A New Algorithm for Recursive Estimation of Controlled ARMA Processes

Mayne, David Q.; Åström, Karl Johan

1982

CODEN: LUTFD2/(TFRT-3162)/1-040/(1981)

A NEW ALGORITHM FOR RECURSIVE ESTIMATION OF PARAMETERS IN CONTROLLED ARMA PROCESSES

K J ÅSTRÖM

D Q MAYNE

DEPARTMENT OF AUTOMATIC CONTROL
LUND INSTITUTE OF TECHNOLOGY
APRIL 1981

LUND INSTITUTE OF TECHNOLOGY
DEPARTMENT OF AUTOMATIC CONTROL
Box 725
S 220 07 Lund 7 Sweden

Document name
Project report
Date of issue
April 1981
Document number
CODEN: LUTFD2/(TFRT-3162)/1-040/(1981)

Author(s)
K J Åström
D Q Mayne

Supervisor

Sponsoring organization

Title and subtitle

A new algorithm for recursive estimation of parameters in controlled ARMA processes

Abstract

A new recursive parameter estimation procedure for ARMA processes is proposed. The algorithm, which is based on convex approximations, is shown to be globally convergent.

Key words

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title

ISBN

Language
Enlgish

Number of pages
40

Security classification

Recipient's notes

A NEW ALGORITHM FOR

RECURSIVE ESTIMATION OF PARAMETERS

IN CONTROLLED ARMA PROCESSES

K.J. Åström and D.Q. Mayne

1. INTRODUCTION

Mathematical models of processes in terms of controlled ARMA processes are of interest in control engineering. Such models are convenient representations of systems, whose input-output relation can be characterized by rational transfer functions, subject to disturbances having rational spectral densities. The problem of estimating the parameters of controlled ARMA processes has also received much attention. The maximum likelihood method was applied in [1] where it was shown that the maximum likelihood estimates were consistent, asymptotically efficient and asymptotically normal. It is a drawback of the maximum likelihood method that the likelihood function is nonlinear. This implies that there may be several local minima and that the optimization may be difficult. Various alternative methods for estimating the parameters in controlled ARMA processes have therefore been proposed, e.g. the generalized least squares [2], the extended least squares [3] and the two stage least squares [4]. A new method was proposed in [5] and [6]. This method is a multistep technique where least squares is used in each step. A recursive version of the method presented in [5] is presented and analysed in this report.

The recursive algorithm is of interest for the design of adaptive regulators and adaptive predictors. A review of recursive estimation methods is given in [7]. It may be legitimally questioned whether it is of any use to add yet another method to a large number of already existing routines. Thus question can be answered as follows. In the case of pure ARMA processes (i.e. no inputs) it is known that the maximum likelihood method is globally convergent. Many of the other recursive methods are, however, not globally convergent even for ARMA processes. There is, moreover, no method which is known to converge globally when inputs are present. One motivation for introducing the method presented in this report is that it is globally convergent. Another motivation is that the corresponding off-line method is consistent and asymptotically efficient.

It is unfortunately only possible to analyse the asymptotic properties of the recursive estimation procedure. Since short sample properties are also important it follows that it is not possible to evaluate estimation methods purely by analysis. For this reason it is thus necessary to explore the proposed method by simulation and also to investigate its numerical properties before it can be judged soberly.

The report is organized as follows. The process model, consisting of a controlled ARMA process, is described in Section 2. The off-line estimation method, which is the basis for the recursive algorithm, is presented in Section 3. The properties of the off-line method are also discussed briefly in that section. The recursive algorithm is presented in Section 4, and its properties are analysed in Section 5. It is shown that the algorithm will under reasonable assumptions always converge.

2. PROCESS DESCRIPTION

The process is described in this section. The assumptions made on the process are also stated.

Only discrete time processes are discussed. It is thus assumed that time T is the set of integers $T = \{\ldots,-1,0,1,\ldots\}$. Signals are functions from T to R. They are denoted by lower case letters like u and y. It is assumed that the process is a controlled ARMA process described by:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t) \tag{2.1}$$

where u is the input signal, y the output signal and $\{e(t)\}$ is a sequence of independent identically distributed random variables with zero mean values and covariances λ^2. In (2.1) $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ are polynomials of degree n in the backward shift operator q^{-1}:

$$A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_nq^{-n}$$

$$B(q^{-1}) = b_1q^{-1} + \dots + b_nq^{-n}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + \dots + c_nq^{-n} \tag{2.2}$$

It is assumed that there are no factors common to A, B and C. There is no loss of generality is assuming that $A(0) = C(0) = 1$. The assumption that $B(0) = 0$ is not important; it is made only to obtain symmetry in certain equations.

The model (2.1) will also be written in the following abbreviated form:

$$Ay = Bu + Ce \tag{2.1'}$$

The problem we will consider is estimation of the parameters of the process (2.1) from observations of input-output records $\{u(t), y(t), t \in T\}$ of the process. For easy reference the parameters are gathered into the vector:

$$\Theta = \text{coeff}\,[A(q^{-1}), B(q^{-1}), C(q^{-1})]$$

$$\triangleq [a_1\ a_2\ \ldots\ a_n\ b_1\ b_2\ \ldots\ b_n\ c_1\ c_2\ \ldots\ c_n] \tag{2.3}$$

Conversely to each parameter vector Θ we can also associate polynomials A, B and C given by:

$$A(q^{-1}) = 1 + \Theta_1 q^{-1} + \ldots + \Theta_n q^{-n}$$

$$B(q^{-1}) = \Theta_{n+1} q^{-1} + \ldots + \Theta_{2n} q^{-n} \tag{2.4}$$

$$C(q^{-1}) = 1 + \Theta_{2n+1} q^{-1} + \ldots + \Theta_{3n} q^{-n} \quad .$$

Assumptions

The following assumptions are made on the process (2.1) which generates the data to be used in the system identification:

A1: There are no factors common to all of A, B and C and $C(q^{-1}) \neq 1$. Furthermore the zeros of the polynomials $z^n A(z^{-1})$ and $z^n C(z^{-1})$ are strictly less than $\gamma < 1$ in magnitude.

A2: The input signal is covariance stationary and persistently exciting of arbitrary order p.

A3: The disturbance e is a sequence of independent random variables with bounded fourth moments. The disturbance is uncorrelated.

3. THE OFF-LINE ALGORITHM

A new algorithm for estimating the parameters of the process (2.1) was proposed in [5] and [6]. The algorithm is a multistep method where least squares is used in each step. The algorithm is described in this section. Some of its properties are also discussed together with minor modifications which lead to different versions of the algorithm.

The Algorithm

Step 1: Estimate A_1 and B_1 in the model

$$A_1 y = B_1 u \tag{M_1}$$

where $\deg A_1 = \deg B_1 = p$ by least squares. Let ε_1 denote the residuals obtained.

Step 2: Estimate A_2, B_2 and C_2 in the model

$$A_2 y = B_2 u + C_2 \varepsilon_1 \tag{M_2}$$

where the polynomials have the form (2.2) and ε_1 is the residual from Step 1, by least squares.

Step 3: Filter the signals u, y and ε_1 by C_2^{-1} where C_2 is the polynomial estimated in Step 2 to obtain

$$\bar{y} = \frac{1}{C_2} y, \quad \bar{u} = \frac{1}{C_2} u, \quad \bar{\varepsilon}_1 = \frac{1}{C_2} \varepsilon_1$$

Then estimate A_3, B_3 and C_3 in the model

$$A_3\bar{y} = B_3\bar{u} + C_3\bar{\varepsilon}_1 \qquad (M_3)$$

where the polynomials A_3, B_3 and C_3 are of the form (2.2), by least squares.

