Brief Paper

An Innovation Representation for Nonlinear Systems with Application to Parameter and State Estimation*

M. S. AHMED†

Abstract—An innovation model is derived for a nonlinear stochastic system described by a state variable representation. The problem of state and system parameter estimation is solved through identification of the innovation model. A recursive prediction error (RPE) algorithm is derived for the joint system parameter and state estimation through minimization of the innovation variance (MIV). The algorithm is robust against the use of an erroneous model. Convergence and stability properties of the algorithm are also analyzed. In order to ensure stability, the algorithm needs an on-line stability check at each iteration.

1. Introduction

LINEAR MODELS are very popular to simplify the analysis and design of control systems. Linear representations could be over simplifications for the phenomena they describe. Controller design based on an oversimplified model not only leads to performance deterioration, but may also make the controlled system unstable. As a result, the analysis and design of nonlinear systems are of importance to system and process control engineers.

This paper considers the on-line state and parameter estimation problem of nonlinear systems in the presence of modeling errors and/or measurement noise. The most popular method of state estimation in nonlinear systems is the extended Kalman filter (EKF) (Jazwinski, 1970). The extensions include linearization at different stages of the algorithm, model linearization and the jump matrix method. All of these methods assume that the covariances of the noise processes are known. It is well known that the performance of EKF deteriorates or may even diverge in the case of erroneous models.

A state estimation procedure for nonlinear systems is presented through an innovation representation. An innovation model is derived for nonlinear stochastic systems given by a state variable description. The innovation representation is exact when the state equation is nonlinear but the output equation is linear.

The state estimation problem is solved through identification of the innovation gain. A recursive prediction error (RPE) algorithm (Ljung and Söderström, 1983) is derived for joint state estimation and innovation gain identification through minimization of the innovation variance (MIV). It does not require the knowledge of the noise covariances. Since the algorithm is based upon direct minimization of the

 Corresponding author Dr M. S. Ahmed. Tel. +1 313 593 5420; Fax +1 313 593 9967; E-mail mahmed@umdsun2.umd.umich.edu.
† Department of Electrical and Computer Engineering, University of Michigan, Dearborn, MI 48128-1491, U.S.A.

innovation variance, it is also robust against modeling errors. MIV can be extended to simultaneous state estimation and system parameter identification. The paper discusses the derivation of the algorithm, its convergence and stability properties, and presents some simulation studies.

Application of the RPE algorithm for simultaneous parameter and state estimation in linear systems is considered by Moore and Weiss (1979), Ljung and Söderström (1983), Omani and Sinha (1987) and Söderström and Stoica (1989). In the present paper, RPE is applied to state and parameter estimation in nonlinear systems. In addition, a continuous–discrete algorithm is presented to incorporate systems in continuous time description.

In a recent paper, Dhingra et al. (1992) proposed a jump matrix method (JMT) for the state estimation problem in nonlinear systems. The method is based on a Kalman estimator which approximates the nonlinear system as a memoryless nonlinearity embedded in a dynamic linear system using fictitious samplers and clamps. Through simulation study, the authors have shown that JMT outperforms the EKF in terms of mean square error. In this simulation study, the proposed MIV is compared with the EKF and JMT, and their advantages and disadvantages are highlighted.

2. An innovation representation

Consider a nonlinear system described by the state variable representation

$$\dot{x}(t) = g(\bar{x}, u, t) + \Gamma(t)w(t) \qquad (1)$$

$$y(t) = h(\bar{x}, u, t) + v(t), \qquad (2)$$

where $\bar{x}(t) \in \mathcal{R}^n$, $u(t) \in \mathcal{R}^r$, $y(t) \in \mathcal{R}^m$, $w(t) \in \mathcal{R}^q$, $v(t) \in \mathcal{R}^m$ and $t \in \mathcal{R}$ are the state, control input, measured output, state disturbance, measurement noise and time, respectively. $\Gamma(t)$ is a time varying $n \times q$ matrix. $w(t)$ and $v(t)$ are assumed to be zero mean, and individually and mutually uncorrelated (white) random sequences with variances $\Sigma_w(t)$ and $\Sigma_v(t)$, respectively. (Even if the state noise is nonwhite, the composite system can still be put into the above form by augmenting the state vector $x(t)$.)

