of the components of Θ_i in equation (3.5.2) and (3.5.7) depends upon exact knowledge of the numerical values of p_i's and q_i's.

If those numerical values are known in advance, the identification algorithm of subsection 3.4.3 can be applied and then the system matrices $\bar{A}$ and $\bar{B}$ can be recovered using the procedure described in subsections 3.5.1 and 3.5.2.

In general, these numerical values are not readily available, and so they and also the order of the system when unknown, must be found. This situation will now be discussed. Define the product PQ as:

$$F = P\,Q \tag{3.5.8}$$

Using the definition in equations (3.2.4) and (3.2.12) for P and Q, matrix F may be written as

$$F = \underset{n \times n}{\begin{bmatrix} h_1^T \\ h_1^T A \\ \vdots \\ h_1^T A^{p_1-1} \\ h_2^T \\ \vdots \\ h_m^T A^{p_m-1} \end{bmatrix}} \left[b_1, Ab_1, \ldots, A^{q_1-1}b_1,\ b_2, \ldots,\ A^{q_r-1}b_r\right]_{nxn} \tag{3.5.9}$$

Assuming that the order of the system and/or the controllability or observability subindices are unknown, initial guesses of these parameters can be made to set up the identification algorithm (3.4.5). It is understood, though, that the order of the system and parameters p_i and q_i are related to each other by equations (3.2.6) and (3.2.14). Depending on the accuracy of these initial guesses, matrix F must satisfy the following conditions when identification estimates from equation (3.4.5) are used in equation (3.5.8).

Conditions for Order and Subindex Identification:

1. If the guess of the system order $n' > n$, or if the guesses of the controllability or observability subindices are wrong, i.e., the matrices Q or P in equations (3.2.12) and (3.2.4) are singular, matrix F will converge to a singular matrix as the identification estimates converge to Θ_i.

2. If the guesses of the system order and the controllability and observability subindices are correct, matrix F will always be nonsingular with probability one. Proof of these two conditions is given in [21].

The following procedure should therefore be observed in case of uncertainty as to the order of the system and controllability and observability subindices, when identification estimates from equation (3.4.5) are used in matrix F:

1. If the controllability and observability subindices are known (thus implying knowledge of the order of the system from equation

(3.2.6) or equation (3.2.14), one can use equation (3.3.2) to determine the identification interval L. Exact values of p_i's and q_i's can then be used in equations (3.5.2) and (3.5.7).

2. If the order of the system is given but the controllability and observability subindices are unknown, one should set L=2n, and arbitrarily assign the p_i's and q_i's, subject to constraints (3.2.6) and (3.2.14). If the set of values is chosen incorrectly, matrix F (3.5.9) converges to a singular matrix as the identification process (3.4.5) converges. The correct set of p_i's and q_i's may thus be found by varying these coefficients (keeping in mind that equations (3.2.6) and (3.2.14) must always be satisfied) until F becomes non-singular. The test for singularity can be performed on-line by monitoring the determinant of the right-hand matrix of equation (3.5.7) and noting whether it converges to zero with time.

3. When the order of the system is unknown, n' can initially be chosen sufficiently large to ensure that $n' \geq n$, where n is the real order of the system. Step 2 is then performed, using this n'. If $n' > n$, there can be no nonsingular matrix under all the possible combinations of p_i's and q_i's. The guessed order of the system n' is then reduced by one and step 2 is repeated. The process will end when $n' = n$, the true controllability and observability subindices being found at this last stage.

CHAPTER 4

RESULTS OF SIMULATION

4.1 Introduction:

In this chapter the normalized stochastic approximation algorithm for the identification of multivariable systems, developed in Chapter 3, will be applied to the identification of two multivariable systems under the realistic conditions of large noise-to-signal ratios up to 66.2% due to imperfect measurements of the outputs. System subindices will be assumed known in advance.

The first system has been identified using Sen and Sinha's algorithm [31] for noise-to-signal ratio of 7% and convergence rate and computational effort are given.

The effect of increasing the identification interval L, discussed in section 3.2 Chapter 3, on the accuracy of the parameter estimates and the computational effort is further examined. The number of arithmetic operations per iteration and the total computation time are given.

4.2 Example 1

The system used for simulation here is a 2-output 1-input 3rd order discrete-time system, which is described by the following

equations:

$$x(k+1) = \begin{bmatrix} 0.0 & 1.0 & 0.0 \\ 0.1 & 0.3 & 0.1 \\ 0.95 & 0.1 & 0.7 \end{bmatrix} x(k) + \begin{bmatrix} 0.12 \\ 0.36 \\ 0.20 \end{bmatrix} u(k)$$

$$z(k) = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} x(k) + v(k)$$

where;

$x(k)$, $u(k)$, $z(k)$, $v(k)$ are the state vector sequence, the scalar input sequence, the measurement output vector sequence and the output noise vector sequence respectively.

The scalar input sequence, $u(k)$ is a zero-mean white noise sequence with unit variance. Each of the two outputs is contaminated with a white noise sequence with variance adjusted to vary the noise-to-signal ratio (defined below) at each output.

The noise-to-signal ratio (N.S.R.) at the ith output is defined as,

$$\text{N.S.R. (i)} = \frac{\sigma_{v_i}}{\sigma_{y_i}}$$

where;

σ_{v_i} is the standard deviation of noise sequence at the ith output.

σ_{y_i} is the standard deviation of the ith output sequence.

The given system is in the required canonical form. Its controllability and observability subindices are easily obtained as;

$$p_1 = 2 \qquad p_2 = 1 \qquad q_1 = 3$$

Hence, the identification interval L is found to be, section 3.3 Chapter 3,

$$L \geq 5$$

The above system will now be identified for two values of L using the proposed algorithm, equation (3.4.5) Chapter 3, which will be updated after each sample interval.

First Case L = 5 :

In this case, the identification interval L has been assigned the minimum allowable value L = 5. Then the system has been simulated for three different noise-to-signal ratios of 7%, 23.4% and 66.2%. In each case the noise-to-signal ratio was kept the same at each of the two outputs.

Figures (4.1), (4.2) and (4.3) show the rate of convergence of the identification algorithm for the different noise-to - signal ratios. The error norm used in the figures is defined as:

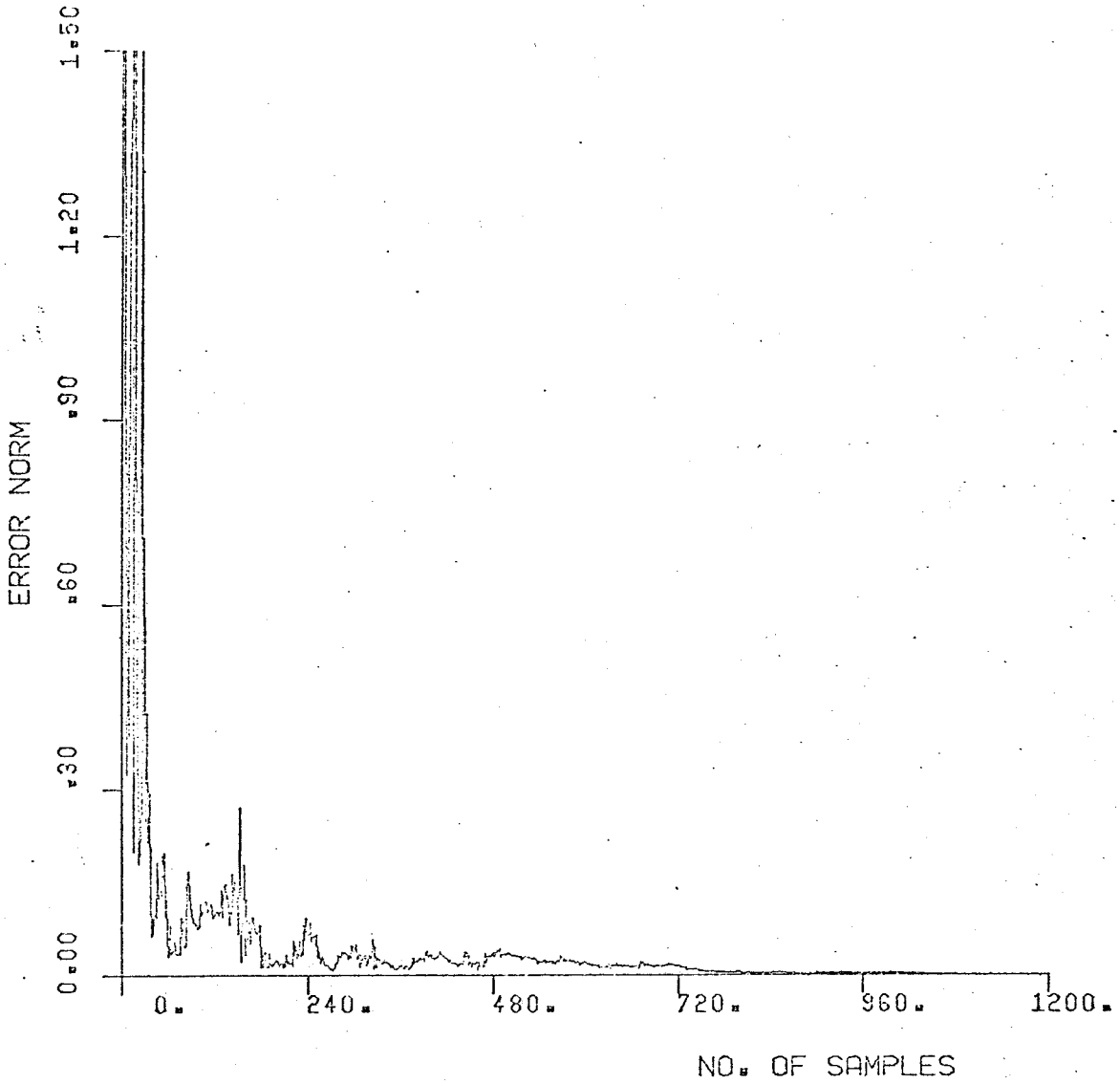

Fig. 4.1 Rate of Convergence of the Proposed Algorithm for Example 1 with L = 5 and White Noise of N.S.R.(at each output) = 7 %

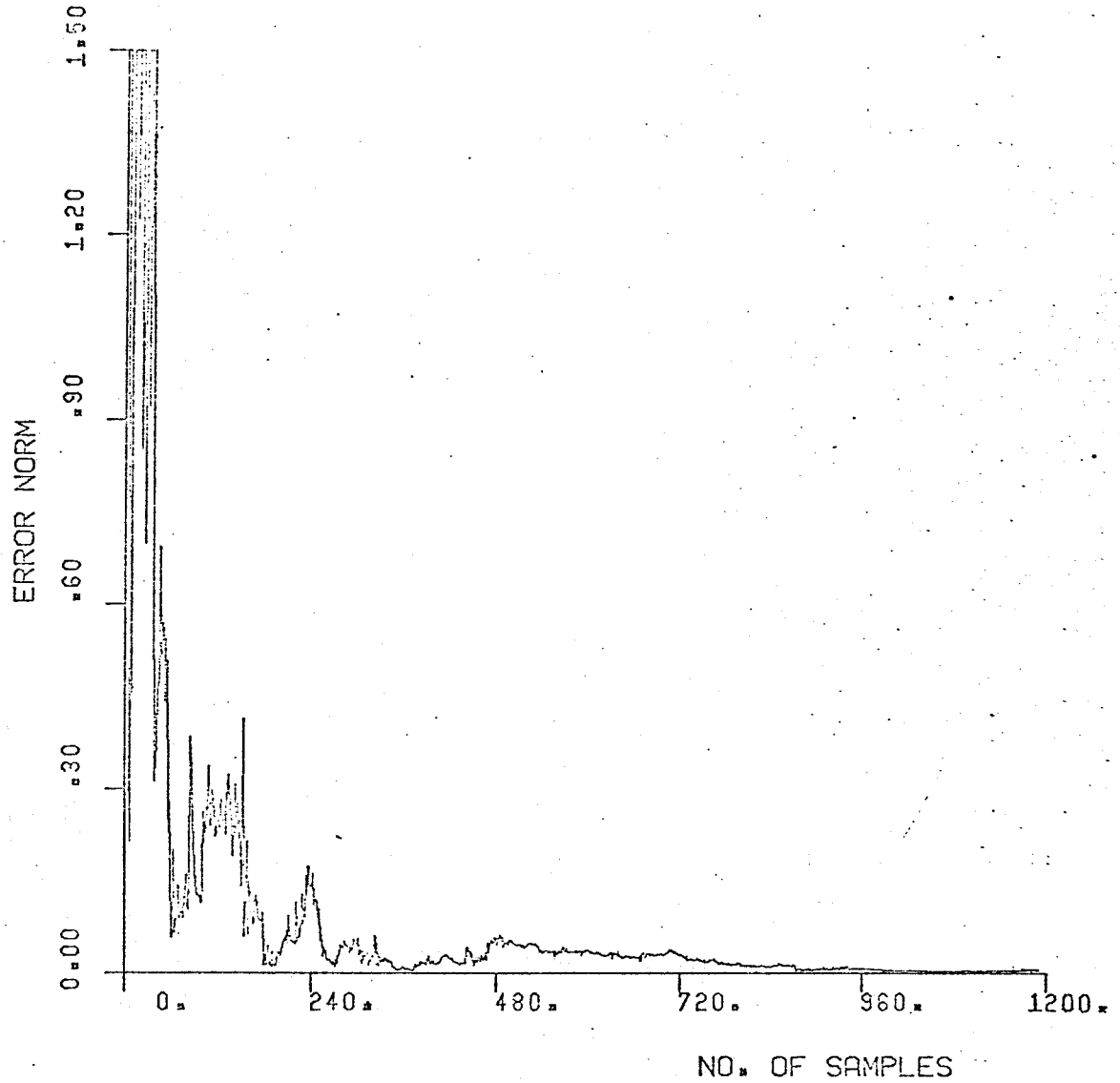

Fig. 4.2 Rate of Convergence of the Proposed Algorithm for Example 1 with L = 5 and White Noise of N.S.R.(at each output) =23.4 %