Minor Modifications

In Step 1 the parameters are determined in such a way that the criterion

$$\sum_t [A_1(q^{-1})y(t)-B_1(q^{-1})u(t)]^2$$

is minimized. Similarly the criterion

$$\sum_t [A_2(q^{-1})y(t)-B_2(q^{-1})u(t)-C_2(q^{-1})\varepsilon_1(t)]^2$$

is used in Step 2. Since the leading coefficient in $C_2(q^{-1})$ is one (c.f. (2.2)) the alternative criterion

$$\sum_t \{A_2(q^{-1})y(t)-B_2(q^{-1})u(t)-[C_2(q^{-1})-1]\varepsilon_1(t)\}^2$$

can also be used in Step 2. This means that the model M_2 is replaced by

$$A_2y = B_2u + (C_2-1)\varepsilon_1 \qquad (M_2')$$

Similarly the model M_3 in Step 3 can be replaced by

$$A_3\bar{y} = B_3\bar{u} + (C_3-C_2)\bar{\varepsilon}_1 \qquad (M_3')$$

It is easily seen that these modifications are asymptotically (for large p and large N) unimportant. They will, however, give versions of the algorithm that are slightly different for finite p and N.

Properties of the Off-Line Algorithm

The residuals ε_1 obtained in Step 1 will be close to the process innovations e if p and N are large. In [5] and [6] it is shown that if all zeros of the polynomial $z^n C(z^{-1})$ have magnitudes less than $\gamma < 1$ then asymptotically for large N

$$E|\varepsilon_1(t)-e(t)|^2 \leq K\,\gamma^{2(p-n)}$$

For large N and p the polynomials A_1 and B_1 obtained in Step 1 are also close to the polynomials obtained by truncating the series expansions in q^{-1} of the rational functions $A(q^{-1})/C(q^{-1})$ and $B(q^{-1})/C(q^{-1})$ in a sense given precisely in [6]. This result is further illustrated by the following simple example:

Example 3.1

Consider the process

$$y(t) = e(t) + ce(t-1), \quad |c| < 1$$

A straightforward solution of the normal equations gives asymptotically, for large N, the following coefficients of the polynomial A_1

$$a_i = (-c)^i \frac{1-c^{2(p-i+1)}}{1-c^{2(p+1)}}, \quad i = 1, \ldots, p$$

The truncation of the rational function $1/C(q^{-1})$ gives a polynomial with the coefficients $(-c)^i$, $i = 1, \ldots, p$. $\square$

An estimate of the parameters of the process (2.1) is already obtained in Step 2. This estimate is the well known two-stage least squares (2SLS) estimate which was originally proposed by Durbin [4], and later in the automatic control literature [8] and [9]. (Note that the term 2LSL is used in the econometrics literature [10] to describe a different estimate, closely related to instrumental variable estimators.) It is shown in [6] and [9] that the bias of the 2SLS estimate can be made arbitrarily small by choosing the parameter p in the first step sufficiently large. The following example from [9] gives the bias of the 2SLS estimate of a first order system:

Example 3.2

Consider the process

$$y(t) + ay(t-1) = bu(t-1) + e(t) + ce(t-1)$$

where the input u is white noise which is independent of e and has variance σ^2. Then [9] as $N \to \infty$, the 2SLS estimate converges to

$$\hat{a} = a - \frac{\delta c^{2p-1}(1-c^2)}{(1-c^{2p})(1+\delta-c^{2p})}$$

$$\hat{b} = b$$

$$\hat{c} = c - \frac{c^{2p-1}(1-c^2)}{(1-c^{2p})}$$

$$\delta = \frac{\lambda^2(1-a^2)(1-c^2)}{\lambda^2(c-a)^2+b^2\sigma^2}$$

If the modified estimate obtained by fitting the model M_2' instead of M_2 then [9] the 2SLS estimate converges instead to

$$\hat{a} = a$$

$$\hat{b} = b$$

$$\hat{c} = c - \frac{c^{2p+1}(1-c^2)}{1+\delta-c^{2p+2}}$$

□

The example thus indicates that it is slightly advantageous (at least for n=1) to use the version of the algorithm where the model M_2' is fitted in the second step.

The 2SLS estimate is known to be inefficient. The purpose of the third step in the algorithm is to reduce the variance of 2SLS estimate.

In the second stage of the algorithm the loss function

$$V_2(\theta_2) = \frac{1}{2N}\sum_{t=1}^{N} [A_2y(t) - B_2u(t) - C_2\varepsilon_1(t)]^2$$

is minimised with respect to θ_2, i.e. with respect to (the coefficients of) A_2, B_2, C_2. In the third stage the loss function

$$V_3(\theta_3) = \frac{1}{2N}\sum_{t=1}^{N} [\frac{A_3}{C_2}\ y(t) - \frac{B_3}{C_2}\ u(t) - \frac{C_3}{C_2}\ \varepsilon_1(t)]^2$$

is minimised with respect to θ_3 (i.e. A_3, B_3, C_3). The likelihood function associated with the problem is:

$$V_{ML}\ (\theta) = \frac{1}{2N}\sum_{t=1}^{N} [\frac{A}{C}\ y(t) - \frac{B}{C}\ u(t)]^2$$

The maximum likelihood estimate θ_{ML} minimises $V_{ML}(\theta)$ with respect to θ. It is shown in [14] that $\theta_3 \to \theta^p$ w.p. 1, as $N \to \infty$, that $\theta^p \to \theta$ as $p \to \infty$, that the asymptotic variance (as $N \to \infty$) of $\sqrt{N}\,[\theta_3 - \theta^p]$ is:

$$[E\nabla^2 V_3(\theta^p)]^{-1} \quad [EN\nabla V_3(\theta^p) \;\nabla V_3^T(\theta^p)]^{-1} \; [E\nabla^2 V_3(\theta^p)]^{-1}$$

and that $E\nabla^2 V_3(\theta^p) \to E\nabla^2 V_{ML}(\theta)$ and $EN\nabla V_3(\theta^p)\; \nabla V_3^T\;(\theta^p) \to E\nabla^2 V_{ML}(\theta)$ as $p \to \infty$. Hence the asymptotic variance of $\sqrt{N}\,[\theta_3 - \theta^p]$ tends to that of $\sqrt{N}\,[\theta_{ML} - \theta]$ as $p \to \infty$; this property is not possessed by θ_2. Since V_3 is a convex (indeed quadratic) function, the estimation scheme can be regarded as based on a convex approximation of the likelihood function.

4. THE RECURSIVE ALGORITHM

Since the algorithm given in the previous section is composed of three least squares steps it is easy to obtain a recursive algorithm simply be replacing the least square estimations by recursive least squares in each step. The recursive algorithm will then be built up of three recursive least squares steps. The equations will be very similar for all steps. The following notation is introduced to describe the algorithm.

$$\theta_1 = \text{coeff}\;(A_1, B_1), \quad \dim\theta_1 = 2p$$

$$\theta_2 = \text{coeff}\;(A_2, B_2, C_2), \quad \dim\theta_2 = 3n$$

$$\theta_3 = \text{coeff}\;(A_3, B_3, C_3), \quad \dim\theta_3 = 3n \tag{4.1}$$

The regressors are denoted as

$$z_1(t) = [-y(t-1)\ldots -y(t-p) \;\; u(t-1)\ldots u(t-p)]^T$$

$$z_2(t) = [-y(t-1)\ldots -y(t-n) \;\; u(t-1)\ldots u(t-n) \;\; \varepsilon_1(t-1)\ldots \varepsilon_1(t-n)]^T$$

$$z_3(t) = [-\bar{y}(t-1)\ldots-\bar{y}(t-n)\ \bar{u}(t-1)\ldots\bar{u}(t-n)\ \bar{\varepsilon}_1(t-1)\ldots\bar{\varepsilon}_1(t-n)]^T \tag{4.2}$$

where $\bar{y}$, $\bar{u}$ and $\bar{\varepsilon}_1$ are the filtered variables. They will be defined precisely below.

