LEAST SQUARES PARAMETER ESTIMATION ALGORITHMS FOR NONLINEAR SYSTEMS

by

S. A. Billings, BEng,PhD,CEng,MIEE,AFIMA,MInstMC

and

W. S. F. Voon, BEng

Department of Control Engineering
University of Sheffield
Mappin Street, Sheffield S1 3JD

Research Report No. 225

April 1983

Abstract

Least squares parameter estimation algorithms for nonlinear systems are investigated based on a nonlinear difference equation model. A modified extended least squares, an instrumental variable and a new suboptimal least squares algorithm are considered. The problems of input sensitivity, structure detection, model validation and input signal selection are also discussed in the nonlinear context.

1. Introduction

Parameter estimation techniques for nonlinear systems depend critically on the choice of model structure, the source of noise within the system and the input excitation. Classically the model structure used in nonlinear systems identification has been the functional series expansions of Volterra or Wiener [Marmarelis, Marmarelis 1978, Billings 1980]. Most of these expansions however map past inputs into the present output and this inevitably means that a very large number of coefficients are required to characterise the process. For example, even a simple quadratic nonlinearity in cascade with a first order linear dynamic system would require typically 400-500 coefficients to specify the first and second order Volterra kernels. Alternative system descriptions such as block-structured models [Billings, Fakhouri 1982] alleviate most of these difficulties provided the system under study belongs to the relevant class of models. Identification based on the functional series or block-oriented models is however often based on correlation methods which dictate the use of Gaussian white input signals.

Input/output descriptions which expand the current output in terms of past inputs and outputs provide models which can be used to represent a broad class of nonlinear systems using a reasonable number of terms. Moreover the coefficients in these models can be estimated using parameter estimation algorithms which are not dependent upon specialized input signals. Some important classes of models which fall into this category include the Hammerstein model [Narendra and Gallman 1966], polynomic state models [Gary and Boziuk 1972, Netravali and De Figueredo 1971] and classes of nonlinear difference

equations [Sontag 1979, Normand-Cyrot 1982, Billings and Leontaritis 1981,1982]. Because the Hammerstein model can only be used to represent a small class of systems and expansions based on the system states imply that all the states can be measured only nonlinear difference equations will be considered in the present study.

Although nonlinear models based on expansions of lagged inputs and outputs provide a very concise system representation any noise on the measurements enters the model as product terms with the system input and output. Most of the parameter estimation algorithms for linear systems cannot therefore be applied directly because the assumption that the noise terms in the model are independent of the input is violated.

While the choice of input excitation is wide if parameter estimation algorithms are employed the input must excite all the dynamic modes of the system and cover the whole amplitude range of interest. The choice of input can influence the parameter estimates and pretreatment of the data can in some instances produce a model which is input sensitive [Billings and Voon 1983].

The present study is an attempt to clarify some of the above problems and to suggest some possible solutions. A nonlinear difference equation model is introduced in Section 2 and the problems of multiplicative noise, input sensitivity and detection of nonlinearities are discussed. A modified extended least squares and an instrumental variable algorithm based on this model expansion are introduced in Section 3, together with a new sub-optimal least squares routine. It is shown that many of these algorithms will only yield unbiased estimates under various assumptions of the noise. The problems of

model validation are reviewed in Section 4 where various simple to compute correlation functions are introduced to analyse the residuals and indicate which of the least squares algorithms will yield unbiased parameter estimates. The choice of input signals for nonlinear systems is considered in Section 5 and simulation results which illustrate some of the points in the text are included in Section 6.

2. System Representation

If a system is linear then it is finitely realizable and can be represented by the linear difference equation model

$$y(t) = -\sum_{i=1}^{n_y} (a_i y(t-i)) + \sum_{i=1}^{n_u} (b_i u(t-i)) \qquad (1)$$

if the Hankel matrix of the system has finite rank. Almost all the parameter estimation algorithms for linear systems have been based on this model [Goodwin and Payne 1977] which provides a very concise system representation. When the system is nonlinear a similar representation can be derived by considering the observability of nonlinear systems and utilizing results from automata theory to yield the nonlinear difference equation model [Billings and Leontaritis 1981, 1982]

$$y(t) = F_*[y(t-1), y(t-2) \ldots y(t-n_y), u(t-1) \ldots u(t-n_u)] \qquad (2)$$

where $F_*[\cdot]$ is some nonlinear function of u and y. Whereas the Volterra series model [Marmarelis and Marmarelis 1978]

$$y(t) = \sum_{i=1}^{\infty} V_{k_i}(u(i),\ldots u(t)) \qquad (3)$$

where V_{k_i} is a homogeneous degree i polynomial, involves an expansion in terms of past inputs only the nonlinear difference equation model expands the present output in terms of lagged inputs and outputs. This reduces considerably the number of parameters required to represent a nonlinear system and provides a suitable basis for the development of parameter estimation algorithms.

The nonlinear difference equation model eqn (2) will be referred to as a nonlinear ARMAX or NARMAX model. The Hammerstein, Wiener, bilinear, state-affine and other well known nonlinear models can be shown to be special cases of this representation [Billings and Leontaritis 1981,1982].