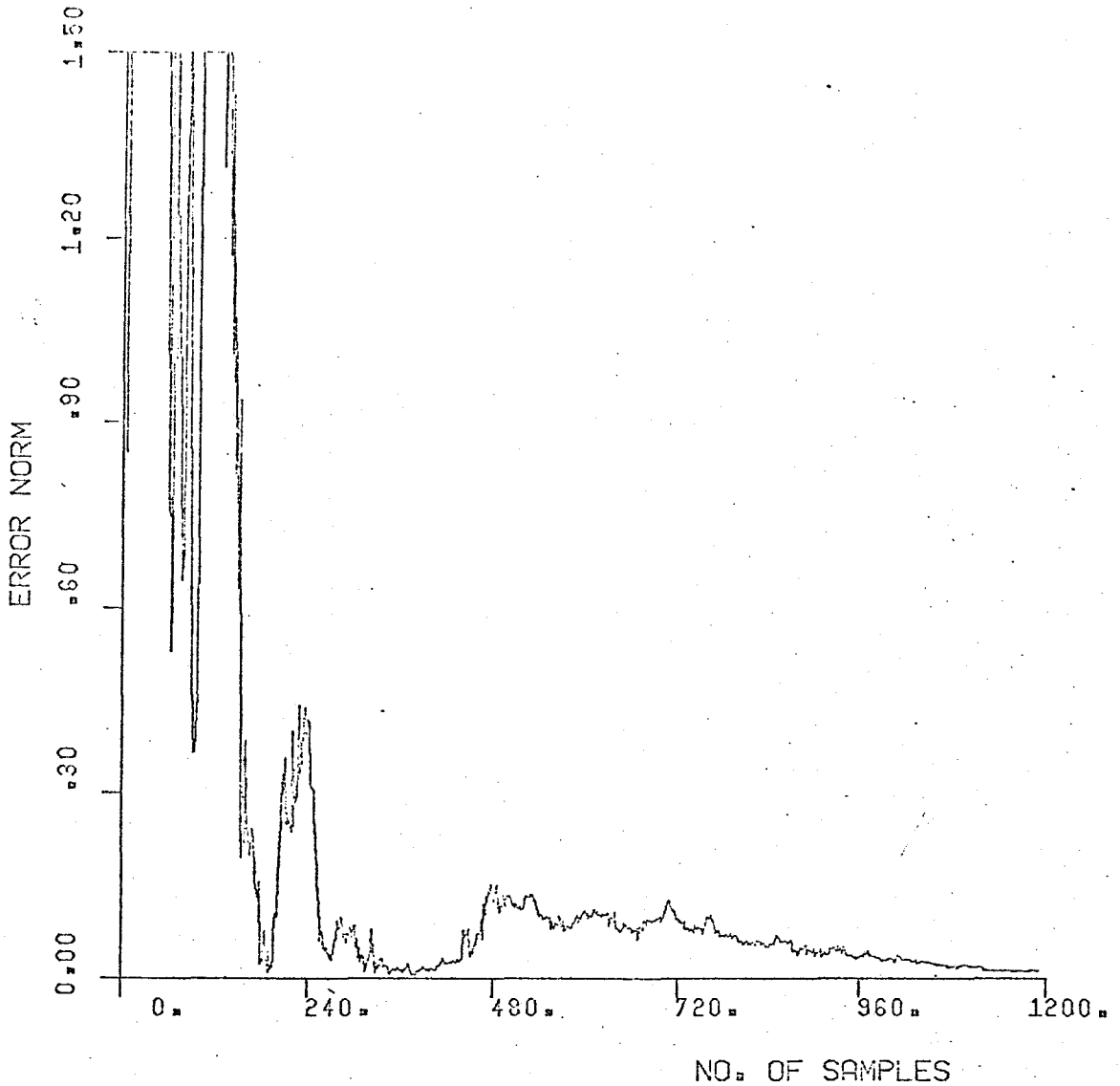

Fig. 4.3 Rate of Convergence of the Proposed Algorithm for Example 1 with L = 5 and White Noise of N.S.R.(at each output) = 66.2 %

$$\text{Error norm} = \frac{|| \hat{\theta}(k)-\theta ||^2}{|| \hat{\theta}(0)-\theta ||^2}$$

where θ is a vector formulated from all the parameters to be identified, that is;

$$\theta = [\ 0.1 \quad 0.3 \quad 0.1 \quad 0.95 \quad 0.1 \quad 0.7 \quad 0.12 \quad 0.36 \quad 0.2]$$

$\hat{\theta}$ (k) is the estimate of the vector θ at the kth interation and $\hat{\theta}$ (0) is the initial estimate which has been given the value of zero.

The estimates of system matrices for the three noise-to-signal ratios after 1200 iterations are given in Table (4.1).

Second Case L =10 :

In this case the identification interval L has been increased to L = 10. Then, the system was simulated with noise-to-signal ratio of 66.2% at each of the two outputs with zero initial estimate.

The convergence rate of the identification algorithm in this case is shown in Fig. (4.4). Also, the estimates of system matrices after 1200 iterations are given in Table 4.1.

To show the main features of the proposed algorithm, the above system has been simulated using Sen and Sinha's algorithm [31] for noise-to-signal ratio of 7% at each of the two outputs with zero initial estimate. The convergence rate of this algorithm is shown in Fig. (4.5) which indicates that the estimates of system parameters possess a bias. The number of arithmetic operations per iteration and the total computation time for 1200 iterations are compared for the

TABLE 4.1

ESTIMATES OF SYSTEM MATRICES FOR EXAMPLE 1

Algorithm	N.S.R.	Matrix A			Vector B
Proposed alg. with white noise and L = 5	7.0% (at each output)	0.000	1.000	0.000	0.120
		0.100	0.301	0.100	0.360
		0.951	0.100	0.700	0.200
	23.4% (at each output)	0.000	1.000	0.000	0.104
		0.112	0.298	0.095	0.361
		0.954	0.102	0.710	0.212
	66.2% (at each output)	0.000	1.000	0.000	0.116
		0.130	0.282	0.081	0.357
		0.977	0.121	0.731	0.237
Proposed alg. with white noise and L = 10	66.2% (at each output)	0.000	1.000	0.000	0.108
		0.117	0.307	0.110	0.362
		0.951	0.101	0.703	0.209