The usual method of joint parameter and state estimation from the representation of equations (1) and (2), is the use of the extended Kalman filter (EKF). However, EKF requires that the variances $\Sigma_w(t)$ and $\Sigma_v(t)$ be known. When the assumed values of these quantities differ substantially from the true values, the filter performance deteriorates.

The above system representation can be discretized by integrating equation (1) in the time interval t_k to t_{k+1}, which gives (Jazwinski, 1970)

$$\bar{x}_{k+1} = \bar{x}_k + f(g, \bar{x}_k, u_k, t_k) + w_k \qquad (3)$$

$$y_{k+1} = h(\bar{x}_{k+1}, u_{k+1}, t_{k+1}) + v_{k+1}, \qquad (4)$$

where

$$\bar{x}_k \equiv \bar{x}(t_k), \quad u_k \equiv u(t_k), \quad y_k \equiv y(t_k), \quad v_k \equiv v(t_k),$$

$$f(g, \bar{x}_k, u_k, t_k) = \int_{t_k}^{t_{k+1}} g(\bar{x}, u, \tau)\,d\tau$$

and the sequences $\{w_k\}$ and $\{v_k\}$ are two zero mean mutually and individually uncorrelated discrete sequences.

Now define $\hat{\bar{x}}_{k/j}$ as the estimate of $\bar{x}_k$ from the observations up to time t_j. The linear estimation of the sampled state values $\bar{x}_k$ from the discrete observations y_k can be obtained as

$$\hat{\bar{x}}_{k+1/k} = \hat{\bar{x}}_{k/k} + f(g, \hat{\bar{x}}_{k/k}, u_k, t_k) \tag{5}$$

$$\hat{\bar{x}}_{k+1/k+1} = \hat{\bar{x}}_{k+1/k} + L_{k+1}\bar{y}_{k+1/k} \tag{6}$$

with

$$\bar{y}_{k+1/k} = y_{k+1} - h(\hat{\bar{x}}_{k+1/k}, u_{k+1}, t_{k+1}), \tag{7}$$

where L_{k+1} is an $n \times m$ gain matrix to be optimally chosen according to a criterion to be specified below.

Next defining $x_k \equiv \hat{\bar{x}}_{k/k-1}$, $x_k^* \equiv \hat{\bar{x}}_{k/k}$ and $e_k \equiv \bar{y}_{k/k-1}$ equations (6), (5) and (7) may be written, respectively, as

$$x_k^* = x_k + L_k e_k \tag{8}$$

$$x_{k+1} = x_k^* + f(g, x_k^*, u_k, t_k) \tag{9}$$

$$y_{k+1} = h(x_{k+1}, u_{k+1}, t_{k+1}) + e_{k+1}. \tag{10}$$

Equations (8)–(10) constitute an alternate representation for the nonlinear system described by equations (1) and (2).

If L_k in the above state estimation scheme is chosen to minimize the mean squared error, then the sequence $\{e_k\}$ becomes individually uncorrelated. Indeed, subtracting equation (10) from equation (4) and using a first-order approximation of $h(.)$, gives

$$e_{k+1} \cong \frac{\partial h(\bar{x}_{k+1}, u_{k+1}, t_{k+1})}{\partial \bar{x}} \tilde{x}_{k+1/k} + v_{k+1} \tag{11}$$

where

$$\tilde{x}_{k+1/k} \equiv \bar{x}_{k+1} - \hat{\bar{x}}_{k+1/k} = \bar{x}_{k+1} - x_{k+1}.$$

The first-order approximation is not necessary if the output equation is linear. Then equation (11) becomes

$$e_{k+1} = \bar{h}(t_{k+1})\tilde{x}_{k+1/k} + v_{k+1}. \tag{12}$$

For a mean square error estimation sequence $\{L_k\}$ must be chosen such that the estimation error $\tilde{x}_{k+1/k}$ is made orthogonal (uncorrelated) to y_j, $j \leq k$ (Papoulis, 1991). Further, v_{k+1} being white, it is also uncorrelated with y_j, $j \leq k$. Therefore, from equation (11) [or equivalently equation (12)] it can be observed that e_{k+1} is also uncorrelated with y_j, $j \leq k$. In other words e_k is white and is an innovation sequence.