Consider the NARMAX model of eqn (2) and introduce a time delay d

$$y(t) = F_*^{\ell}[y(t-1),\ldots y(t-n_y)u(t-d),\ldots u(t-d-n_u+1)] \qquad (4)$$

where n_y is the order of lagged outputs, n_u the order of lagged inputs and ℓ represents the degree of nonlinearity. Expanding $F_*[\cdot]$ as a polynomial and defining $V_1 = y(t-1),\ldots V_{n_y} = y(t-n_y)$, $V_{n_{y+1}} = u(t-d),\ldots V_s = u(t-d-n_u+1)$ where $S = n_y+n_u$, the model of eqn (4) can be expressed as

$$y(t) = F_*^{\ell}[V_1,V_2,\ldots V_s] \qquad (5)$$

which expanded yields

$$y(t) = \sum_{i=1}^{S} \theta_i V_i + \sum_{i=1}^{S}\sum_{j=1}^{S} \theta_{ij} V_i V_j + \ldots + \sum_{i=1}^{S}\sum_{j=1}^{S}\cdots\sum_{m=1}^{S}\sum_{n=1}^{S} \theta_{ij\ldots mn} V_i V_j \ldots V_m V_n \qquad (6)$$

(ℓ times)

Assume that the system output y(t) is corrupted by zero mean additive noise e(t) to yield the measured output signal

$$z(t) = y(t) + e(t) \tag{7}$$

Substituting in eqn (6) yields

$$\begin{aligned} z(t) = & \sum_{i=1}^{n_y} \theta_i (z(t-i)-e(t-i)) + \sum_{i=1}^{n_u} \theta_{n_y+i} (u(t-d-i+1)) \\ & + \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} \theta_{ij} (z(t-i)-e(t-i))(z(t-j)-e(t-j)) \\ & + 2 \sum_{i=1}^{n_y} \sum_{j=1}^{n_u} \theta_{i\, n_y+j} (z(t-i)-e(t-i))(u(t-d-j+1)) \\ & + \sum_{i=1}^{n_u} \sum_{j=1}^{n_u} \theta_{n_{y+i} n_{y+j}} (u(t-d-i+1)(u(t-d-j+1)) + \ldots \\ & + \text{higher order terms up to degree } \ell + e(t) \end{aligned} \tag{8}$$

Inspection of eqn (8) shows that even though the noise enters the process as an additive signal at the output it appears in the model as multiplicative terms with the system input and output. Thus although the model is linear-in-the-parameters the inclusion of lagged process outputs introduces cross-product terms between the noise and the process input-output signals. Equation (8) can be expressed more concisely as

$$\begin{aligned} z(t) = F'^{\ell}[& z(t-1),\ldots z(t-n_y)u(t-d),\ldots u(t-d-n_u+1), \\ & e(t-1),\ldots e(t-n_y)] + e(t) \end{aligned} \tag{9}$$

In general noise may corrupt internal measurements within the system. Because the process is nonlinear the superposition principle does not apply and internal noise cannot be translated to be additive

at the output as is usually assumed in linear identification. This situation can however be represented by the model

$$z(t) = F^{\ell}[z(t-1),\ldots z(t-n_y),u(t-d),\ldots u(t-d-n_u+1), e(t),e(t-1),\ldots e(t-n_y)] \qquad (10)$$

Notice that in eqn (10) the term e(t) is included as an argument within $F^{\ell}[\cdot]$. The special case of eqn (9) when additive output noise always results in an isolated term e(t) will not always exist when the noise source is internal. However, in general eqn (10) can be used to represent a wide class of nonlinear systems whatever the noise source.

2.1. Input Sensitivity

It is vitally important in system identification that the model which is fitted to the data should be valid for a wide range of input excitation signals. This property always holds whenever the system and model are linear. When the system and model are nonlinear however the model may become input sensitive. That is the model is only valid for a very small class of input signals. To illustrate this phenomena consider a noisefree first order dynamic model expanded as a second degree polynomial with an input signal $u(t)+b$, $\overline{u(t)} = 0$

$$y(t) = \theta_1 y(t-1)+\theta_2(u(t-1)+b)+\theta_{11}y^2(t-1) +\theta_{12}y(t-1)(u(t-1)+b)+\theta_{22}(u(t-1)+b)^2 \qquad (11)$$

Following the standard procedure of removing the output mean yields

$$y'(t) = y(t)-\bar{y} = \theta_1(y(t-1)-\bar{y})+\theta_2 u(t-1)+\theta_{11}(y^2(t-1)-\overline{y^2}) +\theta_{12}(y(t-1)u(t-1)-\overline{yu} - b\bar{y}) +\theta_{22}(u^2(t-1)-\overline{u^2}+2bu(t-1)) \qquad (12)$$

Inspection of eqn (12) shows that any model relating y'(t) and u(t) will be input sensitive. That is the model will be dependent upon the mean level b and the variance $\overline{u^2(t)} = \sigma_u^2$ of the input and will only yield the correct predicted output for inputs with exactly these statistics. This is a severe limitation and means a new model must be identified every time it is required to compute the predicted process output for any input with a variance unequal to σ_u^2.

This problem can be alleviated by measuring the average output $\overline{y_b(t)}$ of the system with zero input u(t) = 0

$$\overline{y_b} = \theta_1\overline{y_b} + \theta_2 b + \theta_{11}\overline{y^2}_b + \theta_{12}\overline{y_b}\, b + \theta_{22}b^2 \qquad (13)$$

to yield the input/output description

$$y'_b(t) = y(t)-\overline{y_b} = \theta_1(y(t-1)-\overline{y_b})+\theta_2 u(t-1)+\theta_{11}(y^2(t-1)-\overline{y^2}_b)$$

$$+\theta_{12}(y(t-1)u(t-1)+b(y(t-1)-\overline{y_b})) + \theta_{22}(u^2(t-1)+2u(t-1)b) \qquad (14)$$

The dependence of the model on the higher order moments (≥ 2) of the input has therefore been removed but the model will only be valid for inputs around the operating point b.

2.2. Detecting Nonlinearity

Prior to applying parameter estimation algorithms the data should be analysed to determine if it is worthwhile fitting a nonlinear model. This can easily be achieved either by comparing $\overline{z(t)}$ and $\overline{z_b(t)}$ or by computing $\phi_{z'z'^2}(\tau)$ [Billings and Voon 1983]. Inspection of eqns (11) and (13) shows that $\overline{z_b(t)} = \overline{z(t)}$ iff the system is linear and this can

be used as a very simple test for detecting nonlinearities. Alternatively, if it is possible to inject an input u(t)+b into the process where all the odd order moments of u(t) are zero the correlation test $\phi_{z'z'^2}(\tau) = E[z'(t+\tau)(z'(\tau))^2]$ will only be zero for all τ iff the process is linear. Whenever $\phi_{z'z'^2}(\tau) \neq 0$ therefore this indicates that the system under test is nonlinear. It has been found in practice that $\phi_{z_b'z_b'^2}(\tau)$ usually provides a much clearer indication of nonlinear effects compared with $\phi_{z'z'^2}(\tau)$ and the former test should therefore be used whenever $z_b'(t)$ is available.