It is straightforward to obtain recursive equations for the first step simply by replacing the least squares calculation by recursive least squares. Such a recursive algorithm will give exactly the same estimates as the off-line algorithm provided that the recursion is initialized properly. The same trick can, however, not be used in Steps 2 and 3. To calculate the regressor $z_2(t)$ it is necessary to know the past residuals $\varepsilon_1(t-n)$. These, however, depend on the estimate θ_1 which, in turn, depends on all data. This estimate is clearly not available at time t. Similarly evaluation of the regressor $z_3(t)$ requires knowledge of the filter polynomial $\hat{C}_2$, which is an estimate based on all data. This is not available at time t. To obtain a recursive algorithm it is therefore necessary to approximate. One possibility is to approximate the residuals by

$$\varepsilon_1(t) = y(t) - z_1^T(t)\hat{\theta}_1(t-1)$$

where $\hat{\theta}_1(t)$ is the estimate of θ_1 in (M_1) based on data available at time t. Similarly the filtered values $\bar{y}$, $\bar{u}$ and $\bar{\varepsilon}_1$ in $z_3(t)$ are approximated by

$$\bar{y}(t) = y(t) - \hat{c}_1(t)\bar{y}(t-1) - \ldots - \hat{c}_n(t)\bar{y}(t-n)$$

$$\bar{u}(t) = u(t) - \hat{c}_1(t)\bar{u}(t-1) - \ldots - \hat{c}_n(t)\bar{u}(t-n)$$

$$\bar{\varepsilon}_1(t) = \varepsilon_1(t) - \hat{c}_1(t)\bar{\varepsilon}_1(t-1) - \ldots - \hat{c}_n(t)\bar{\varepsilon}_1(t-n) \tag{4.3}$$

where $\hat{c}_2(q^{-1}) \triangleq 1 + \hat{c}_1(q^{-1}) + \ldots + \hat{c}_n(q^{-n})$, that is $\hat{c}_i(t)$ is the estimate of the parameter c_i based on data available up to time t. The equations (4.3) are difference equations. Since $\hat{c}_i$, $i = 1, \ldots, p$, is an estimate it may happen that the difference equations are unstable. This difficulty can be overcome by testing for stability and reflecting in the unit circle when necessary. Summarizing the recursive estimation algorithm can thus be written as follows:

Algorithm

Step 1: The recursive equations are given by

$$\hat{\theta}_1(t+1) = \hat{\theta}_1(t) + P_1(t+1)z_1(t+1)\varepsilon_1(t+1)$$

$$\varepsilon_1(t+1) = y(t+1) - z_1^T(t+1)\hat{\theta}_1(t)$$

$$P_1^{-1}(t+1) = P_1^{-1}(t) + z_1(t+1)z_1^T(t+1) \tag{4.4a}$$

where the regressors are defined by (4.1). The number p of components of θ_1 is larger than 2n.

Step 2: The recursive equations are given by

$$\hat{\theta}_2(t+1) = \hat{\theta}_2(t) + P_2(t+1)z_2(t+1)\varepsilon_2(t+1)$$

$$\varepsilon_2(t+1) = y(t+1) - \varepsilon_1(t+1) - z_2^T(t+1)\hat{\theta}_2(t)$$

$$P_2^{-1}(t+1) = P_2^{-1}(t) + z_2(t)z_2^T(t) \tag{4.4b}$$

where the regressor $z_2(t)$ is defined by (4.2) and $\varepsilon_1(t)$ by (4.4a).

Step 3: The recursive equations are given by

$$\hat{\theta}_3(t+1) = \hat{\theta}_3(t) + P_3(t+1)z_3(t+1)\varepsilon_3(t+1)$$

$$\varepsilon_3(t+1) = \bar{y}(t+1) - \bar{\varepsilon}_1(t+1) - z_3^T(t+1)\hat{\theta}_3(t)$$

$$P_3^{-1}(t+1) = P_3^{-1}(t) + z_3(t)z_3^T(t) \tag{4.4c}$$

where the regressor $z_3(t)$ is defined by (4.2) and the filtering by (4.3) where $\hat{C}_2$ is obtained from (4.4b) i.e. $\hat{c}_i$, the i^{th} component of $\hat{C}_2$ is given by:

$$\begin{aligned}\hat{c}_i(t) &= \hat{\theta}_2^{2n+i} \\ &= (2n+i)\text{:th component of } \hat{\theta}_2(t)\end{aligned} \tag{4.5}$$

Remark 1

It is well known that the matrices $P_i(t)$ satisfy the following recursive equations

$$P_i(t+1) = P_i(t) - P_i(t)z_i(t+1)[1+z_i^T(t+1)P_i(t)z_i(t+1)]^{-1}z_i^T(t+1)P_i(t) \tag{4.6}$$

If the matrices P_i are badly conditioned square root algorithms [10] could also be used. It is, however, more convenient for the analysis to use the equations given for $P_i^{-1}(t)$. Since the number of parameters estimated in the first step may be quite large it is useful to use a fast algorithm, [11], [12] for solving the least squares equations at least in the first step. Neglecting numerical errors the fast algorithms will give the same estimates as the ordinary algorithm. The analysis can therefore be based on the normal algorithms.

Remark 2

The algorithm given can be modified slightly by redefining the residuals in the second and third steps as

$$\varepsilon_2(t+1) = y(t+1) - z_2^T(t+1)\hat{\theta}_2(t)$$

and

$$\varepsilon_3(t+1) = \bar{y}(t+1) - z_3^T(t+1)\hat{\theta}_3(t) - \hat{C}_2(q^{-1})\varepsilon_2(t+1)$$

This corresponds to fitting the models (M_2') and (M_3') in the off-line case.

Remark 3

The third step can be repeated many times as for the off-line algorithm.

5. CONVERGENCE ANALYSIS

The recursive algorithm proposed in Chapter 4 will now be analysed. It has been shown by Ljung [13] that the convergence of certain recursive algorithms are closely related to the stability of an ordinary differential equation. In this chapter the ordinary differential equation associated with the recursive algorithm of Chapter 4 is first derived. It is then shown that this differential equation has a unique stationary solution which is globally asymptotically stable. The convergence of the recursive parameter estimation algorithm then follows from Ljung's theorem.

The Associated Ordinary Differential Equations

Ljung [13] considers algorithms of the form

$$x(t) = x(t-1) + \gamma(t)Q(x(t-1),\phi(t)) \tag{5.1}$$

where $\phi(t)$ may depend on past $x(t)$ through

$$\phi(t) = A(x(t-1))\phi(t-1)+B(x(t-1))e(t)$$

Furthermore $\gamma(t)$ are numbers that converge to zero, e.g. as $1/t$, as $t \to \infty$, and $\{e(t)\}$ is a sequence of independent random variables. It is shown in [13] that the "estimates" x generated by (5.1) are close to the solutions of the ordinary differential equation

$$\frac{dx}{d\tau} = f(x) \tag{5.2}$$

where

$$f(x) \triangleq EQ(x,\mathring{\phi}(x,t))$$

and

$$\mathring{\phi}(x,t) = A(x)\mathring{\phi}(x,t-1) + B(x)e(t)$$

A precise statement of the results and the conditions required are given in [13].