With $x_k = \hat{\bar{x}}_{k/k-1}$ being the mean square estimate of $\bar{x}_k$, the representation given by equations (8)–(10) becomes an innovation representation. This representation is exact if the output equation is linear. However, in the presence of nonlinearity in the output equation, it involves an approximation as shown in equation (11).

3. Application to parameter and state estimation

The innovation representation can be used advantageously for parameter identification and/or state estimation. Gain sequence L_k is, in general, time varying. However, assuming that gains change slowly with time, algorithms can be developed for simultaneous parameter and state estimation.

Consider the innovation description given by equations (8)–(10) and the following criterion function

$$V(\theta) = \frac{1}{2}\mathscr{E}\{\hat{e}_k^T\hat{\Sigma}_e^{-1}(k)\hat{e}_k\}, \tag{13}$$

where $\mathscr{E}\{.\}$ denotes the expectation, $\{\hat{.}\}$ denotes the estimate, $\Sigma_e(k)$ is the covariance of e_k, and the row vector θ of dimension n_θ contains all the parameters to be estimated including any unknown system parameters and the elements of gain matrix L_k. Since sequence $\{e_k\}$ is white with covariance $\Sigma_e(k)$, the identification problem may be formulated as minimization of $V(\theta)$. In single output systems this corresponds to minimization of the innovation variance (MIV). Hence this approach will be called the MIV algorithm.

Define

$$\hat{x}_k^* = \hat{x}_k + \hat{L}_k\hat{e}_k \tag{14}$$

$$\hat{x}_{k+1} = \hat{x}_k^* + f(g, \hat{x}_k^*, u_k, t_k, \hat{\theta}_k) \tag{15}$$

$$\hat{y}_{k+1} = h(\hat{x}_{k+1}, u_{k+1}, t_{k+1}, \hat{\theta}_k) \tag{16}$$

$$\hat{e}_{k+1} = y_{k+1} - \hat{y}_{k+1} \tag{17}$$

$$\Psi_k = -\frac{d\hat{e}_k}{d\theta}\bigg|_{\theta=\hat{\theta}_k} = \frac{d\hat{y}_k}{d\theta}\bigg|_{\theta=\hat{\theta}_k} \quad (\text{dim: } m \times n_\theta), \tag{18}$$

where $\hat{\theta}_k$ is the estimated value of θ after the kth observation and $\hat{y}_k$ is the estimated output from the assumed parameter values. The gradient matrix Ψ_k provides a descent direction for the minimization of $V(\theta)$. In order to compute Ψ_k the following variables need to be defined.

$$\Lambda_k^* \equiv \frac{d\hat{x}_k^*}{d\theta}\bigg|_{\theta=\hat{\theta}_k} \quad \text{and} \quad \Lambda_{k+1} \equiv \frac{d\hat{x}_{k+1}}{d\theta}\bigg|_{\theta=\hat{\theta}_k} \quad (\text{dim: } n_\theta \times n). \tag{19}$$

Further defining

$$F_k^x \equiv \frac{\partial f}{\partial x}\bigg|_{x=\hat{x}_k^*, \theta=\hat{\theta}_k} \quad (\text{dim: } n \times n) \tag{20}$$

$$H_k^x \equiv \frac{\partial h}{\partial x}\bigg|_{x=\hat{x}_k, \theta=\hat{\theta}_k} \quad (\text{dim: } m \times n) \tag{21}$$

$$F_k^\theta \equiv \frac{\partial f}{\partial \theta}\bigg|_{x=\hat{x}_k^*, \theta=\hat{\theta}_k} \quad (\text{dim: } n \times n_\theta) \tag{22}$$

$$H_k^\theta \equiv \frac{\partial h}{\partial \theta}\bigg|_{x=\hat{x}_k, \theta=\hat{\theta}_k} \quad (\text{dim: } m \times n_\theta) \tag{23}$$

$$L_k^\theta \equiv \frac{\partial(L_k e_k)}{\partial \theta}\bigg|_{x=\hat{x}_k, \theta=\hat{\theta}_k} \quad (\text{dim: } n \times n_\theta). \tag{24}$$

where $h \equiv h(\hat{x}_k, u_k, t_k, \theta)$ and $f \equiv f(g, x_k^*, u_k, t_k, \theta)$, one gets the following recursive relations

$$\Lambda_k^* = \Lambda_k - L_k\Psi_k + L_k^\theta \tag{25}$$

$$\Lambda_{k+1} = \Lambda_k^* + F_k^x\Lambda_k^* + F_k^\theta \tag{26}$$

$$\Psi_{k+1} = H_{k+1}^x\Lambda_{k+1} + H_{k+1}^\theta. \tag{27}$$

3.1. *The discrete estimation algorithm.* With the above definitions and following the approach of Ljung and Soderström (1983) an RPE algorithm for the recursive minimization of $V(\theta)$ may be derived as follows.