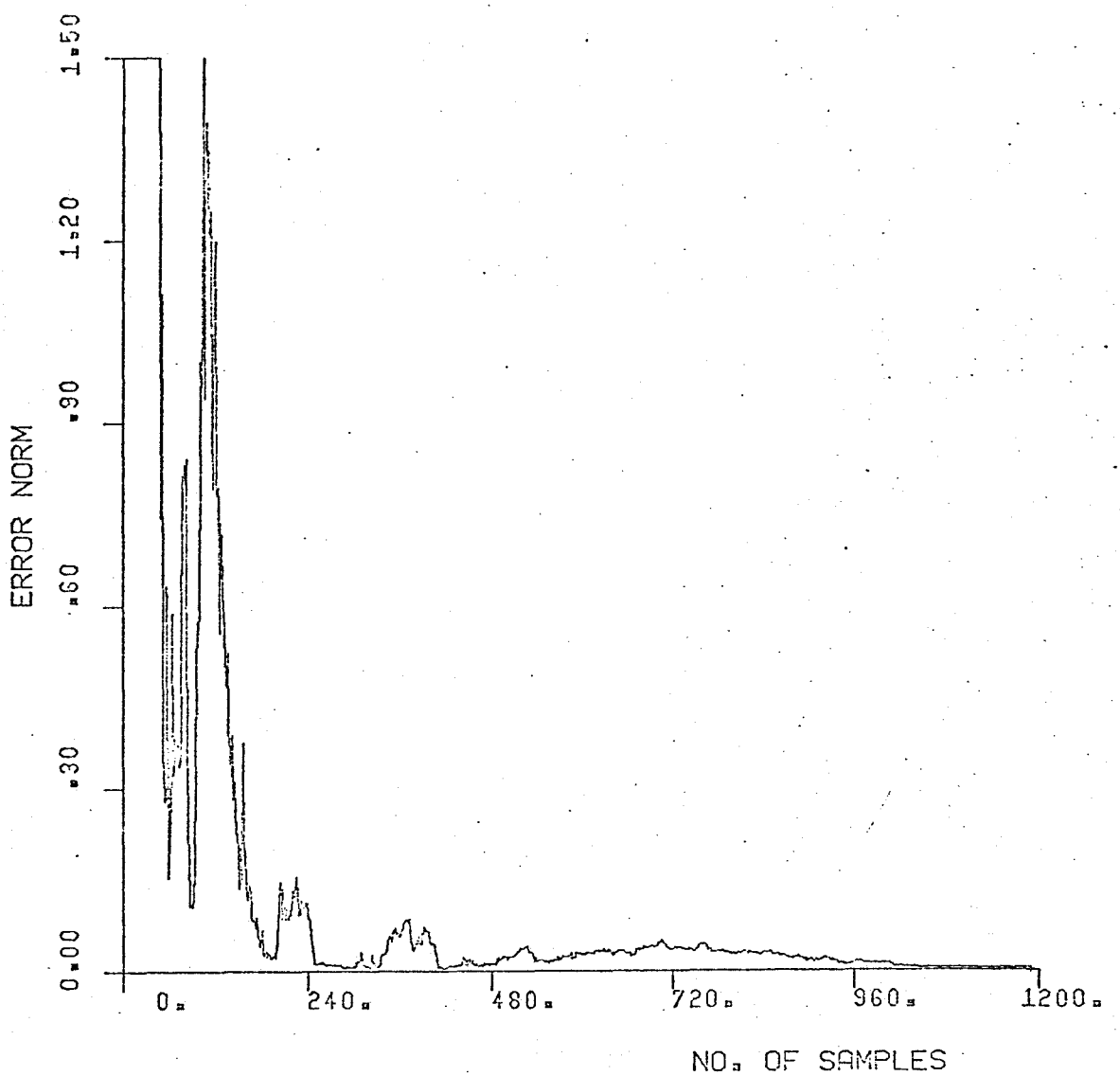

Fig. 4.4 Rate of Convergence of the Proposed Algorithm for Example 1 with L = 10 and White Noise of N.S.R.(at each output) = 66.2 %

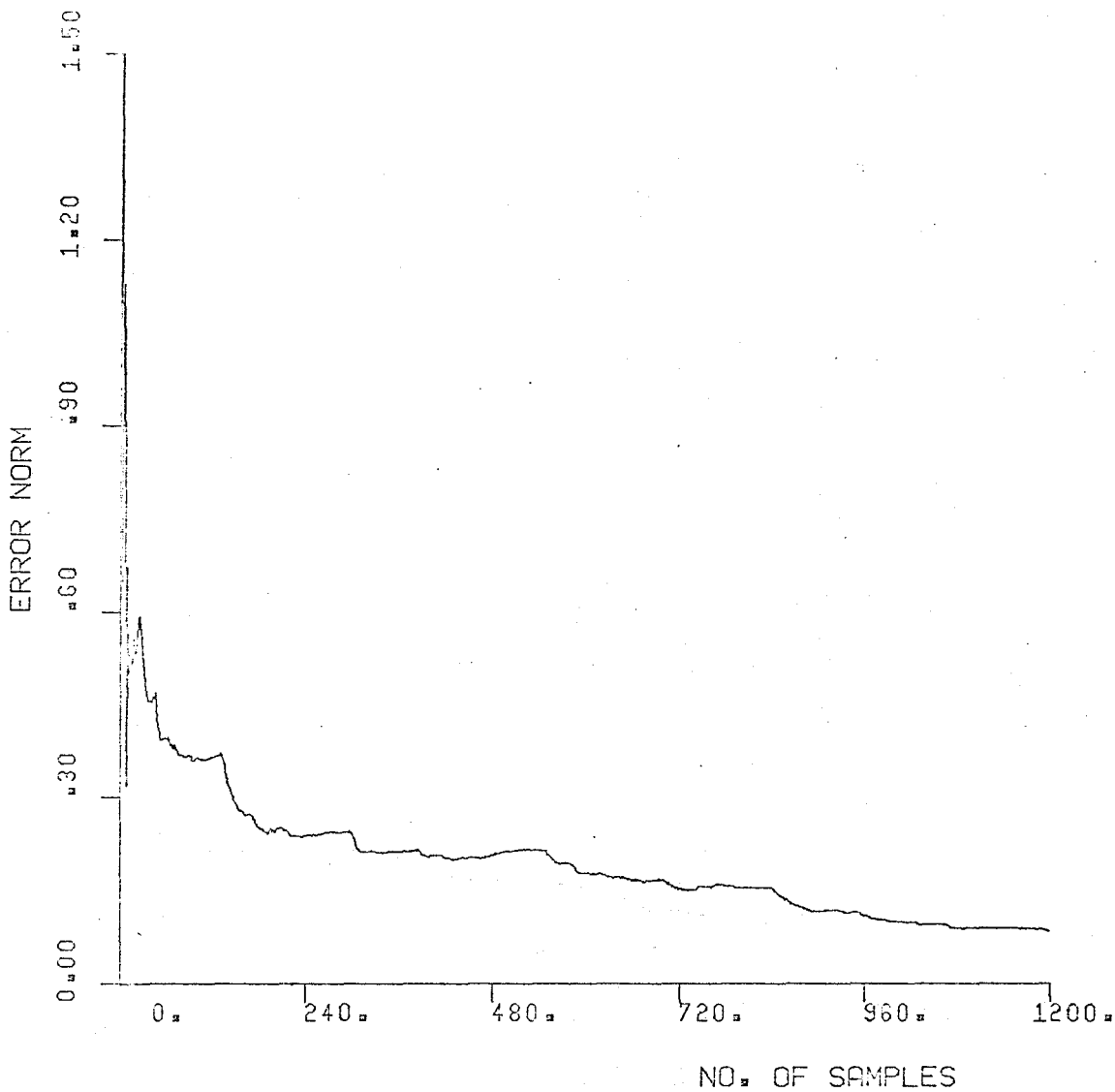

Fig. 4.5 Rate of Convergence of Sen and Sinha's Algorithm for Example 1 With White Noise of N.S.R.(at each output) = 7 %

proposed algorithm with Sen and Sinha's algorithm which is given in Table (4.2).