The recursive parameter estimation algorithm described by equations (4.4) can be written in the form (5.1). The components of the r vector x are the components of the vectors $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ and the elements of the matrices P_1^{-1}, P_2^{-1} and P_3^{-1}. To write equations (4.4) in the form (5.1) introduce $R_i(t) = tP_i(t)$, i = 1, 2, 3. The equations (4.4) can then be written as

$$\hat{\theta}_i(t+1) = \hat{\theta}_i(t) + \frac{1}{t+1} R_i(t+1)z_i(t+1)\varepsilon_i(t+1)$$

$$R_i^{-1}(t+1) = R_i^{-1}(t) + \frac{1}{t+1}\left[z_i(t+1)z_i^T(t+1) - R_i^{-1}(t)\right]$$

for i = 1, 2, and 3. These equations are clearly of the form (5.1). To obtain the ordinary differential equation (5.2) introduce the following functions.

$$S_1 = Ez_1(t+1)z_1^T(t+1)$$

$$S_2(\Theta_1) = Ez_2(t+1)z_2^T(t+1)$$

$$S_3(\Theta_1,\Theta_2) = Ez_3(t+1)z_3^T(t+1)$$

$$s_1 = Ez_1(t+1)y(t+1)$$

$$s_2(\Theta_1) = Ez_2(t+1)[y(t+1)-\varepsilon_1(t+1)]$$

$$s_3(\Theta_1,\Theta_2) = Ez_3(t+1)[\bar{y}(t+1)-\bar{\varepsilon}_1(t+1)] \tag{5.3}$$

The expectation in (5.3) should be calculated under the assumption that the data is generated by the process (2.1) in stationary equilibrium. The ordinary differential equations which are associated with the difference equations (4.4), (4.5) and (4.6) can then be written as

$$\begin{cases} \dfrac{d\hat{\Theta}_1}{d\tau} = R_1[s_1 - S_1\hat{\Theta}_1] \\ \\ \dfrac{dR_1^{-1}}{d\tau} = S_1 - R_1^{-1} \end{cases} \tag{5.4a}$$

$$\begin{cases} \dfrac{d\hat{\theta}_2}{d\tau} = R_2[s_2(\hat{\theta}_1) - S_2(\hat{\theta}_1)\hat{\theta}_2] \\ \\ \dfrac{dR_2^{-1}}{dt} = S_2(\hat{\theta}_1) - R_2^{-1} \end{cases} \tag{5.4b}$$

$$\begin{cases} \dfrac{d\hat{\theta}_3}{d\tau} = R_3[s_3(\hat{\theta}_1,\hat{\theta}_2) - S_3(\hat{\theta}_1,\hat{\theta}_2)\hat{\theta}_3] \\ \\ \dfrac{dR_3^{-1}}{d\tau} = S_3(\hat{\theta}_1,\hat{\theta}_2) - R_3^{-1} \end{cases} \tag{5.4c}$$

These ordinary differential equations, which are associated with the three steps of the estimation procedure, will now be analysed. Notice that the equations have a "triangular" structure. The equations (5.4a) can thus be integrated independently of the other equations to give $\hat{\theta}_1$ and R_1 as functions of τ. Knowing $\hat{\theta}_1$ the equations (5.4b) can then be integrated to give $\hat{\theta}_2$ and R_2. Finally when $\hat{\theta}_1$ and $\hat{\theta}_2$ are known the equation (5.4c) can be integrated to give $\hat{\theta}_3$ and R_3. The special structure of the equations simplifies the analysis substantially.

Stationary Solutions

The equilibrium solutions to the differential equations will now be investigated. The main result is:

Theorem 1

Assume that $p > 2n$ then the differential equations (5.4) have a unique stationary solution given by

$$\begin{cases} \theta_1^0 = S_1^{-1} s_1 \\ \\ R_1^0 = S_1^{-1} \end{cases} \tag{5.5a}$$

$$\begin{cases} \theta_2^0 = S_2(\theta_1^0)^{-1} s_2(\theta_1^0) \\ \\ R_2^0 = S_2(\theta_1^0)^{-1} \end{cases} \tag{5.5b}$$

$$\begin{cases} \theta_3^0 = S_3(\theta_1^0,\theta_2^0)^{-1} s_3(\theta_1^0,\theta_2^0) \\ \\ R_3 = S_3(\theta_1^0,\theta_2^0)^{-1} \end{cases} \tag{5.5c}$$

where the matrices S_1, S_2 and S_3 are positive definite.

Proof

The "triangular" structure of the equations will be used by analysing the equations one at a time. First consider the equations (5.4a). It is a well-known result in least squares theory that the matrix S_1 is positive definite; this follows from u and e being uncorrelated and from u being persistently exciting. A full proof is given as Lemma B1 in Appendix B. It then follows that (5.5a) is the stationary solution of (5.4a). It is also shown in Lemma B2 that the matrices $S_2(\theta_1^0)$ and $S_3(\theta_1^0,\theta_2)$ are positive definite and it thus follows that the stationary solutions to (5.4b) and (5.4c) are given by (5.5b) and (5.5c). □

Remark

It is shown in Appendix B that the matrices S_1 and $S_2(\theta_1^0)$ are bounded. For small p it may happen that the estimate θ_2^0 is such that its associated polynomial C_2 is unstable S_3 would then become unbounded. It follows from Example 3.2 that this will not happen in the first order case because there

$$\hat{c} = c[1 - \frac{c^{2p}(1-c^2)}{1-c^{2p}}]$$

and clearly

$$|\hat{c}| < |c|$$

For large p the estimate C_2 will be close to C which is stable.

Global Stability

Conditions have thus been found which imply that the differential equations (5.4) have a unique stationary solution. It is then natural to investigate the stability of the stationary solution. The following result can then be established.

Theorem 2

Assume that the initial conditions to the differential equations (5.4) are such that $R_1(0)$, $R_2(0)$ and $R_3(0)$ are positive definite, then the solutions will converge exponentially to the stationary solutions given by Theorem 1.

Proof

The continuity of the functions s_1, s_2, s_3, S_1, S_2 and S_3 appearing in the right hand side of the equations is established in Appendix C. This in essence establishes the existence of the differential equation. The global stability is proven by exploiting the "triangular" structure of the equations. The proof is thus composed of three almost identical parts each dealing with the equations (5.4a), (5.4b) and (5.4c).

Part 1

First consider the equation (5.4a). The equation

$$\frac{dR_1^{-1}}{d\tau} + R_1^{-1} = S_1$$

has the solution

$$R_1(\tau)^{-1} = e^{-\tau}R_1(0)^{-1} + \int_0^{\tau} e^{-(\tau-s)} S_1 ds$$

Since S_1 is a constant matrix we get

$$R_1(\tau)^{-1} = e^{-\tau}R_1(0)^{-1} + (1-e^{-\tau})S_1 \tag{5.6}$$

It was assumed that $R_1(0)$ was positive definite and it was shown in Lemma B1 that S_1 is positive definite. It thus follows that $R_1(\tau)^{-1}$ is positive definite for all $\tau \in [0,\infty]$. The matrix $R_1(\tau)$ thus exists and is bounded for all $\tau \in [0,\infty]$. Furthermore it follows from (5.6) that $R_1(\tau)^{-1}$ converges exponentially to S_1. Then $R_1(\tau)S_1$ is also bounded and

$$R_1(\tau)S_1 \to I$$

exponentially with rate $\exp(-\tau)$ as $\tau \to \infty$.

The equation

$$\frac{d\hat{\theta}_1}{d\tau} = R_1(s_1 - S_1\hat{\theta}_1) = -R_1S_1\theta_1 + R_1s_1$$

is a linear timevarying differential equation. The equation is exponentially stable and the solution $\hat{\theta}_1(t)$ thus converges to the equilibrium solution θ_1^0 with rate $\exp(-(1-\varepsilon)\tau)$.