(i) Compute the prediction error and the innovation covariance matrix

$$\hat{e}_k = y_k - \hat{y}_k \quad (\text{dim: } m \times 1) \tag{28}$$

$$\Sigma_e(k) = \Sigma_e(k-1) + \frac{1}{k}[\hat{e}_k\hat{e}_k^T - \Sigma_e(k-1)]$$

$$(\text{dim: } m \times m). \tag{29}$$

(ii) Compute the parameter adaptation gain

$$M_k = [\lambda\Sigma_e(k) + \Psi_k P_{k-1}\Psi_k^T]^{-1}\Psi_k P_{k-1} \quad (\text{dim: } m \times n_\theta). \tag{30}$$

(iii) Update the parameters

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \hat{e}_k^T M_k \quad (\text{dim: } 1 \times n_\theta). \tag{31}$$

(iv) Update the parameter covariance matrix

$$P_k = \frac{1}{\lambda}[P_{k-1} - P_{k-1}\Psi_k^T[\lambda\Sigma_e(k) + \Psi_k P_{k-1}\Psi_k^T]^{-1}\Psi_k P_{k-1}]$$

$$(\text{dim: } n_\theta \times n_\theta). \tag{32}$$

(v) Compute the next state and output estimates using equations (14)–(16).

(vi) Compute the gradient Ψ_{k+1} using equations (25)–(27).

(vii) Set $k = k + 1$ and go to (i).

λ is a forgetting factor with $0 \leq \lambda \leq 1.0$, which accommodates for the time-varying nature of the parameters. λ may be held constant or may be generated by the following equation

$$\lambda(k) = \lambda_\infty - \lambda_r[\lambda_\infty - \lambda(k-1)]; \quad \lambda(0) = \lambda_0, \tag{33}$$

where λ_0 and λ_∞ are the initial and final values of λ, and λ_r controls the rate at which the transition takes place. To start the iteration P_0 may be set to αI ($\alpha > 0$), elements of $\hat{\theta}_0$ may be set to the best guess or to small values (≈ 0.1) and elements of Λ_0, Λ_0^* and Ψ_0 may be set to zero.

The above algorithm requires discretization of $g(.)$ as shown in equation (3). Using Euler's integration technique one may write

$$f(g, \bar{x}_k, u_k, t_k) = [t_{k+1} - t_k] g(\bar{x}_k, u_k, t_k).$$

However, discretization using other integration procedures is also possible. A continuous–discrete algorithm that does not require the plant discretization is presented below.

3.2. *The continuous–discrete estimation algorithm.* Using the relationships following equation (3), a continuous–discrete algorithm can be developed as well. The continuous–discrete equivalent of equations (14)–(16) and equations (25)–(27) can be obtained, respectively, as

$$\begin{cases} \hat{x}_k^* = \hat{x}_k + \hat{L}_k \hat{e}_k; \quad \xi_k = \hat{x}_k^* & (34) \\[2mm] \xi_{k+1} = \xi_k + \int_{t_k}^{t_{k+1}} g(\xi, u, \tau, \hat{\theta}_k)\,d\tau; \quad \hat{x}_{k+1} = \xi_{k+1} & (35) \\[2mm] \hat{y}_{k+1} = h(\hat{x}_{k+1}, u_{k+1}, t_{k+1}, \hat{\theta}_k) & (36) \end{cases}$$

$$\begin{cases} \Lambda_k^* = \Lambda_k - L_k \Psi_k + L_k^\theta & (37) \\[2mm] \Lambda_{k+1} = \Lambda_k^* + \left(\int_{t_k}^{t_{k+1}} G_k^\xi\,dt \right)\Lambda_k^* + \int_{t_k}^{t_{k+1}} G_k^\theta\,dt, & (38) \\[2mm] \Psi_{k+1} = H_{k+1}^\xi \Lambda_{k+1} + H_{k+1}^\theta & (39) \end{cases}$$

where

$$G_k^\xi \equiv \left.\frac{\partial g}{\partial \xi}\right|_{\theta = \hat{\theta}_k} \quad (\text{dim: } n \times n) \tag{40}$$

$$G_k^\theta \equiv \left.\frac{\partial g}{\partial \theta}\right|_{\theta = \hat{\theta}_k} \quad (\text{dim } n \times n_\theta) \tag{41}$$

and $g \equiv g(\xi, u, t, \theta)$.