3. Parameter Estimation

When the system under test is nonlinear it will in general be impossible to solve eqn (10) for e(t) and consequently the noise source e(t) and the prediction errors $\varepsilon(t)$ will not be equal. Reformulating eqn (10) into a prediction error model gives the general form

$$z(t) = F^{\ell}[z(t-1),\ldots z(t-n_y),u(t-d),\ldots u(t-d-n_u+1), \varepsilon(t-1),\ldots \varepsilon(t-n_y)]+\varepsilon(t) \qquad (15)$$

where $$\varepsilon(t) = z(t)-\hat{z}(t) \qquad (16)$$

and $$E[\varepsilon(t)\,|\,z(t-1),z(t-2),\ldots,u(t-d),u(t-d-1)..] = 0 \qquad (17)$$

Since in general the noise source e(t) can neither be measured nor computed from eqn (10) all the parameter estimation algorithms will be developed based on eqn (15).

Expanding eqn (15) and re-grouping terms yields

$$z(t) = G^{zu}[z(t-1)\ldots,z(t-n_y),u(t-d)\ldots u(t-d-n_u+1)]+$$
$$G^{zu\varepsilon}[z(t-1),\ldots z(t-n_y),u(t-d),\ldots u(t-d-n_u+1),\varepsilon(t-1),\ldots\varepsilon(t-n_y)]$$
$$+G^{\varepsilon}[\varepsilon(t-1),\ldots\varepsilon(t-n_y)]+\varepsilon(t) \qquad (18)$$

where $G^{zu}[\cdot]$ is a function of z and u only, $G^{zu\varepsilon}[\cdot]$ represents all the cross product terms involving $\varepsilon(t)$ and $G^{\varepsilon}[\cdot]$ is a polynomial function of the prediction errors only. Separating out the unknown parameters gives

$$z(t) = \psi(t)^T\theta(t-1)+\varepsilon(t)$$
$$= [\psi_{zu}{}^T(t)\psi_{zu\varepsilon}{}^T(t)\psi^T{}_{\varepsilon}(t)] \begin{pmatrix} \theta_{zu}(t-1) \\ \theta_{zu\varepsilon}(t-1) \\ \theta_{\varepsilon}(t-1) \end{pmatrix} +\varepsilon(t) \qquad (19)$$

where $G^{zu}[\cdot] = \psi_{zu}{}^T(t)\theta_{zu}(t-1)$

$G^{zu\varepsilon}[\cdot] = \psi_{zu\varepsilon}{}^T(t)\theta_{zu\varepsilon}(t-1)$

$G^{\varepsilon}[\cdot] = \psi_{\varepsilon}{}^T(t)\theta_{\varepsilon}(t-1)$ (20)

and the definitions of the ψ's and θ's follows easily. Grouping all terms involving $\varepsilon(t)$ and defining

$$\xi(t) = \psi^T{}_{zu\varepsilon}(t)\theta_{zu\varepsilon}(t-1)+\psi^T{}_{\varepsilon}(t)\theta_{\varepsilon}(t-1)+\varepsilon(t) \qquad (21)$$

gives

$$z(t) = \psi^T{}_{zu}(t)\theta_{zu}(t-1)+\xi(t) \qquad (22)$$

Inspection of eqn (21) shows that $\xi(t)$ is highly correlated with the elements of $\psi^T{}_{zu}(t)$. Direct application of least squares will therefore

yield biased estimates and alternative algorithms must be developed to overcome this difficulty.

All the algorithms presented in the following sections will be based on the following unified algorithm [Soderstrom, Ljung and Gustavsson 1974]

$$\hat{\theta}(t+1) = \hat{\theta}(t)+K(t+1)\varepsilon(t+1) \tag{23}$$

$$K(t+1) = \frac{P(t)zz(t+1)}{\lambda(t+1)+\psi(t+1)^{T}P(t)zz(t+1)} \tag{24}$$

$$P(t+1) = \left[P(t) - \frac{P(t)zz(t+1)\psi(t+1)^{T}P(t)}{\lambda(t+1)-\psi(t+1)^{T}P(t)zz(t+1)}\right] /\lambda(t+1) \tag{25}$$

$$\lambda(t+1) = \lambda_{o}\lambda(t)+(1-\lambda_{o}) \tag{26}$$

$$\varepsilon(t) = z(t+1)-\psi(t+1)^{T}\hat{\theta}(t) \tag{27}$$

where $\lambda(t)$ is a variable forgetting factor and the definition and computation of the quantities $\hat{\theta}(t)$, $\varepsilon(t)$, $\psi(t)$, and $zz(t)$ are given below.