4.3 Example 2:

The system used for simulation in this example is a 2-input 2-output 5th order discrete time system, which is described by the following equations:

$$x(k+1) = \begin{bmatrix} 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ -0.125 & 0.75 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.1 & -0.5 & 1.25 \end{bmatrix} x(k) + \begin{bmatrix} -0.5 & 0.3 \\ 0.65 & 1.0 \\ 0.2 & -0.7 \\ 0.25 & 0.4 \\ 0.125 & 0.5 \end{bmatrix} u(k)$$

$$z(k) = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \end{bmatrix} x(k) + v(k)$$

where,

x(k), u(k), z(k), v(k) are the state vector sequence, the input vector sequence, the output measurement vector sequence and the output noise vector sequence respectively.

Each of the two inputs is a zero-mean white noise sequence with unit variance.

The system described above is in the required canonical form.

TABLE 4.2

THE COMPUTATION EFFORT FOR EXAMPLE 1

Algorithm	No. of additions and subtractions per iteration	No. of multiplications and divisions per iteration	Total computation time for 1200 iterations (sec.)	Matrix inversion per iteration
Proposed Alg. L = 5	13	21	0.689	None
Proposed Alg. L = 10	23	41	1.283	None
Sen & Sinha 's Alg.	538	594	18.270	2x2 matrix

Its controllability and observability subindices are easily obtained as;

$$p_1 = 2, \quad p_2 = 3, \quad q_1 = 2, \quad q_2 = 3$$

Hence, the identification interval L is found to be, [section 3.3 Chapter 3]

$$L \geq 6$$

The above system will now be identified for a three different cases using the proposed algorithm, equation (3.4.5) Chapter 3, which will be updated after each sample interval.

(1) White Noise with L = 6:

In this case the identification interval L has been assigned the minimum allowable value L = 6. Each of the two outputs has been contaminated with a zero-mean white noise sequence. Then the above system is identified using the proposed algorithm for a noise-to-signal ratio of 12% at each of the two outputs.

The rate of convergence of the identification algorithm is shown in Fig. (4.6) and the estimates of system matrices after 1200 iterations are given in Table (4.3).

(2) White Noise with L = 9:

In this case the identification interval L is increased to L = 9. The output noise sequences are the same as in case (1).

TABLE 4.3

ESTIMATES OF SYSTEM MATRICES FOR EXAMPLE 2

Algorithm	N.S.R.	Matrix A	Matrix B
Proposed alg. with white noise and L = 6	12% (at each output)	$\begin{bmatrix} 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ -0.134 & 0.761 & 0.000 & 0.009 & 0.018 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \\ 0.004 & 0.001 & 0.110 & -0.521 & 1.275 \end{bmatrix}$	$\begin{bmatrix} -0.435 & 0.309 \\ 0.640 & 1.002 \\ 0.209 & -0.681 \\ 0.260 & 0.407 \\ 0.129 & 0.512 \end{bmatrix}$
Proposed alg. with white noise and L = 9	12% (at each output)	$\begin{bmatrix} 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ -0.124 & 0.751 & 0.001 & 0.007 & 0.017 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \\ 0.003 & 0.000 & 0.103 & -0.505 & 1.249 \end{bmatrix}$	$\begin{bmatrix} -0.498 & 0.300 \\ 0.648 & 1.001 \\ 0.204 & -0.689 \\ 0.247 & 0.404 \\ 0.120 & 0.502 \end{bmatrix}$

Table 4.3--continued

Algorithm	N.S.R.	Matrix A	Matrix B
Proposed alg.with white noise and L = 9	50% (at each output)	$\begin{bmatrix} 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ -0.141 & 0.778 & 0.007 & -0.021 & 0.075 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \\ 0.008 & 0.001 & 0.106 & -0.521 & 1.271 \end{bmatrix}$	$\begin{bmatrix} -0.457 & 0.287 \\ 0.629 & 0.993 \\ 0.202 & -0.667 \\ 0.237 & 0.406 \\ 0.131 & 0.503 \end{bmatrix}$
Proposed alg. with coloured noise and L = 6	Output 1=23.73% Output 2=25.72%	$\begin{bmatrix} 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ -0.118 & 0.744 & -0.009 & 0.001 & 0.021 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \\ 0.001 & 0.030 & 0.114 & -0.510 & 1.182 \end{bmatrix}$	$\begin{bmatrix} -0.503 & 0.304 \\ 0.690 & 1.008 \\ 0.198 & -0.693 \\ 0.262 & 0.412 \\ 0.123 & 0.509 \end{bmatrix}$