Equation (35) involves the solution of differential equations that can be implemented using any numerical procedure. In addition, the evaluation of Λ_{k+1} using equation (38) may be incorporated with the above numerical procedure. An alternate efficient but simpler method of computing Λ_{k+1} is the use of the trapezoidal integration. With appropriate modifications, the discrete algorithm given by equations (28)–(33) can be extended to accommodate the above continuous–discrete equations.

3.3. *Known system parameters.* When the system parameters are known, the derivative computation by equations (25)–(27) is simplified. In this case, θ contains only the elements of L_k. As a result, F_k^θ (or equivalently G_k^θ) and H_k^θ become null matrices. Further denoting

$$\theta = [l_1, \ldots, l_n],$$

where l_i denotes the ith row of L_k, L_k^θ becomes

$$L_k^\theta = \begin{bmatrix} \hat{e}_k^T & & \\ & \ddots & \\ & & \hat{e}_k^T \end{bmatrix} \quad (\text{dim: } n \times n_\theta). \tag{42}$$

The convergence and stability properties of the proposed algorithm are analyzed in the Appendix.

4. Simulation studies

In order to verify the theoretical findings and compare the proposed algorithm (MIV) with other methods, five examples are considered. The first two examples are taken from Dhingra et al. (1992). The other three examples are modified forms of the first one. The other methods for parameter and state estimation considered in the simulation are the extended Kalman filter (EKF) (Jazwinski, 1970) and the jump matrix technique (JMT) (Dhingra et al., 1992).

Example 1. State estimation in Van der Pol's equation. In this example a form of Van der Pol's equations is considered.

Van der Pol's equation originally arose in the theory of circuits containing a triode valve. The solution of many other nonlinear equations can also be converted into this form. Van der Pol's equation driven by a Gaussian wideband input $w(t)$ can be obtained as (Dhingra et al., 1992)

$$\ddot{y} - \mu(1 - y^2)\dot{y} + 9y = w(t) \tag{43}$$

which can be written as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -9x_1 + \mu(1 - x_1^2)x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t) \tag{44}$$

$$y(t) = x_1(t) + v(t), \tag{45}$$

where $v(t)$ depicts the additive measurement noise. In our study μ is taken as 2.0. Two values of the sampling interval h are considered. They are 0.01 and 0.05 s. The differential equations are solved using a fourth-order Runge–Kutta method. In all the simulations involving Van der Pol's equation, x_0 and $\hat{x}_0$ are taken as $[5\ 0]^T$ and $[6\ 2]^T$, respectively. For the MIV and EKF the continuous–discrete versions are implemented.

All the system parameters are assumed to be known in this example. Therefore, MIV requires estimation of L_k only. P_0 and L_0 are taken as $0.1I$ and $[0.1\ 0.1]^T$, respectively. Other values of P_0, i.e. I, $10I$ and $100I$, are also considered. Large values of P_0 caused relatively large initial errors in the state estimates. But in all cases, the estimated states settled down near the true states within 40 iterations. It may be emphasized here that in MIV, P_0 corresponds to the initial confidence on $\hat{\theta}_0$ (not on $\hat{x}_0$). A small value of P_0 adjusts the unknown parameters conservatively and prevents initial oscillation of the estimated parameters. For the MIV, λ has been fixed at 0.99 and the integration in equation (38) is performed using a trapezoidal rule.

Figure 1 shows the actual and estimated state values with MIV when h is set to 0.05. The variances of $w(t)$ and $v(t)$ both are taken as 1. In order to get an appreciation of the noise levels, it may be mentioned here that the noise free output $x_1(k)$ had a mean square value ≈ 1. After the initial convergence period the tracking ability of the algorithm is clearly evident. The on-line estimation of L_k is depicted in Fig. 2. The time varying nature of these parameters is clearly evident from the figure. When the simulation is carried on for a longer time, these parameters are found to swing slowly confirming their nonstationary nature.