3.1. Extended Least Squares

Extended least squares for linear systems has been widely studied and is often referred to as Panuska's method, the extended matrix method of approximate maximum likelihood [Goodwin and Payne 1977]. The extension of the algorithm to nonlinear systems is straightforward. From eqn (19) the algorithm eqn (23) to (27) yields the extended least squares estimate with the following definitions

$$\psi(t+1)^{T} = [\Psi^{T}_{zu}(t+1)\Psi^{T}_{zu\varepsilon}(t+1)\Psi_{\varepsilon}{}^{T}(t+1)] \tag{28}$$

$$\theta(t) = [\theta_{zu}(t)\theta_{zu\varepsilon}(t)\theta_{\varepsilon}(t)]^T \tag{29}$$

$$zz(t) = \psi(t) \tag{30}$$

For example, a first order dynamic system with second degree polynomial expansion would be represented as

$$\psi(t+1)^T = [z(t),u(t),\varepsilon(t),z^2(t),z(t)u(t),z(t)\varepsilon(t), u^2(t),u(t)\varepsilon(t),\varepsilon^2(t)]$$

$$\hat{\theta}(t) = [\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_{11},\hat{\theta}_{12},\hat{\theta}_{13},\hat{\theta}_{22},\hat{\theta}_{21},\hat{\theta}_{33}]^T \tag{31}$$

The convergence properties of this algorithm when applied to linear systems is well documented and is known to depend upon the positive realness of the noise pulse transfer function [Soderstrom, Ljung and Gustavvson 1978]. These results do not carry over when the model is nonlinear because the noise model includes multiplicative terms between the prediction errors $\varepsilon(t)$ and the measured input/output sequences. Although simulations have shown that the algorithm has good convergence properties a theoretical study of the method when applied to nonlinear systems is in progress.

The major disadvantage of the extended least squares algorithm when applied to nonlinear systems is the need to include noise or prediction error terms in the estimation vector. The maximum number of entries in the $\hat{\theta}(t)$ vector is given by

$$n = \sum_{i=1}^{\ell} [n_{i-1}(n_y+n_u+n_\varepsilon+i-1)]/i \tag{32}$$

$$n_o = 1$$

and if the nonlinearity within the system is severe the dimension of $\hat{\theta}(t)$ increases rapidly. In an attempt to limit the dimension of the vector $\hat{\theta}(t)$ both instrumental variables and a new suboptimal least squares algorithm were developed.

3.2. Instrumental Variables

The principle of instrumental variables has been applied to linear system identification in several ways [Goodwin and Payne 1977]. The off-line description of these algorithms is based on the linear model $Y = \phi\theta+e$ and the selection of an instrument matrix V^T which satisfies the conditions

$$\lim \frac{1}{N} V^T\phi = R \quad \text{a nonsingular matrix}$$

$$\lim \frac{1}{N} V^T(Y-\phi\theta_o) = 0 \tag{33}$$

where θ_o denotes the true parameter vector and lim refers to limit in probability. The conditions of eqn (33) ensure that the estimate

$$\hat{\theta} = (V^T\phi)^{-1}V^TY \tag{34}$$

is unbiased. There are a number of ways of satisfying eqn (33) the most popular use either delayed inputs to form V^T or define V^T to have the same structure as ϕ^T but with the measured outputs replaced by predicted outputs [Young 1970]. This latter algorithm is often referred to as the auxiliary model algorithm. Unfortunately instrumental variables can only be applied to nonlinear systems providing certain properties of the system noise are satisfied. Consider the NARMAX model of eqn (22)

$$z(t) = \psi^T_{zu}(t)\theta_{zu}(t-1) + \xi(t) \tag{35}$$

where

$$\xi(t) = \psi^T_{zu\varepsilon}(t)\theta_{zu\varepsilon}(t-1) + \psi^T_{\varepsilon}(t)\theta_{\varepsilon}(t-1) + \varepsilon(t) \tag{36}$$

For a sequence of N output measurements define

$$\underline{Z} = [z(1),\ldots z(N)]^T$$

$$\underline{\Psi}_{zu} = [\psi^T{}_{zu}(1)\ldots\psi^T{}_{zu}(N)]^T \qquad (37)$$

$$\underline{\xi} = [\xi(1),\ldots\xi(N)]^T$$

to yield the description

$$\underline{Z} = \underline{\Psi}_{zu}\theta_{zu} + \underline{\xi} \qquad (38)$$

Forming an instrument matrix V^T using the auxiliary model algorithm gives the instrumental variable estimate

$$\hat{\theta}_{zu} = (V^T\underline{\Psi}_{zu})^{-1}V^T\underline{Z} \qquad (39)$$

which will in general be biased whenever the process under test is nonlinear. A similar conclusion follows even if V^T is formed from delayed inputs. This problem arises because in general

$$\lim \frac{1}{N} V^T(\underline{Z}-\underline{\Psi}_{zu}\theta_{zu}{}^o) = \lim \frac{1}{N} V^T(\underline{\Psi}^T{}_{zu\varepsilon}\theta_{zu\varepsilon}+\underline{\Psi}_{\varepsilon}\theta_{\varepsilon}+\underline{\varepsilon}) \neq 0 \quad (40)$$

A typical term in eqn (40) takes the form

$$\lim \frac{1}{N} V^T[z^i u^j \varepsilon^k] \qquad \text{for some i,j,k and}$$

will not in general be zero even when $\varepsilon(t)$ is a zero mean white noise sequence. For example consider a simple NARMAX model

$$y(t) = \theta_1 y^2(t-1) + \theta_2 u(t-1) + \theta_3 u^2(t-1) \qquad (41)$$

with output additive noise

$$z(t) = y(t) + e(t) \qquad (42)$$

The model to be estimated takes the form

$$z(t) = \theta_1 z^2(t-1)+\theta_2 u(t-1)+\theta_3 u^2(t-1)-2\theta_1 z(t-1)\varepsilon(t-1)$$
$$+\theta_1 \varepsilon^2(t-1)+\varepsilon(t) \qquad (43)$$

and the matrices in eqn (37) are given by

$$\Psi_{zu}(t) = [z^2(t-1), u(t-1), u^2(t-1)]^T \tag{44}$$

$$\theta_{zu} = [\theta_1, \theta_2, \theta_3]^T \tag{45}$$

$$\Psi^T_{zu\varepsilon}(t)\theta_{zu\varepsilon}(t-1) = -2\theta_1 z(t-1)\varepsilon(t-1) \tag{46}$$

$$\Psi_{\varepsilon}(t)\theta_{\varepsilon}(t-1) = \theta_1 \varepsilon^2(t-1) \tag{47}$$