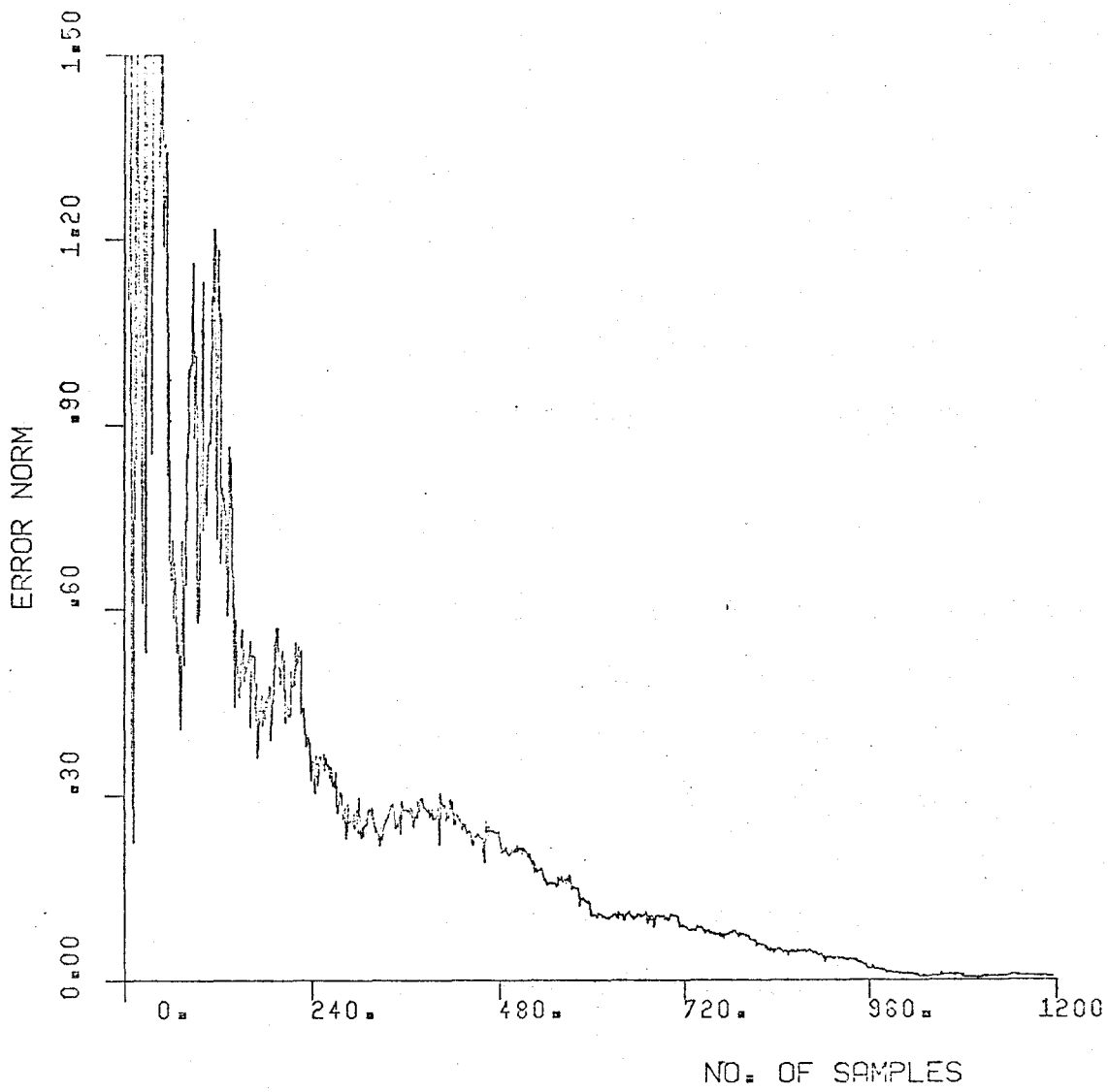

Fig. 4.6 Rate of convergence of the Proposed Algorithm for Example 2 with L = 6 and White Noise of N.S.R.(at each output) = 12 %

The above system is then identified, using the proposed algorithm, for two cases of noise-to-signal ratios of 12% and 50%. In the two cases the noise-to-signal ratio was the same at each of the two outputs. The rate of convergence of the identification algorithm in each case is shown in Fig. (4.7) and Fig. (4.8), also the estimates of system matrices after 1200 iterations are given in Table (4.3).

(3) Coloured Noise with L = 6:

In this case the identification interval L has been assigned the value L = 6. But, each of the two output noise sequences has been taken as a coloured noise sequence which is otbained as the output of a first-order digital filter with white-noise input i.e.,

$$v_1(k+1) = 0.8v_1(k) + .7\xi_1(k)$$
$$v_2(k+1) = 0.8v_2(k) + .7\xi_2(k)$$

where,

$\xi_1(k)$ and $\xi_2(k)$ are a zero-mean white noise sequences uncorrelated with the input sequences. The variances of sequences $\xi_1(k)$ and $\xi_2(k)$ are to be varied in order to change the noise-to-signal ratio at each output.

The above system is then identified using the proposed algorithm with noise-to-signal ratios of 23.73% and 25.72% at the first and second output respectively. The rate of convergence of the identification algorithm in this case is shown in Fig. (4.9) and the estimates of system matrices after 1200 iterations are given in

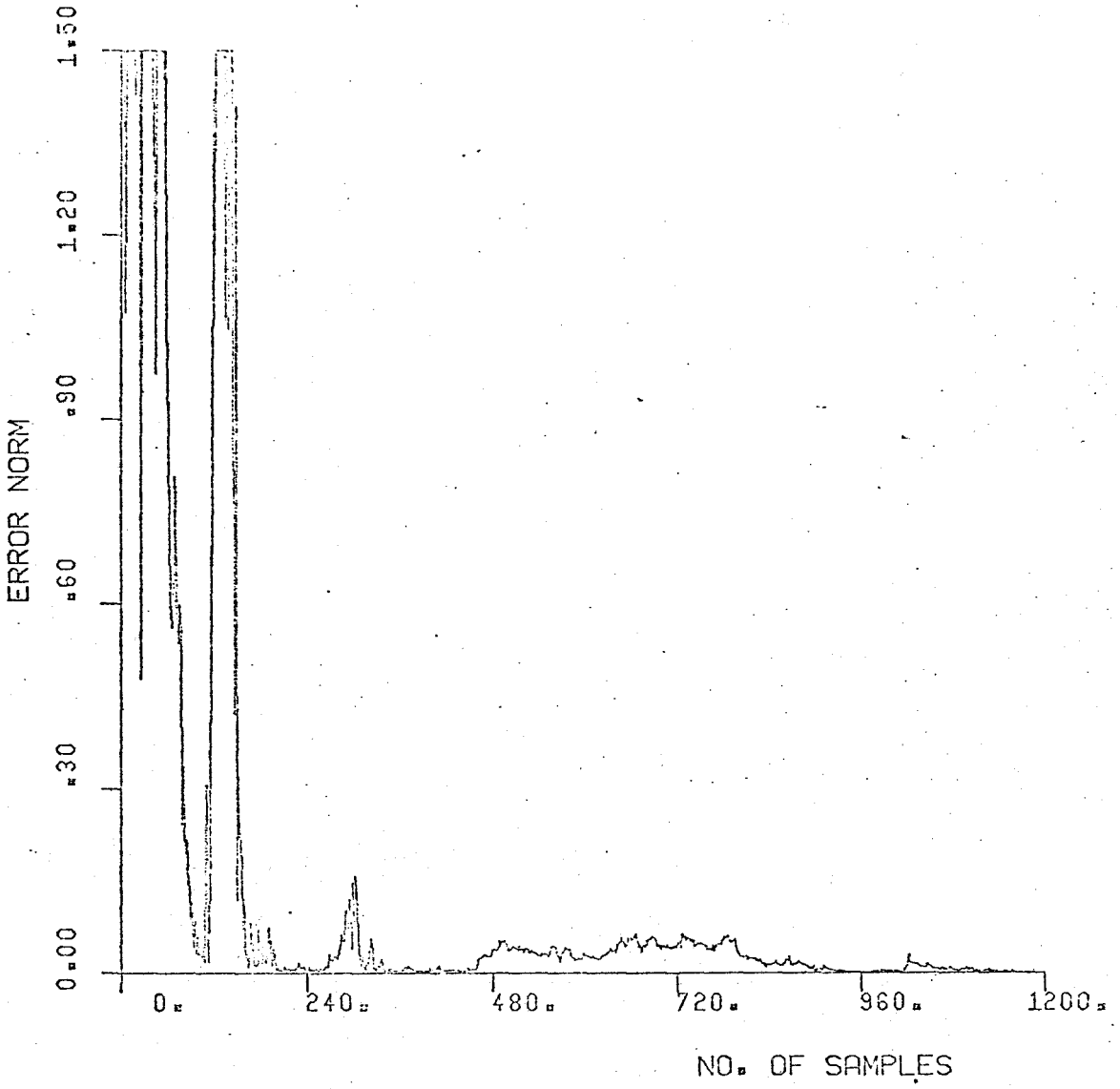

Fig. 4.7 Rate of convergence of the Proposed Algorithm for Example 2 with $L = 9$ and White Noise of N.S.R.(at each output) = 12 %