Tables 1 and 2, respectively, compare the three methods in terms of the mean square error between the estimated and the real state values for $h = 0.01$ and $h = 0.05$. The values of $\hat{\Sigma}_w$ and $\hat{\Sigma}_v$ needed by the EKF and JMT, are both set to $10I$. Table 1 indicates that at low noise levels the performance of EKF is best with MIV being comparable, but at high noise levels JMT outperforms EKF and the performance of MIV becomes worst.

As the value of h was increased to 0.05, the performance of JMT deteriorated. The algorithm started showing instability. Although in some cases it has been possible to stabilize it by varying the initial state covariance matrix, $\hat{x}_2$ showed large initial oscillations in the range ± 100. All these oscillations needed about 40 iterations to die out. In other cases no values of the initial covariance matrix succeeded to stabilize it. These results are shown in Table 2, where two columns of JMT results are given, one corresponding to the total mean square error, the other corresponding to the mean square error computed after the oscillations has died out. The results in these tables show that JMT is effective at a high sampling rate but is unsuitable for low frequency sampling. This observation is consistent with the JMT principle that it approximates the dynamic nonlinearity by a static nonlinearity embedded into a dynamic linear equations. Such an approximation may only be justified for a high sampling frequency. The performance of EKF and MIV in this table follows the same explanation as in Table 1.

However, the last rows of Tables 1 and 2 show superiority of the MIV. The inferior performance of EKF and JMT in these rows is due the fact that the ratio $(\hat{\Sigma}_w/\hat{\Sigma}_v)$ has been too different from the actual values. The performance of EKF
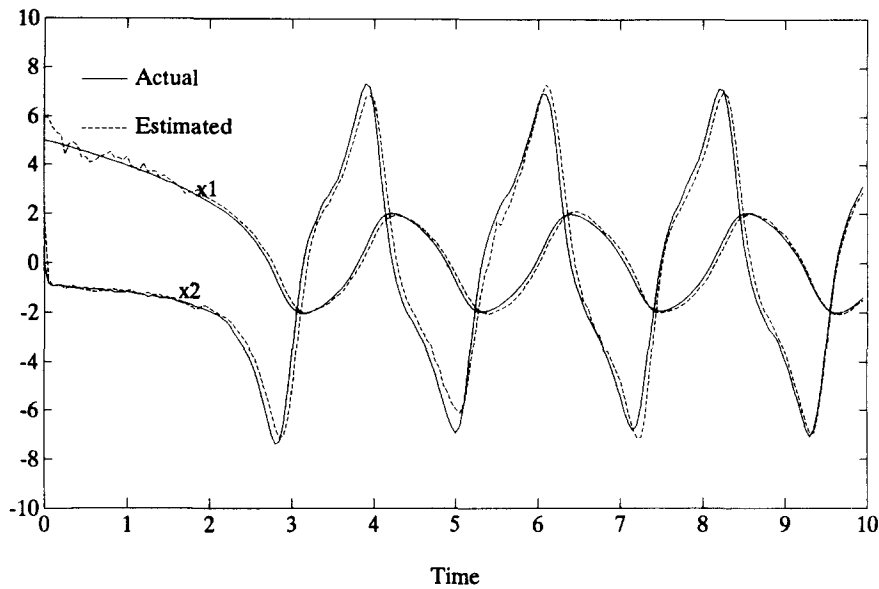
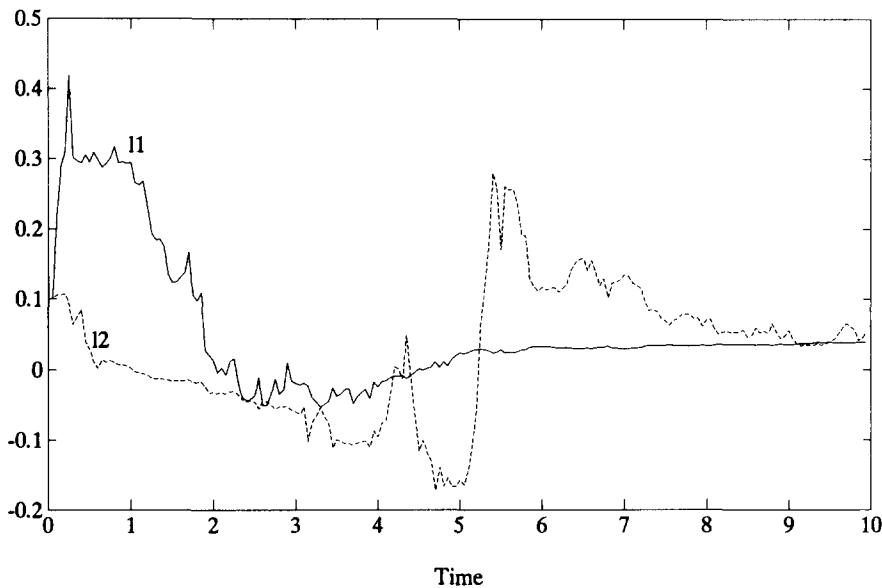FIG. 1. Actual and estimated states (MIV): Van der Pol's equation.