Formulate the instrument matrix V^T using the auxiliary model technique and assume that an unbiased estimate of the predicted output $\hat{z}(t) = y(t)$ is available. A typical column V_t in V^T takes the form

$$V_t = [\hat{z}^2(t-1), u(t-1), u^2(t-1)]^T$$

Substituting into eqn (40)

$$\lim \frac{1}{N} V^T(\Psi^T_{zu\varepsilon}\theta_{zu\varepsilon} + \Psi^T_{\varepsilon}\theta_{\varepsilon} + \varepsilon)$$

$$= \lim \frac{1}{N} \begin{bmatrix} \sum_{t=1}^{N} [-2\theta_1 \hat{z}^2(t-1)z(t-1)\varepsilon(t-1) + \theta_1 \hat{z}^2(t-1)\varepsilon^2(t-1) + \hat{z}^2(t-1)\varepsilon(t)] \\ \sum_{t=1}^{N} [-2\theta_1 u(t-1)z(t-1)\varepsilon(t-1) + \theta_1 u(t-1)\varepsilon^2(t-1) + u(t-1)\varepsilon(t)] \\ \sum_{t=1}^{N} [-2\theta_1 u^2(t-1)z(t-1)\varepsilon(t-1) + \theta_1 u^2(t-1)\varepsilon^2(t-1) + u^2(t-1)\varepsilon(t)] \end{bmatrix} \tag{48}$$

which taking the most favourable assumptions that u(t) is zero mean and $\varepsilon(t)$ is a zero mean white noise sequence which is independent of u(t) reduces to

$$= \begin{bmatrix} -\theta_1 \phi_{yy}(o)\phi_{\varepsilon\varepsilon}(o) \\ -\theta_1 \phi_{uu}(o)\phi_{\varepsilon\varepsilon}(o) \end{bmatrix} \tag{49}$$

Clearly the matrix in eqn (49) will not be zero as required by eqn (33) and biased estimates of the process parameters will therefore result. In general therefore the instrumental variable algorithm will yield biased estimates when the system under investigation is nonlinear. The degree of bias does depend upon the particular model expansion used in each case and may even go to zero if k in the expression $\lim \frac{1}{N} V^T[z^i u^j \varepsilon^k]$ is odd or the terms of $\xi(t)$ are small compared to Ψ_{zu}. One particular special case of the NARMAX model which always satisfies these conditions arises when the noise terms are represented within the model as a linear map $L[\cdot]$

$$\begin{aligned} z(t) &= F^{\ell}[z(t-1),\ldots z(t-n_y),u(t-d),\ldots u(t-d-n_u+1)] \\ &\quad + L[\varepsilon(t-1),\ldots \varepsilon(t-n_\varepsilon)] + \varepsilon(t) \end{aligned} \tag{50}$$

These conditions can however only be tested after the parameter estimation is complete. Details of how this can be achieved are presented in Section 4.

The off-line instrumental variable estimator as defined in eqn (39) can be implemented using the recursive equations (23)-(27) with the following definitions of the matrices

$$\begin{aligned} \psi(t+1)^T &= [\Psi^T{}_{zu}(t+1)] \\ \hat{\theta}(t) &= [\theta_{zu}(t)] \\ zz(t) &= [\Psi_{\hat{zu}}(t+1)] \end{aligned} \tag{51}$$

3.3. Suboptimal Least Squares

The number of parameters to be estimated in the NARMAX model increases significantly if the noise model is included in the estimation vector, as in the extended least squares algorithm for example. It would therefore be advantageous if unbiased process parameter estimates could be obtained without specifically estimating a noise model. This can be achieved for the NARMAX model whenever the noise enters as an additive signal at the output by using a new sub-optimal least squares algorithm described below.

Consider the NARMAX model with output additive noise defined by eqns (6),(7) and expressed concisely as

$$z(t) = F'^{\ell}[z(t-1),\ldots z(t-n_y),u(t-d),\ldots u(t-d-n_u+1), e(t-1),\ldots e(t-n_y)]+e(t) \quad (52)$$

The expansion of eqn (52) in terms of a polynomial function induces cross-product terms between the output additive noise and the measured process inputs and outputs. If however y(t) the noisefree process output or predicted output could be monitored eqn (52) could be expressed as

$$z(t) = F'^{\ell}[y(t-1),\ldots y(t-n_y),u(t-d)\ldots u(t-d-n_u+1)]+e(t) \quad (53)$$

and all the cross-product noise terms are eliminated. Parameter estimation based on equation (53) would therefore require significantly less computational effort compared with extended least squares. Although the signal y(t) is unavailable for measurement it can be estimated by recursively computing the predicted output using

$$\hat{y}(t) = \Psi_{\hat{y}u}^{T}(t)\hat{\theta}_{yu}(t-1) \quad (54)$$

The noisefree output $y(t)$ in eqn (53) is therefore effectively replaced by the estimate $\hat{y}(t)$. This algorithm which was derived independently by the authors specifically for nonlinear systems has recently been published by Moore (1982) who considered the linear case and established convergence results.

The algorithm is computationally simple and can be implemented using eqns (23)-(27) with the following definitions

$$\begin{aligned} \psi(t+1)^T &= \Psi_{\hat{y}u}{}^T(t+1) \\ \theta(t) &= \left[\theta_{zu}(t)\right] \\ zz(t) &= \psi(t) \end{aligned} \tag{55}$$

Notice that if the noise enters the system internally, from eqn (10), cross product terms between the input, output and noise are inherently a part of the output signal. These cross product terms if they exist cannot be eliminated, they appear in $\Psi^T_{zu\varepsilon}(t)$ and may induce bias. Fortunately this situation can be detected using the model validity tests discussed in Section 4.