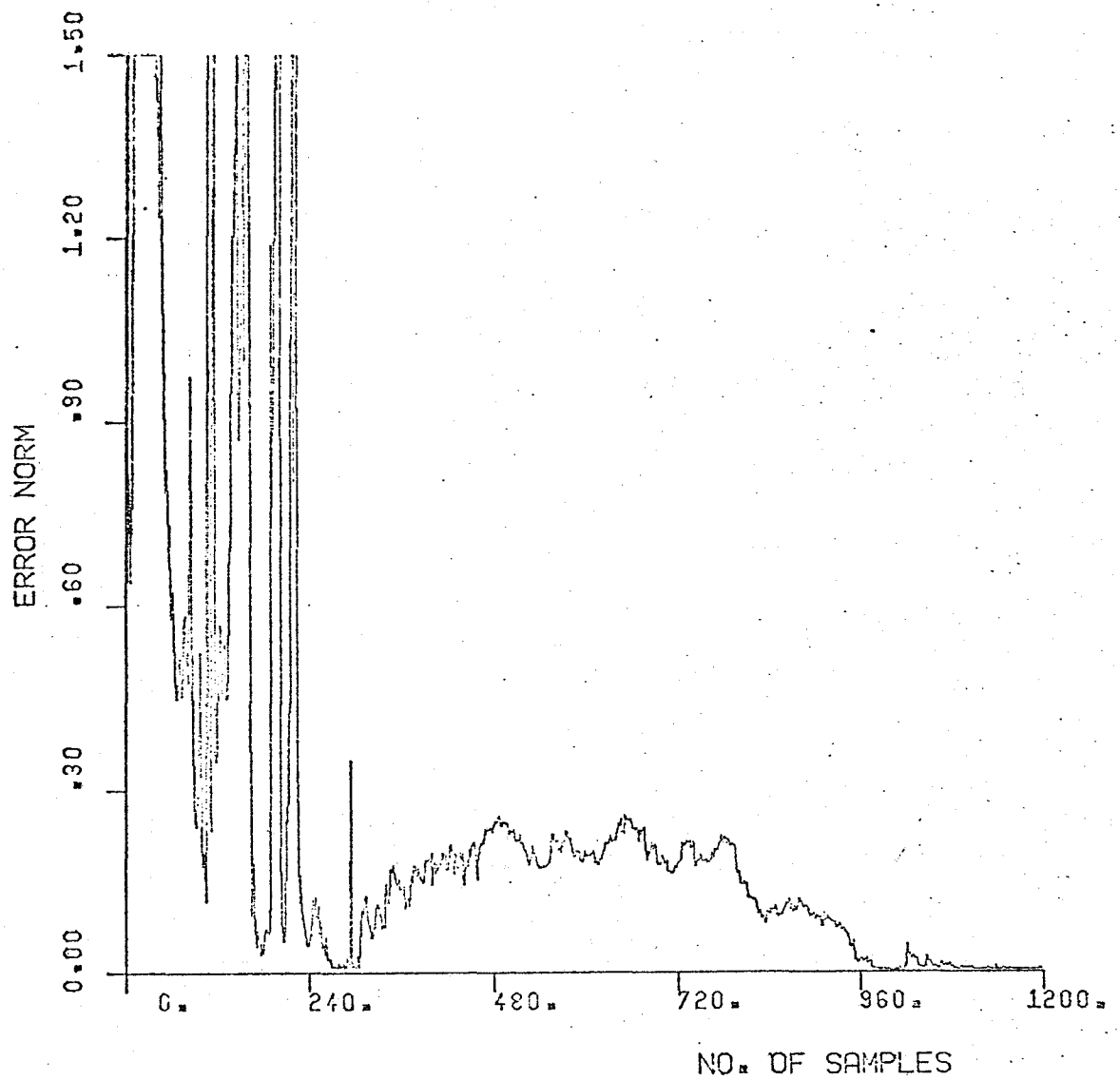

Fig. 4.8 Rate of Convergence of the Proposed Algorithm for Example 2 with L = 9 and White Noise of N.S.R.(at each output) = 50 %

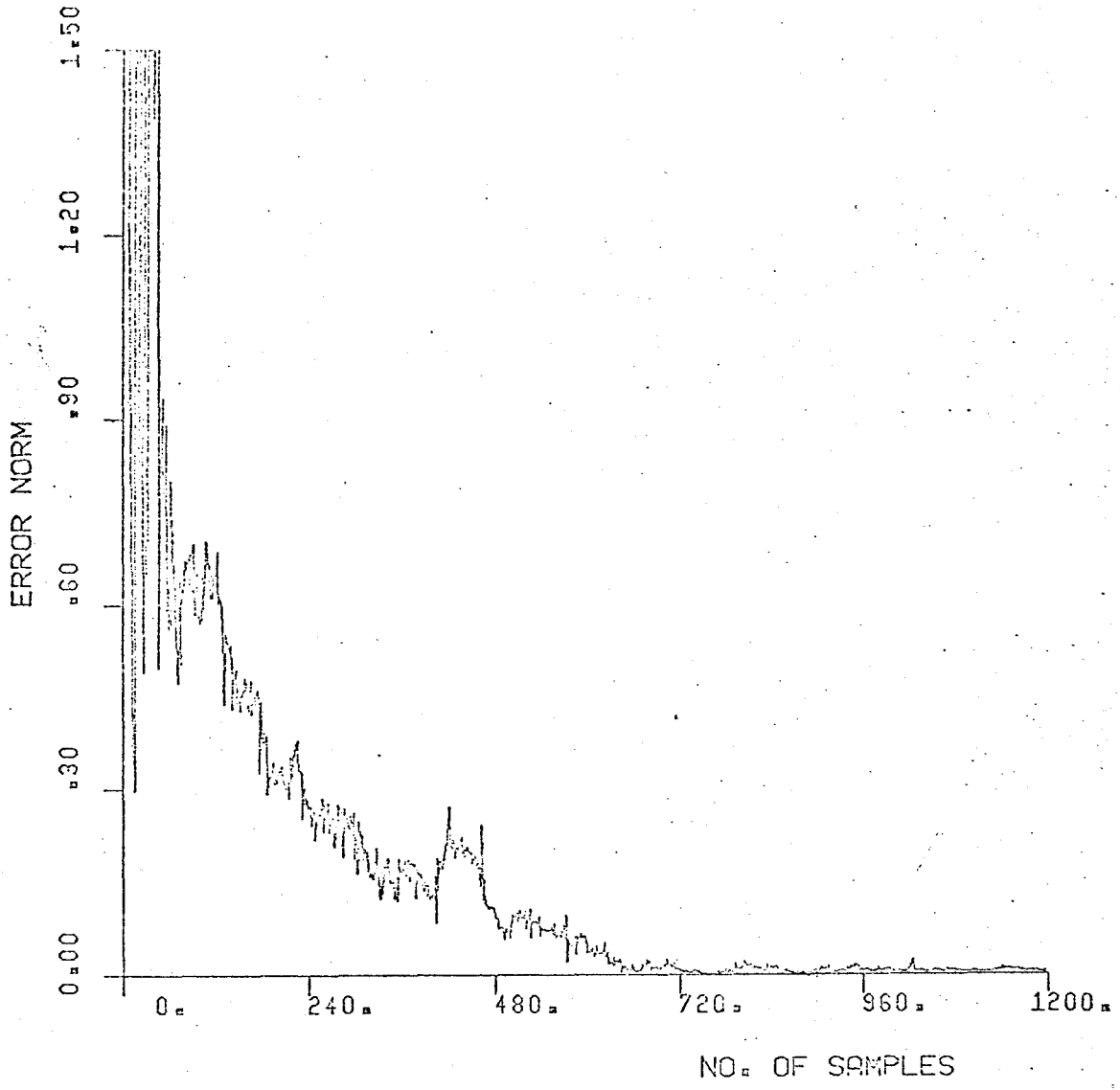

Fig. 4.9 Rate of Convergence of the Proposed Algorithm for Example 2 with L = 6 and Coloured Noise of N.S.R.(1) = 23.73 % N.S.R.(2) = 25.72 %

Table (4.3).