Part 2

Now consider equation (5.4b). The solution of the equation

$$\frac{dR_2^{-1}}{d\tau} = S_2(\hat{\theta}_1(\tau)) - R_2^{-1}$$

can be represented as

$$R_2(\tau)^{-1} = e^{-\tau}R_2(0)^{-1} + \int_0^{\tau} e^{-(\tau-s)} S_2(\hat{\theta}_1(s))ds \tag{5.7}$$

It was assumed that $R_2(0)$ was positive definite. The matrix $S_2(\hat{\theta}_1(s))$ is nonnegative definite for all s. It follows from Lemma B2 that $S_2(\theta_1^0)$ is positive definite and from Lemma C1 that S_2 is continuous at θ_1^0. Since it has been shown in Part 1 of this proof that $\hat{\theta}_1(s)$ converges to θ_1^0 it thus follows that $S_2(\hat{\theta}_1(s))$ is positive definite for large s. The matrix $R_2(\tau)^{-1}$ is thus uniformly positive definite, in $[0,\infty]$ and its inverse $R_2(\tau)$ thus exists and is bounded in $[0,\infty]$. Moreover it follows from (5.7) that $R_2(\tau)^{-1} \to S_2(\theta_1^0)$, exponentially with rate $\exp(-(1-\varepsilon)\tau)$ as $\tau \to \infty$ and consequently that

$$R_2(\tau)S_2 \to I$$

The same argument as in Part 1 of the proof now establishes that $\hat{\theta}_2(\tau)$ converges exponentially to θ_2^0 with rate $\exp(-(1-\varepsilon)\tau)$ as $\tau \to \infty$.

Part 3

Consider the equation (5.4c). The solution of the equation

$$\frac{dR_3^{-1}}{d\tau} = S_3(\hat{\theta}_1(\tau),\hat{\theta}_2(\tau)) - R_3^{-1}$$

can be represented as

$$R_3(\tau)^{-1} = e^{-\tau}R_3(0)^{-1} + \int_0^{\tau} e^{-(\tau-s)} S_3(\hat{\theta}_1(s),\hat{\theta}_2(s))ds$$

Using arguments analogous to those used in Part 2 of the proof it follows that $R_3(\tau)^{-1}$ is positive definite uniformly in $[0,\infty]$, that $R_3(\tau)$ exists and that

$$R_3(\tau)S_3(\hat{\Theta}_1(\tau),\hat{\Theta}_2(\tau)) \to I$$

exponentially with rate $\exp(-(1-\varepsilon)\tau)$. As before it then follows that $\hat{\Theta}_3(\tau)$ also converges exponentially to Θ_3^0. □

Remark

In view of the properties of the corresponding off-line algorithm it is natural to use a large value of p to make the limiting estimates close to the true parameters of the process model. The assumption for the initial values of R_1, R_2 and R_3 to be positive definite is also very natural.

In practice this will not be any real restriction.

Convergence of the Parameter Estimates

Having established the conditions for the ordinary differential equation (5.4) to converge to its unique stationary solution we will now return to the recursive parameter estimation procedure. The following result then holds.

Theorem 3

Consider the recursive parameter estimation algorithm with initial conditions such that P_1, P_2 and P_3 are positive definite. Assume that the algorithm is applied to data generated by the process (2.1) subject to assumptions A1, A2 and A3. Then the parameter estimates converge with probability one to the values given by equation (5.5).

Proof

Since the input signal is mean square bounded and the system is stable the output is also mean square bounded. Under the assumptions of the theorem the differential equation

Under the assumptions of the theorem the differential equation (5.4) will converge (exponentially) to the unique stationary solution (5.5). Assumptions A1, A2 and A3 imply that the conditions for Theorem 1 of Ljung [13] are satisfied. The probability one convergence then follows from Ljung's theorem. □

Remark

Notice that the limiting values (5.5) are the same as the limiting values of the off-line algorithm.

6. Simulations

A range of simulation studies were carried out by Davis [15]. The initial variances (for each estimator) where set equal to $10^6 I$. A square root version of the recursive least square estimator was employed. To avoid poor initial estimates of θ_1 adversely affecting the estimators $\hat{\theta}_2$ and $\hat{\theta}_3$ the following "variable forgetting factor strategy" was employed. For $\hat{\theta}_1$ a constant forgetting factor of unity was employed. For $\hat{\theta}_2$ and $\hat{\theta}_3$ the forgetting factor $\gamma(t)$ was defined to be $[1 - \beta^{-1}(t)]^2$ if $\beta(t) > 1$ and δ otherwise, where $\beta(t)$ is the maximum diagonal element of $P_1^{-1}(t)^{\frac{1}{2}}$. Effectively the updating equations for $P_i(t)$ in (4.4b) and (4.4c) are replaced by:

$$P_i^{-1}(t+1) = \gamma(t)[P_i^{-1}(t) + z_i(t)z_i^T(t)],\ i=2,\ 3.$$

While the estimate of θ_1 is poor (i.e. $P_1(t)$ is large) the forgetting factor is small. The signal to noise ratio Sn in the simulations is defined to be the ratio of the variance of the output when the input is u(t) only to the variance of the output when the input is e(t) only. The simulations in [15] include (among many others) the following examples:

S1 $\quad A = 1 - 0.8q^{-1},$

$$B = 1.0q^{-1},$$

$$C = 1 + 0.7q^{-1}$$

$Sn = 1,\ p = 8$

S2 $\quad A = 1 - 1.5q^{-1} + 0.7q^{-2}$,

$B = 1.0q^{-1} + 0.5q^{-2}$,

$C = 1 - 1.0q^{-1} + 0.2q^{-2}$,

$Sn = 1,\ p = 9$

For S1, a simulation with N = 1000 yielded the following results:

$\hat{\theta}_2$: $\quad \hat{a}_1 = -0,7992 \pm 0.0194$,

$\hat{b}_1 = 1.0068 \pm 0.0182$,

$\hat{c}_1 = 0.6852 \pm 0.0393$

$\hat{\theta}_3$: $\quad \hat{a}_1 = 0.8007 \pm 0.0136$,

$\hat{b}_1 = 1.0023 \pm 0.0133$,

$\hat{c}_1 = 0.6880 \pm 0.0214$

The standard deviations were (crudely) estimated using ten different trials, each with N = 1000.

For S2, using the same procedure, the following results were obtained:

$\hat{\theta}_2$:	$\hat{a}_1 = -1.4982 \pm 0.0055$,	$\hat{a}_2 = 0.6988 \pm 0.0052$,
	$\hat{b}_1 = 1.0035 \pm 0.0081$,	$\hat{b}_2 = 0.5004 \pm 0.0198$,
	$\hat{c}_1 = -0.9522 \pm 0.0492$,	$\hat{c}_2 = 0.1907 \pm 0.0672$
$\hat{\theta}_3$:	$\hat{a}_1 = -1.4995 \pm 0.0036$,	$\hat{a}_2 = 0.6988 \pm 0.0030$
	$\hat{b}_1 = 1.0018 \pm 0.0075$,	$\hat{b}_2 = 0.5011 \pm 0.0144$
	$\hat{c}_1 = -0.9775 \pm 0.0447$,	$\hat{c}_2 = 0.1803 \pm 0.0455$

<u>Table 1</u>

The variation of $\hat{\theta}_3(t)$ with t for S1 is shown in Fig. 1. If in S1 C is replaced by $1 + 0.99z^{-1}$ (so that zC has a zero at -0.99 and thus very close to the unit circle) then $\hat{c}_2$ has a bias of 15% with p = 8 and 8% with p = 15. Similarly, a second order system with z^2C having zeros of magnitude 0.98 yielded biases of 7.8% in $\hat{c}_1$ and 15% in $\hat{c}_2$ with p = 15. In both cases the bias in A and B was negligible. In both S1 and S2 the standard deviations of the estimates were better than published results for a recursive maximum likelihood estimate (N sufficiently large). The results in Table 1 illustrate the reduction of variance achieved by the third step. It can tentatively be concluded that the algorithm is worth considering if the zeros of $z^n C$ do not lie near the unit circle.