4. Model Validation

It can easily be shown that the traditional linear covariance tests of computing the autocorrelation of the residuals and cross-correlating the input and residuals often fail to indicate a deficient model when the process is nonlinear [Billings and Voon 1983]. This may mislead the investigator into believing that there is no further information in the residuals when additional nonlinear terms may exist which if omitted will induce biased estimates. When the system is nonlinear the residuals should be unpredictable from all linear and nonlinear combinations of past inputs and outputs [Billings

and Voon 1983] and this condition will hold iff

$$\begin{aligned} \phi_{\xi\xi}(\tau) &= \delta(\tau) \\ \phi_{u\xi}(\tau) &= 0 \;\forall\; \tau \\ \phi_{\xi\xi u}(\tau) &= E\left[\xi(k)\xi(k-1-\tau)u(k-1-\tau)\right] = 0 \;\forall\; \tau \end{aligned} \tag{56}$$

where $\xi(t)$ represents the residuals.

The conditions of eqn (56) must always be satisfied when using the Extended Least Squares algorithm which includes a noise model in the estimation vector. Instrumental variables and suboptimal least squares do not specifically estimate a noise model and consequently the residuals may be coloured. Specific tests are required for use in conjunction with these two methods therefore which test the validity of the process model without testing the whiteness of the residuals. When using the instrumental variables routine or sub-optimal least squares any terms in $G^{zu}[\cdot] = \Psi_{zu}^{T}(t)\theta_{zu}(t-1)$ or $G^{zu\varepsilon}[\cdot] = \Psi^{T}_{zu\varepsilon}(t)\theta_{zu\varepsilon}(t-1)$ as defined in eqn (20) which appear in the residuals will induce biased estimates. The condition that none of these terms exists in the residuals will hold iff

$$\begin{aligned} \phi_{u\xi}(\tau) &= 0 \;\forall\; \tau \\ \phi_{u^2\xi}(\tau) &= \phi_{u^2{}'\xi}(\tau) = 0 \;\forall\; \tau \\ \phi_{u^2{}'\xi^2}(\tau) &= 0 \;\forall\; \tau \end{aligned} \tag{57}$$

5. Choice of Input Excitation

The choice of input excitation can significantly influence the results of any identification study and whenever possible the input should be selected to excite all the modes and amplitudes of interest in the system. Because of the ease of generation of pseudo random sequences and the fact that they can be tailored to the process and exhibit an autocorrelation function which approaches an impulse they have become widely used in linear system identification.

Unfortunately, this success has not been carried over for nonlinear systems. Initially pseudo random sequences were studied as a replacement for Gaussian white noise which forms the basis of many nonlinear identification algorithms based on correlation analysis. However, it can be shown [Barker and Pradisthayon 1970] that unless the system has a particular structure and a compound input is used [Billings and Fakhouri 1979,1980] anomalies which occur in the higher even order autocorrelation functions induce errors in the estimates. This problem arises because the higher order moments of pseudo random sequences do not approach those of Gaussian white noise. Whilst this latter property is of great importance in correlation analysis it is not a requirement when using parameter estimation techniques. Indeed one of the advantages of parameter estimation methods is that the choice of inputs is very wide. Even with this relaxation of requirements a pseudo random input sequence appears to be inappropriate whenever the system is nonlinear. The input excitation for nonlinear systems must excite all the dynamic modes over the complete amplitude range of interest. The first requirement specifies the spectral density of the input the second the probability density function.

It is this latter requirement which is vitally important whenever the system under test is nonlinear. If a system with a nonlinearity at the input was excited by a pseudorandom binary sequence for example only the two points corresponding to the amplitude levels of the prbs would be excited on the nonlinear characteristic and the overall model fitted may well turn out to be linear because the probability density function of the input consists of two impulse functions. In a similar manner a Gaussian input to the same system effectively weights the parameter estimates with the probability density function such that characteristics around the mean of the input will be estimated with a greater accuracy than those either side. The ideal input in this situation would be a uniformly distributed signal with a flat probability density function over the whole amplitude range of interest. More work is required to establish suitable inputs for nonlinear system identification but in the meantime the uniformly distributed sequence appears to offer a reasonable compromise.

6. Simulation Results

The parameter estimation routines developed in previous sections have been tested on the two models defined below.

(i) A Hammerstein model with first order dynamics, third degree nonlinearity and output additive noise

$$y(t) = 0.8y(t-1) + 0.4\left[u(t-1)+u^2(t-1)+u^3(t-1)\right]$$
$$z(t) = y(t) + e(t) \qquad (58)$$

was simulated to generate 500 input/output data pairs where u(t) was a uniformly distributed signal with $p(u) = \frac{1}{2}$, $-1 \leq u(t) \leq 1$ and e(t) was a Gaussian white noise sequence N(0.0,0.2).

The mean levels $\bar{z}_b = 0$, $\bar{z} = 0.7242$ indicate that the process is nonlinear and this is confirmed by inspection of the structure detection tests illustrated in Fig.1, $\phi_{z'z'^2}(\tau) \neq 0$, $\phi_{z_b'z_b'^2}(\tau) \neq 0$. The application of recursive least squares produced biased estimates (indicated in this example by $\phi_{\xi\,\xi}(\tau) \neq \delta(\tau)$, $\phi_{u\xi}(\tau) \neq 0$, Fig.2 $\phi_{\xi\,\xi\,u}(\tau) = 0$) whereas the extended least squares algorithm gave unbiased results ($\phi_{\xi\,\xi}(\tau) = \delta(\tau)$, $\phi_{u\xi}(\tau) = 0, \phi_{\xi\,\xi\,u}(\tau) = 0 \forall \tau$) as did the instrumental variable and suboptimal least squares routines ($\phi_{u\xi}(\tau) = 0$, $\phi_{u^2'\,\xi}(\tau) = 0$, $\phi_{u^2'\,\xi^2}(\tau) = 0 \forall \tau$). These results which agree with the theoretical predictions are summarised in table 1.