The number of arithmetic operations per iteration and the total computation time for 1200 iterations used in identifying the above system are given in Table (4.4) for the two cases of L = 6 and L = 9.

TABLE 4.4

THE COMPUTATION EFFORT FOR EXAMPLE 2

Algorithm	No. of additions and subtraction per iteration	No. of multiplications and divisions per iteration	Total computation time for 1200 iterations (sec.)
Proposed Alg. L = 6	27	49	1.75
Proposed Alg. L = 9	39	73	2.575

CHAPTER 5

CONCLUSIONS

In this thesis a non-parametric normalized stochastic approximation algorithm has been developed for the identification of multivariable systems from noisy measurements. This algorithm does not require the knowledge of the statistics of the noise contaminating the measurements.

The proposed identification algorithm processes live information at the rate it comes in and therefore it should be suitable for on-line applications.

Two systems have been selected for simulation studies in Chapter 4, which exhibit most of the salient properties of the identification process. The results of simulation indicate that the proposed algorithm has desirable convergence properties even for high noise-to-signal ratios, also they show the initial convergence of the identification algorithm which is due to the use of the normalized mean-square error criterion in the identification algorithm. As it is shown in Tables (4.2) and (4.4) Chapter 4, the number of arithmetic operations per iteration and the total computation time used in identifying the two examples using the proposed algorithm are very small.

In spite of the large number of samples used for identifying

the parameters of the system; the total computation time taken is quite small, as shown in Table (4.2) Chapter 4, compared to Sen and Sinha's algorithm. This is actually the main advantage of the stochastic approximation algorithms.

It can be seen that, the proposed algorithm is suitable for identifying systems which require fast identification in a very short time. On the other hand it is not suitable for the identification of systems with slow modes.

The proposed algorithm can be used only for the identification of open-loop systems. In the case where feedback is present, a difficulty arises because the feedback control signal produces additional correlations in the identification algorithm developed for open-loop systems. In such cases the proposed algorithm must be modified to identify closed-loop systems. This is a topic for further study.

REFERENCES

[1] H. Robbins and S. Monro, "A stochastic approximation method", Ann. Math. Statist., vol. 22, no. 1, pp.400-407, 1951.

[2] A. Dvoretzky, "On stochastic approximation", in Proc. 3rd Berkeley Symp. Math. Stat. and Probability, vol. 1, pp.39-55, 1956.

[3] D. J. Sakrison, "Stochastic Approximation, a Recursive Method for Solving Regression Problems", in Advances in Communication Theory, vol. 2, ch. 2, A.V. Balakishnan , Ed. New York: Academic Press, 1966.

[4] H.G. Kwatny and D.W.C. Shen, "Identification of nonlinear systems using a method of stochastic approximation", in Proc. 8th Joint Automatic Control Conf., 1967, pp. 814-826.

[5] D.J. Sakrison, "The use of stochastic approximation to solve the system identification problem", IEEE Trans. Automatic Control, vol. AC-12, pp.563-567, Oct. 1967.

[6] G.N. Saridis and G. Stein, "Stochastic approximation algorithms for discrete-time systems identification", IEEE Trans. Automatic Control, vol. AC-13, pp. 515-523, Oct. 1968.

[7] G.N. Saridis and G. Stein, "A new algorithm for system identification", IEEE Trans. Automatic Control (corresp.), vol. AC-13, pp. 592-594, Oct. 1968.

[8] V. Panuska, "A stochastic approximation method for identification of linear systems using adaptive filtering", Joint Automatic Control Conference, 1968, pp.1014-1021.

[9] Y. Bar-Shalom and S.C. Schwartz, " A non-parametric recursive identification algorithm for finite-memory linear systems", Proc. of 7th Annual Allerton Conference on Circuit and System Theory, 1969, pp.523-532.

[10] J.K. Holmes, "Two stochastic approximation procedures for identifying linear systems", IEEE Trans. Automatic Control (Short Papers), vol.14, pp.292-295, June 1969.

[11] G.N. Saridis and R.N. Lobbia, "Parameter identification and control of linear discrete-time systems", IEEE Trans. Automatic Control, vol. 17, pp.52-61, Feb. 1972.

[12] N.K. Sinha, "System identification using stochastic approximation", Reprints of 3rd IFAC Symp. on Identification and System Parameter Estimation, Hague, 1973, pp. 755-758.

[13] N.K. Sinha and M.P. Griscik, "A stochastic approximation method", IEEE Trans. on Systems, Man and Cybernetics ,vol. SMC-1, Oct. 1971, pp.338-344.

[14] H.G. Kwatny, "A note on stochastic approximation algorithms in system identification", IEEE Trans. Automatic Control, vol. 17, pp. 570-572, 1972.

[15] G.N. Saridis, "Stochastic approximation methods for identification and control- A Survey", IEEE Trans. Automatic Control, vol. 19,

pp.798-809, Dec. 1974.

[16] D. Graupe, Identification of Systems. Princeton, N.J.: Van Nostrand Reinhold, 1972.

[17] A.P.Sage & J.L.Melsa, System Identification. New York: Academic, 1971.

[18] J. Mendel, Discrete Techniques of Parameter Estimation. New York: Marchel Dekker, 1973.

[19] J.M. Mendel and K.S. Fu, Eds., Adaptive, Learning and Pattern Recognition Systems. New York: Academic, 1970.

[20] A. Sen and N.K. Sinha, "On line system identification using stochastic approximation", Proc. 8th Annual Princeton Conference on Identification Sciences and Systems. 1974.

[21] R.N. Lobbia and G.N. Saridis, "Identification and control of multivariable stochastic feedback systems", J. Cybern., pp.40-59. Jan./Mar. 1973.

[22] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function", Ann. Math. Statist., vol 23, no. 3, pp. 462-466, 1952.

[23] A. Albert and L. Gardner, Stochastic Approximation and Non-Linear Regression. Cambridge, Mass.: M.I.T. Press, 1967.

[24] H. Kesten, "Accelerated stochastic approximation methods", Ann. Math. Statist., vol. 29, no. 1, pp. 41-59, 1958.

[25] V. Dupac,"A dynamic stochastic approximation method", Ann. Math. Statist., vol. 36, pp. 1695-1702, 1965.

[26] D.G. Luenberger, "Canonical forms for linear multivariable systems", IEEE Trans. Automatic Control, vol. 12, no. 3, June 1967.

[27] N.K. Sinha and P. Rozsa, "Canonical forms for linear multivariable systems", Int. J. Control, vol. 23, no. 6, pp.865-883, 1976.

[28] W.A. Wolovich, Linear Multivariable Systems, Springer-Verlag, New York, 1974.

[29] C.T. Chen, Introduction to Linear System Theory, Holt Rinehart Winston, 1970.

[30] G. Hadley, Linear Algebra, Chapter 4, Reading Mass.: Addison-Wesley, 1961.

[31] A. Sen and N.K. Sinha, "On-line estimation of the parameters of a multivariable system using matrix pseudo-inverse", Int. J. of System Science, vol. 17, pp. 461-471, 1976.