FIG. 2. Estimates of the innovation gains (MIV): Van der Pol's equation.

and JMT therefore may deteriorate when knowledge of the noise variances is imprecise. MIV, however, is free from this problem.

Example 2. State estimation in a nonlinear control system. In this example a nonlinear relay control system provided by

Kuo (1970) and used by Dhingra *et al.* (1992) is considered. The block diagram of the control system is given in Fig. 3. It consists of an amplifier followed by a second-order linear system connected in a feedback configuration. The slope of

TABLE 1. MEAN SQUARE ERROR IN STATE ESTIMATION. VAN DER POL'S EQUATION ($h = 0.01$)

Σ_w	Σ_v	EKF	JMT	MIV
0.01	0.01	0.0076	0.0168	0.0090
0.10	0.01	0.0083	0.0149	0.0099
0.10	0.10	0.0091	0.0147	0.0265
1.00	0.10	0.0168	0.0141	0.0133
1.00	1.00	0.0244	0.0194	0.1534
10.0	1.00	0.0984	0.0696	0.3627
1.00	10.0	0.0925	0.0801	2.6109
10.0	10.0	0.1706	0.1340	3.1623
10.0	0.01	0.0872	0.0602	0.0301

TABLE 2. MEAN SQUARE ERROR IN STATE ESTIMATION. VAN DER POL'S EQUATION ($h = 0.05$)

Σ_w	Σ_v	EKF	JMT	JMT*	MIV
0.01	0.01	0.0300	22.586	0.3173	0.0397
0.10	0.01	0.0317	22.628	0.3326	0.0411
0.10	0.10	0.0368	107.64	0.3515	0.0563
1.00	0.10	0.0562	108.96	0.4246	0.0813
1.00	1.00	0.1066	333.00	1.6647	0.4075
10.0	1.00	0.3708	unstable		0.9053
1.00	10.0	0.6613	unstable		3.4566
10.0	10.0	0.9512	unstable		3.6103
10.0	0.01	0.2754	23.514	0.9244	0.1906

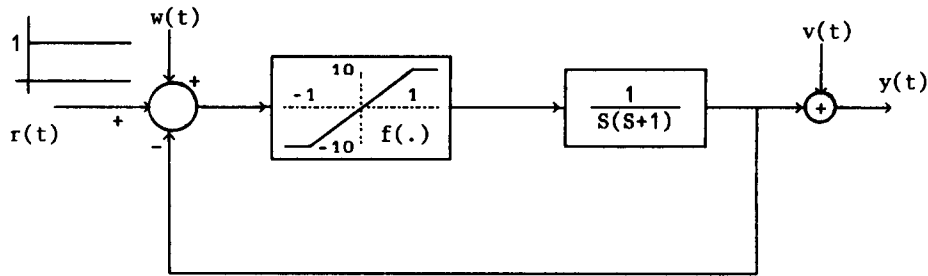*When first 40 iterations are excluded from the MSE calculation.

FIG. 3. Block diagram of the nonlinear control system.

the linear part of the amplifier characteristic is taken as 10. The equation describing the system is given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -x_2 + f(r - x_1 + w) \end{bmatrix} \quad (46)$$

$$y(t) = x_1(t) + v(t), \quad (47)$$

where f is the amplifier characteristic and, $w(t)$ and $v(t)$ are the feedback noise and measurement noise, respectively.