7. CONCLUSION

A new multi-step method for estimating the parameters in a controlled ARMA process was proposed in [6]. The first step is a least squares fit of a high order (p) LS model to obtain estimates of the residuals. The following steps are combinations of least squares and filtering. The method proposed in this report is a recursive version of the method proposed in [6]. The properties of the proposed recursive method are investigated by analysis. In particular it has been established that the method is globally convergent. The estimates will converge to the limiting values for large data sets of the off-line method. Since it has been shown that the estimates obtained by the off-line method can be made arbitrarily close to the true process parameters by choosing the dimension p of the first LS fit sufficiently large it is clear that the proposed recursive method has nice computational properties. It thus seems worthwhile to explore the algorithm further. The next logical step is to explore the properties of the estimate for finite data sets. This problem is unfortunately not amenable to analysis. It has to be investigated by simulation. The proposed method thus has to be compared with other recursive methods for a large range of different systems. To do this the computational aspects of the algorithm should also be explored. In particular the possibilities of using fast algorithms and square root algorithms should be investigated.

Provided that the method compares well also for short data sets there are numerous extensions that can be considered. For example

- extensions to multivariable systems
- modifications to the third step of the algorithm
- extensions to state-space models.

More attention needs to be given to the performance of the algorithm for short data sets. The variable forgetting function employed to prevent errors in $\hat{\theta}_1$ from adversely affecting $\hat{\theta}_2$ and $\hat{\theta}_3$ is common to both $\hat{\theta}_2$ and $\hat{\theta}_3$ and depends solely on P_1. It may be preferable to use one variable forgetting factor for $\hat{\theta}_2$ based on P_1 and a second for $\hat{\theta}_3$ based on P_1 and P_2.

However the computational results in [15] suggest that the algorithm is worth considering for those cases when the zeros of $z^n C$ do not lie near the unit circle. Moreover the structure of the algorithm, consisting of a sequence of parallel estimators, the estimator in one level utilizing the estimate yielded in a lower level, may suggest a means for improving other recursive algorithms or even developing entirely new algorithms. The various "levels" represent the on-line version of successive "iterations" of an off-line algorithm. One possibility (the authors are grateful to T J Hannan for information on this point) is to employ any consistent estimator $\hat{\theta}_1$ of (A, B) in the first level, and to employ an on-line approximation to a Newton-step from $\hat{\theta}_1$ for minimizing the likelihood function in the second step.

ACKNOWLEDGEMENTS

The idea for this work originated when one of the authors, D.Q. Mayne, visited Lund Institute of Technology in the spring of 1977. The details were worked out and this report was written while the other author, K.J. Åström, was visiting Imperial College on a Science Research Council Senior Visiting Fellowship. This support from the SRC is gratefully acknowledged.

REFERENCES

[1] Åström K.J. and Bohlin T., Numerical Identification of Linear Dynamic Systems from Normal Operating Records, IFAC Teddington Symposium 1965. Also in Hammond P.H. (Ed.), Adaptive Control, Plenum Press, New Tork, 1966.

[2] Clarke D.W., Generalized Least-Squares Estimation of the Parameters of a Dynamic Model, Paper 3.17, Proc. IFAC Symposium on Identification in Automatic Control Systems, Prague, Czechoslovakia, 1967.

[3] Panuska V., An Adaptive Recursive Least Squares Identification Algorithm, Proc. JACC 1968.

[4] Durbin J., The Fitting of Time-Series Models, Revue Inst. Int. de Stat., 28, (1961) 233-243.

[5] Mayne D.Q. and Firoozan F., An Efficient, Multi-Stage, Linear Identification Method for ARMA Processes, Proc. JACC 1967 p. 435-438.

[6] Mayne D.Q. and Firoozan F., Linear Estimation of ARMA Systems Proc. IFAC 7th World Congress, Helsinki 1978

[7] Söderström T., Ljung L. and Gustavson I., A Comparative Study of Recursive Identification Methods, Report 7427, Department of Automatic Control, Lund Inst. of Technology, Lund, Sweden.

[8] Pandya R.N., Two Stage Least Squares Estimators and Their Recursive Approximations, 3rd IFAC Symposium on Identification and System Parameter Estimation, The Hague, 1973.

[9] Söderström T., Some Notes on Extensions of the Least Squares Identification Method, Report UPTEC 75/OR, Teknikum Institute of Technology, Uppsala University, Uppsala, Sweden, 1975.

[10] Lawson C.L. and Hanson R.J., Solving Least Squares Problems, Prentice Hall, Englewood Cliffs, NJ, 1974.

[11] Lindquist A., A New Algorithm for Optimal Filtering of Discrete-Time Processes, SIAM J. Control, 12 (1974), 736-746.

[12] Morf M., Sidhu G.S. and Kailath T., Some New Algorithms for Recursive Estimation in Constant Linear Discrete-Time Systems, IEEE Trans. AC-19, (1974), 315-323.

[13] Ljung L., Analysis of Recursive Stochastic Algorithms, IEEE Trans AC-22, (1977), 551-575.

[14] Mayne D.Q. and Firoozan F., Linear Estimation of ARMA Systems, Parts I and II, Publications 77/13 and 77/14, Department of Computing and Control, Imperial College of Science and Technology, London.

[15] Davis K.H., MSc Thesis, Imperial College, London 1978.

APPENDIX A

Lemma A1

Assume that $C(q^{-1}) \neq 1$. If $p > 2n$ then the polynomials A_1 and B_1 associated with θ_1^0 have the property $\deg A_1 > n$ or $\deg B_1 > n$.

Proof:

The proof is by contradiction. Hence assume that $\deg A_1 \leq n$ and $\deg B_1 \leq n$. This means that the vector θ_1^0 has the form

$$\theta_1^0 = [\alpha_1\alpha_2\ldots\alpha_n \; 0\ldots0 \; \beta_1\beta_2\ldots\beta_n \; 0\ldots0]^T \tag{A1}$$

where the non zero components have been denoted by α_i and β_i. Introduce also the vector

$$\Psi = [\alpha_1\alpha_2\ldots\alpha_n\beta_1\beta_2\ldots\beta_n] \tag{A2}$$

whose components are the non zero components of θ_1^0.

Furthermore introduce

$$V_1(t) = [-y(t-1)\ldots-y(t-n) \;\; u(t-1)\ldots u(t-n)]^T$$

and

$$V_2(t) = [-y(t-n-1)\ldots-y(t-2n) \;\; u(t-1)\ldots u(t-n)]^T$$

Since the residuals are uncorrellated with the regressors:

$$E[V_2(t)V_1^T(t)]\Psi = EV_2(t)y(t) \tag{A3}$$

It also follows from the process model (2.1) that

$$E[V_2(t)V_1^T(t)]\Psi^0 = EV_2(t)y(t) \tag{A4}$$

where

$$\Psi^0 = [a_1 \ldots a_n \; b_1 \ldots b_n]$$

because $C(q^{-1})e$ is a moving average of order n. Combination of (A3) and (A4) now give

$$E[V_2(t)V_1^T(t)][\Psi-\Psi^0] = 0 \tag{A5}$$

Assuming for the moment that the matrix $EV_2V_1^T$ is nonsingular then it follows from (A5) that

$$\Psi = \Psi^0$$

The least squares residuals are then given by

$$\varepsilon(t) = e(t) + c_1 e(t-1) + \ldots + c_n e(t-n) \tag{A6}$$

Since Θ_1^0 is a least squares estimate the residuals are uncorrelated with the regressors. Hence

$$E\varepsilon(t)y(t-1) = 0$$

$$E\varepsilon(t)y(t-2) = 0$$

$$\vdots$$

$$E\varepsilon(t)y(t-n) = 0$$

Using (A6) it follows from the last of the equations that

$$E\varepsilon(t)y(t-n) = \lambda^2 c_n = 0 .$$

Hence $c_n = 0$.