(ii) An implicit NARMAX model with first order dynamics and second degree nonlinearity described as

$$\begin{aligned} y(t) &= 0.4y^2(t-1) + 0.2u(t-1) + 0.6u^2(t-1) \\ z(t) &= y(t) + e(t) \end{aligned} \tag{59}$$

was simulated with a sinusoidal input of amplitude 1.0 and a Gaussian white noise sequence e(t) N(0.0,0.2) to generate 500 data pairs. The structure detection test illustrated in Fig.3 shows that $\phi_{z_b'z_b'^2}(\tau) \neq 0$, indicating that a nonlinear model is appropriate and this is confirmed by $\bar{z}_b = 0$ and $\bar{z} = 0.4543$. As expected the model validity tests showed that the recursive least squares estimates were biased, $\phi_{\xi\,\xi}(\tau) \neq \delta(\tau)$, $\phi_{u\xi}(\tau) \neq 0$, Fig.4 $\phi_{\xi\,\xi\,u}(\tau) = 0$. Because the system contains a nonlinearity in the output the instrumental variable estimate is biased as predicted. In this particular example this was indicated by the results $\phi_{u^2'\,\xi^2}(\tau) = 0$, $\phi_{u^2'\,\xi}(\tau) \neq 0$, $\phi_{u\,\xi}(\tau) \neq 0$. The model validity tests indicated that both the extended least squares and suboptimal least squares estimates were both unbiased. The results are summarised in table 2.

7. Conclusions

Least squares parameter estimation techniques have been investigated based on a nonlinear difference equation model. Although initially the problems appear to be a trivial extension of the linear case further investigation shows that noise in nonlinear systems whether internal or additive at the output considerably complicates the estimation problem. A modified extended least squares and a new suboptimal least squares algorithm were introduced as one possible solution to yield unbiased estimates in the presence of multiplicative noise terms. Instrumental variables which is widely used for linear systems was shown to yield biased estimates whenever the noise model cannot be expressed as a linear map.

Deficiencies in the well used residual tests based on first order covariance functions may lead the experimentor to believe that his models are correct when they are in fact strongly biased due to omitted nonlinear effects. Two simple to compute second degree correlation functions were introduced to overcome this problem. Methods of detecting nonlinearities in the data prior to estimation, input sensitivity and the selection of input signals were also discussed.

Coefficient	Theoretical Value	RLS	ELS	RIV	SOLS
z(t-1)	0.8	0.7483	0.7865	0.7884	0.7867
u(t-1)	0.4	0.4315	0.4198	0.4323	0.4232
$u^2(t-1)$	0.4	0.4552	0.4241	0.4056	0.4235
$u^3(t-1)$	0.4	0.3703	0.4001	0.3726	0.3909
e(t-1)	-0.8	-	-0.8084	-	-

Table 1. Parameter Estimates for model 1

Coefficient	Theoretical Value	RLS	ELS	RIV	SOLS
u(t-1)	0.2	0.2574	0.2066	0.2165	0.2067
$z^2(t-1)$	0.4	0.2798	0.3851	0.3634	0.3949
$u^2(t-1)$	0.6	0.6628	0.5890	0.6023	0.6014
z(t-1)e(t-1)	-0.8	-	-0.7321	-	-
$e^2(t-1)$	0.4	-	0.6600	-	-

Table 2. Parameter Estimates for model 2

8. References

Barker, H.A., Pradisthayon, T., 1970, Proc. Inst. Elect. Engs., 117, 1857.

Billings, S.A., 1980, Proc. Inst. Elec. Engs., 127, 272.

Billings, S.A., Fakhouri, S.Y., 1979, Proc. Inst. Elect. Engs., 126, 1018; 1980, Int. J. Systs. Sci., 11, 261; 1982, Automatica, 18, 15.

Billings, S.A., Leontaritis, I.J., 1981, Proc.IEE Conference, Control and its Applications, Warwick Univ., 183; 1982, 6th IFAC Symp. Ident. & Syst. Par. Est., Washington DC, 427.

Billings, S.A., Voon, W.S.F., 1983, Proc. Inst. Elect. Engs., to appear.

Garg, D.P., Boziuk, J.S., 1972, Int. J. Control, 15, 1121.

Goodwin, G.C., Payne, R.L., 1977, Dynamic system identification: Experiment design and data analysis, Academic Press, N. York.

Marmarelis, P.Z., Marmarelis, V.Z., 1978, Analysis of physiological systems - the white noise approach, Planum Press.

Moore, J.V., 1982, IEEE Trans. Aut. Control, AC-27, 1189.

Narendra, K.S., Gallman, P.G., 1966, IEEE Trans. Aut. Control, AC-11, 546.

Netravali, A.H., De Figuieredo, R.J.P., 1971, IEEE Trans. Aut. Control, AC-16, 28.

Normand-Cyrot, D., 1982, Proc. American Control Conf., Washington DC, 466.

Soderstrom, T., Ljung, L., Gustavsson, I., 1974, Rp 7427 Lund Inst. Tech.; 1978, Automatica, 14, 231.

Sontag, E.D., 1979, Polynomial response maps, Springer-Verlag.

Young, P.C., 1970, Automatica, 6, 271.

Acknowledgements

The authors gratefully acknowledge that this work is supported by SERC under grant GR/B/31163.