Similarly it can be shown from the other equations that

$$c_1 = c_2 = c_3 = \ldots = c_{n-1} = 0 .$$

Hence we find that $C(q^{-1}) \equiv 1$ which is the desired contradiction. To complete the Lemma it now remains to show that the matrix $EV_2V_1^T$ is regular. For this purpose introduce the vectors

$$y_1 = [y(t-1)\ldots y(t-n)]^T$$

$$y_2 = [y(t-n-1)\ldots y(t-2n)]^T$$

$$u = [u(t-1)\ldots u(t-n)]^T.$$

The matrix $EV_2V_1^T$ can then be written as

$$EV_2(t)V_1(t)^T = \begin{bmatrix} A & -B \\ -C & D \end{bmatrix} = A_1 \tag{A7}$$

where

$$A = Ey_2y_1^T$$

$$B = Ey_2u^T$$

$$C = Euy_1^T$$

$$D = Euu^T .$$

The matrix D is invertible because it was assumed that the input was persistently exciting (Assumption A2). Transforming the matrix (A7) by the nonsingular transformation matrix

$$T_1 = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}$$

gives

$$T_1 A_1 = \begin{bmatrix} A-BD^{-1}C & 0 \\ -C & D \end{bmatrix}$$

The matrix $A-BD^{-1}C$ can be written as

$$A - BD^{-1}C = Ey_2y_1^T - Ey_2u^T(Euu^T)^{-1}Euy_1^T$$

$$= E\{y_2-Ey_2u^T(Euu^T)^{-1}\}\{y_1-Ey_1u^T(Euu^T)^{-1}\}^T$$

$$= E\{y_2-E[y_2|u]\}\{y_1-E[y_1|u]\}^T = E\tilde{y}_2\tilde{y}_1$$

Furthermore it follows from (2.1) that

$$\tilde{y}_1 = \Phi^k\tilde{y}_2 + L_1e(t) + \ldots + L_ne(t-n+1)$$

where Φ is a nonsingular transition matrix associated with (2.1). Hence

$$E\tilde{y}_2\tilde{y}_1 = \Phi^u E\tilde{y}_2\tilde{y}_2^T$$

But $\tilde{y}_2$ is the error in predicting y_2 based on u. The matrix $E\tilde{y}_2\tilde{y}_2^T$ is thus clearly positive.

APPENDIX B

This appendix contains the lemmas used in the proof of Theorem 1.

<u>Lemma B1</u>

The matrix set $S_1 = Ez_1z_1^T$ is positive definite and bounded.

<u>Proof</u>

The proof is by contradiction. Hence assume that S_1 is not positive definite. Then there exist polynomials

$$F(q^{-1}) = f_1q^{-1} + \dots + f_pq^{-p}$$

and

$$G(q^{-1}) = g_1q^{-1} + \dots + g_pq^{-p}$$

not both zero, such that

$$\text{Var } [F(q^{-1})y(t)-G(q^{-1})u(t)] = 0$$

But since

$$y = \frac{B}{A} u + \frac{C}{A} e$$

it follows that

$$0 = \text{Var } [(F \frac{B}{A} - G)u + \frac{FC}{A} e]$$

$$= \text{Var } [(F \frac{B}{A} - G)u] + \text{Var } (\frac{FC}{A} e)$$

because u and e are assumed to be uncorrelated. Hence

$$\text{Var } (\frac{FC}{A} e) = 0 \tag{B1}$$

and

$$\text{Var } \left(\frac{FB}{A} - G\right)u = 0 \tag{B2}$$

Since e is white noise equation (B1) implies that $F = 0$. Since u is persistently exciting of arbitrary order, equation (B2) implies $G = 0$ and we have thus obtained a contradiction.

That S_1 is bounded follows from our assumptions on u and the stability of A. □

<u>Lemma B2</u>

The matrix $S_2(\Theta_1) = Ez_2z_2^T$ is positive definite and bounded for all $\Theta_1 \in R^{2p}$.

The matrix $S_2(\Theta_1,\Theta_2) = Ez_3z_3^T$ is positive definite and bounded for all $\Theta_1 \in R^{2p}$ such that the degree of A_1 or the degree of B_1 exceeds n and for all $\Theta_2 \in R^{3n}$ such that C_2 is stable.

<u>Proof</u>

If Θ_2 is such that $C_2(q^{-1}) = 1$, then $S_3(\Theta_1,\Theta_2) = S_2(\Theta_1)$. Hence it is only necessary to consider S_3. If S_3 is not positive definite, there exist polynomials

$$F(q^{-1}) = f_1q^{-1} + \ldots + f_nq^{-n}$$

$$G(q^{-1}) = g_1q^{-1} + \ldots + g_nq^{-n}$$

$$H(q^{-1}) = h_1q^{-1} + \ldots + h_nq^{-n}$$

not all zero, such that

$$\mathrm{Var}\ [F(q^{-1})\bar{y}(t)-G(q^{-1})\bar{u}(t)-H(q^{-1})\bar{\varepsilon}_1(t)] = 0$$

But since

$$y = \frac{B}{A}u + \frac{C}{A}e$$

it follows that

$$0 = \mathrm{Var}\ [\frac{F}{C_2}\ (\frac{B}{A}u + \frac{C}{A}e) - \frac{G}{C_2}u - \frac{H}{C_2}\ (A_1y-B_1u)]$$

$$= \mathrm{Var}\ [(\frac{FB}{C_2A} - \frac{G}{C_2} - \frac{HA_1B}{C_2A} + \frac{HB_1}{C_2})u + (\frac{FC}{C_2A} - \frac{HA_1C}{C_2A})e]$$

Because u and e are assumed to be uncorrelated, we obtain

$$\mathrm{Var}\ [\frac{C}{C_2A}\ (F-HA_1)e] = 0$$

and

$$\mathrm{Var}\ [\{\frac{B}{C_2A}\ (F-HA_1) - \frac{1}{C_2}\ (G-HB_1)\}u] = 0$$

Hence

$$F = HA_1$$

and

$$G = HB_1$$

But this contradicts the fact that the degree of A_1 or B_1 exceeds n. Hence S_3 (and S_2) is positive definite for the permitted values of Θ_1 and Θ_2.

The boundedness of S_3 (and S_2) follows from our assumptions on u and the stability of the polynomials A and C_2. □

APPENDIX C

In this appendix the continuity of the functions S_2 and S_3 is shown.

Lemma C1

The function $S_2 : R^{2p} \to R^{3nx3n}$ is continuous. The function $S_3 : R^{2p} \times R \to R^{3nx3n}$ is continuous at all (Θ_1,Θ_2) such that the corresponding polynomials satisfy Assumptions A1.

Proof

As in Lemma B2, we need only consider S_3. $S_3(\Theta_1,\Theta_2)$ contains terms of the following types

(i)	$E\bar{y}(i)\bar{y}(j)$
(ii)	$E\bar{y}(i)\bar{u}(j)$
(iii)	$E\bar{y}(i)\bar{\varepsilon}_1(j)$
(iv)	$E\bar{u}(i)\bar{u}(j)$
(v)	$E\bar{u}(i)\bar{\varepsilon}_1(j)$
(vi)	$E\bar{\varepsilon}_1(i)\bar{\varepsilon}_1(j)$

where $\bar{y} = (1/C_2)y$, $\bar{u} = (1/C_2)u$, and $\bar{\varepsilon}_1 = (1/C_2)(A_1y-B_1u)$.

Consider the first term and let F and G denote the (infinite degree) polynomials satisfying $F = B/AC_2$ and $G = C/AC_2$. Then

$$\bar{y} = Fu + Ge$$

As before we let f denote coeff $[F(q^{-1})]$, and g denote coeff $[G(q^{-1})]$, recalling that the coefficiant of q^0 in F and G is unity. Let ℓ_p denote the metric space of infinite dimensional vectors or infinite sequences with the metric $[\sum_{i=1}^{\infty}(x_i)^p]^{1/p}$, where p is any positive integer. It is known [14], Proposition 11 of Part II) that any map of the type

$$\text{coeff}\,[C(q^{-1})] \rightarrow \text{coeff}\,[C(q^{-1})^{-1}]$$

from R^n to ℓ_1 is continuous at any point where C is stable. It follows from ([14], Proposition 10 of Part II) that the maps $c_2 \rightarrow f$ and $c_2 \rightarrow g$ (where $c_2 \triangleq \text{coeff}\,[C_2(q^{-1})]$) from R^n to ℓ_1 (and, hence, from R^n to ℓ_2) are continuous where C_2 is stable. Since

$$\bar{y}(i) = u(i) + f_1u(i-1) + f_2u(i-2) + \ldots$$

$$+ e(i) + g_1e(i-1) + g_2e(i-2) + \ldots$$

it follows that

$$|E\bar{y}(i)\bar{y}(j)| \leq |E\bar{y}(i)^2|$$

$$\leq \|R_u\|_2\ [1 + \|f\|_2]^2$$

$$+ \lambda^2[1 + \|g\|_2]^2 .$$

where R_u denotes the infinite matrix whose ijth element in Eu(i)u(j). By assumption $||R_u||_2$ is finite. It follows from the continuity of $c_2 \rightarrow f$ and $c_2 \rightarrow g$ that the map $c_2 \rightarrow E\bar{y}(i)\bar{y}(j)$ is continuous. The continuity of $c_2 \rightarrow E\bar{y}(i)\bar{u}(j)$ and $c_2 \rightarrow E\bar{u}(i)\bar{u}(j)$ can be similarly established.

Consider now the second term. Clearly

$$\bar{\varepsilon}_1 = Hu + Ke$$

where H and K are infinite degree polynomials satisfying

$$H = A_1B/C_2A - B_1/C_2$$

$$K = A_1C/C_2A$$

The coefficient of q^0 in H and K is unity. Arguing as above it is seen that the maps $(\Theta_1, C_2) \rightarrow h$, from $R^{2p} \times R^n$ to ℓ_2 and $(a_1, c_2) \rightarrow k$ from $R^p \times R^n$ to ℓ_2 where

$$\Theta_1 \triangleq (a_1, b_1) = \text{coeff}\ [A_1(q^{-1}), B_1(q^{-1})]$$

$$h \triangleq \text{coeff}\ [H(q^{-1})]$$

$$k \triangleq \text{coeff}\ [K(q^{-1})]$$

are continuous where C_2 is stable. Now

$$|E\bar{y}(i)\bar{\varepsilon}_1(j)| \le ||R_u||_2\ [1 + ||f||_2]\ [1 + ||h||_2]$$

$$\lambda^2[1 + ||g||_2]\ [1 + ||h||_2].$$

The continuity of $(\Theta_1, \Theta_2) \rightarrow E\bar{y}(i)\varepsilon_1(j)$ follows. The continuity of $(\Theta_1, \Theta_2) \rightarrow E\bar{u}(i)\bar{\varepsilon}_1(j)$ and $(\Theta_1, \Theta_2) \rightarrow E\bar{\varepsilon}_1(i)\bar{\varepsilon}_1(j)$ can be similarly established.

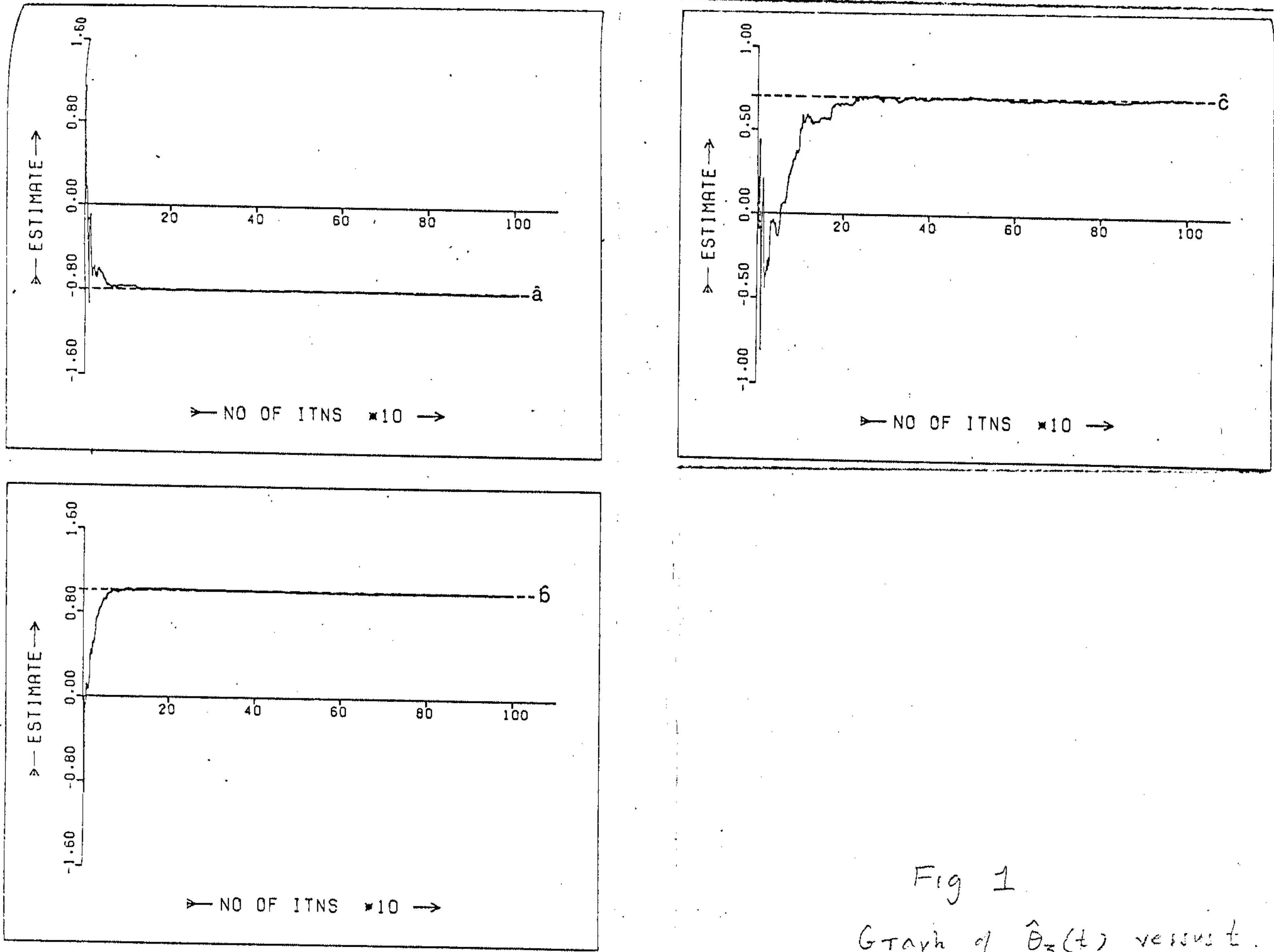

Fig 1

Graph of $\hat{\theta}_3(t)$ versus t.