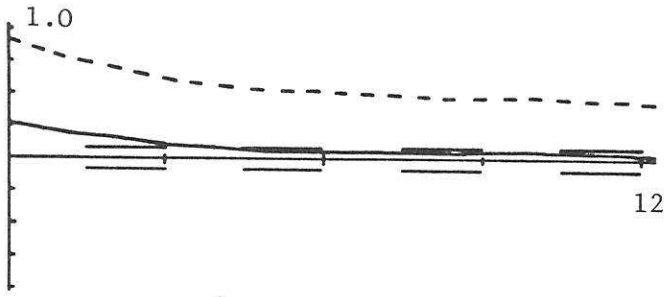

Fig. 1. Structure Detection Test

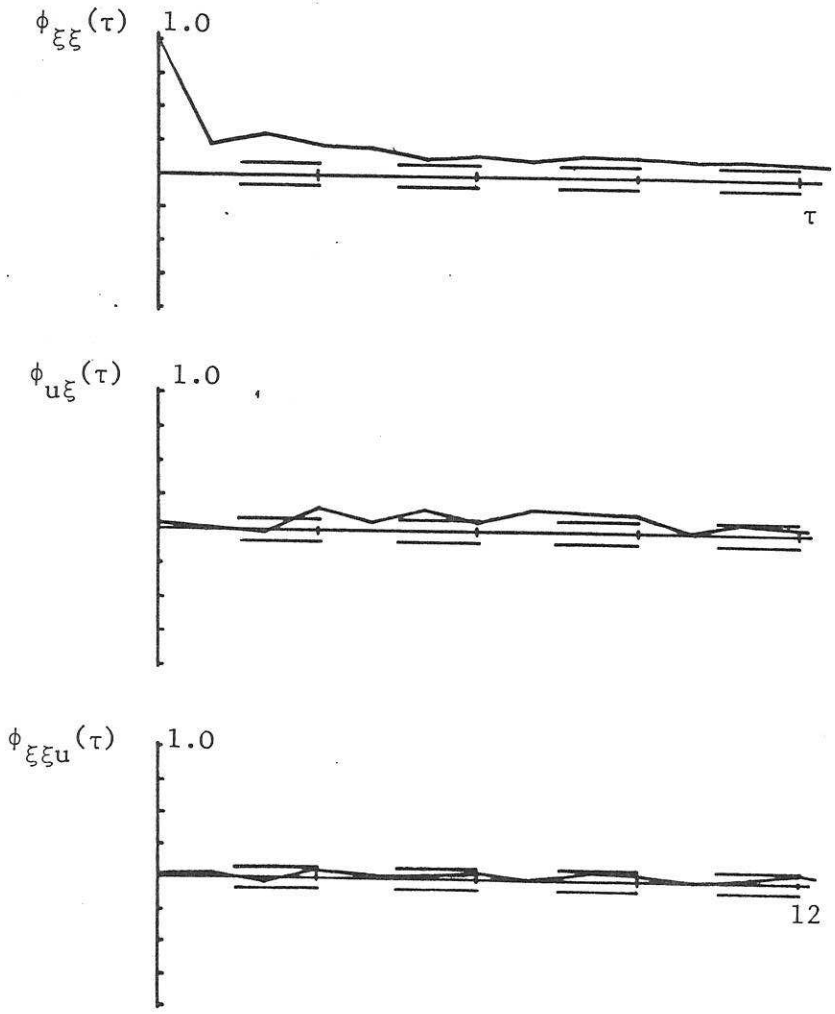

Fig. 2. Model Validity Tests for Recursive Least Squares

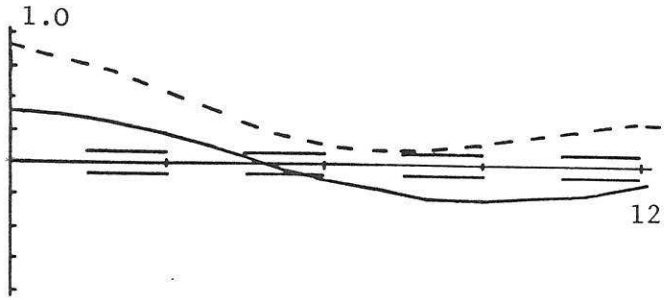

Fig. 3. Structure Detection Test

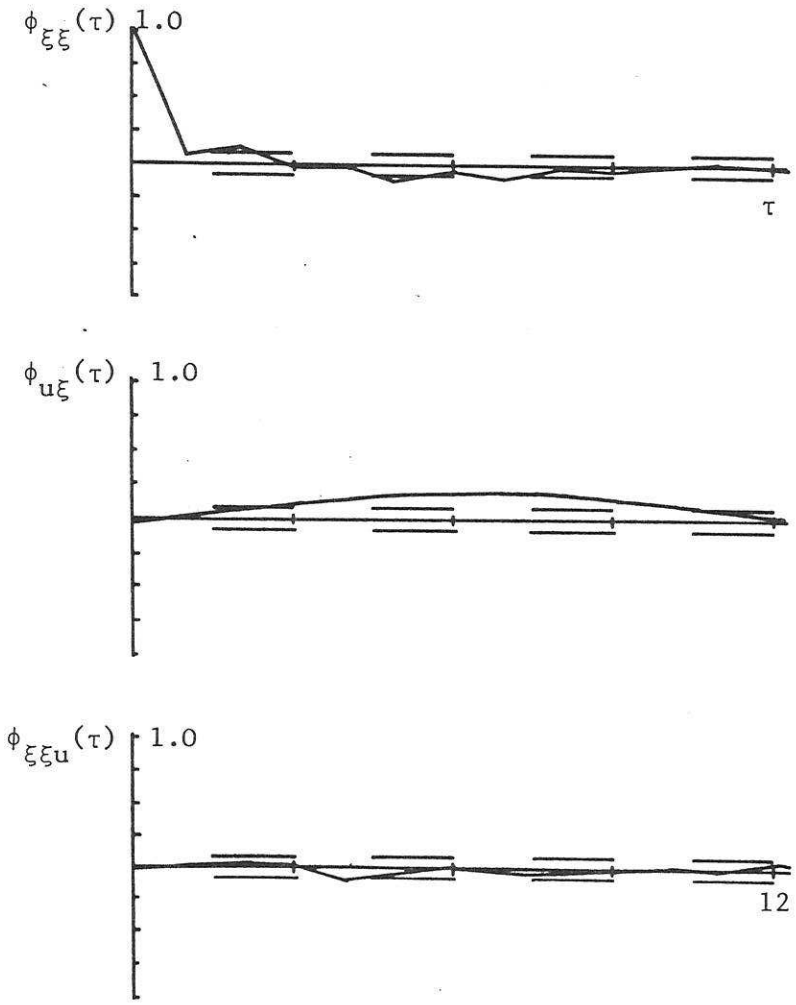

Fig. 4. Model Validity Tests for Recursive Least Squares