The system has been initially simulated using a sampling period of 0.05. Figure 4 shows the system and estimated states by MIV when the variances of $w(t)$ and $v(t)$ have been 1.0 and 0.1, respectively. Tables 3 and 4, respectively, compare the three methods for $h = 0.01$ and $h = 0.05$. It can be observed that MIV outperformed the other two methods. In this example, the partial derivatives can not be computed accurately due to the piecewise linear amplifier saturation curve. In addition, the term $w(t)$ being an argument of the nonlinear function provided nonstationary state noise. It is possible that these facts are responsible for the inferior performance of EKF and JMT. On the other hand, MIV adjusts the innovation gain to minimize the model mismatch and therefore has superior modeling error tolerance.

Example 3. State estimation with an erroneous model. In this example Van der Pol's equation is reconsidered. However, an erroneous model is used for the state estimation. Although $\mu = 2$ has been used for system simulation, the state estimation algorithms assumed μ as 0.5. Figures 5 and 6 show the actual and estimated states due to the EKF, JMT and MIV, respectively. The variances of $w(t)$ and $v(t)$ both are taken as 1.0. h is taken as 0.05. It can be observed that EKF and JMT failed to track the states due to the imprecise

TABLE 3. MEAN SQUARE ERROR IN STATE ESTIMATION. NONLINEAR CONTROL SYSTEM ($h = 0.01$)

Σ_w	Σ_v	EKF	JMT	MIV
0.01	0.01	0.0401	0.0908	0.0209
0.10	0.01	0.0627	0.1146	0.0337
0.10	0.10	0.0646	0.1163	0.0395
1.00	0.10	0.0924	0.1281	0.0841

TABLE 4. MEAN SQUARE ERROR IN STATE ESTIMATION. NONLINEAR CONTROL SYSTEM ($h = 0.05$)

Σ_w	Σ_v	EKF	JMT	MIV
0.01	0.01	0.1130	0.2051	0.0427
0.10	0.01	0.1418	0.2244	0.0604
0.10	0.10	0.1519	0.2324	0.0643
1.00	0.10	0.2266	0.3283	0.1304

modeling assumption. On the other hand, although a little bumpy, MIV estimated the states reasonably well. The estimated innovation gains are shown in Fig. 7. A comparison of Figs 2 and 7 shows that the gains assumed higher values in the latter, which has been needed to compensate for the modeling error.

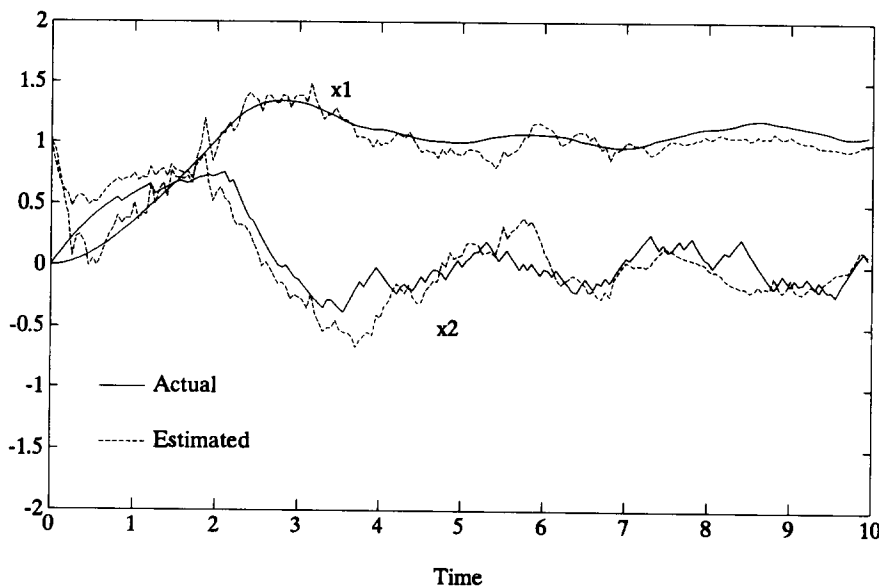*Example 4. State estimation with an unknown system parameter.* In this example Van der Pol's equation is

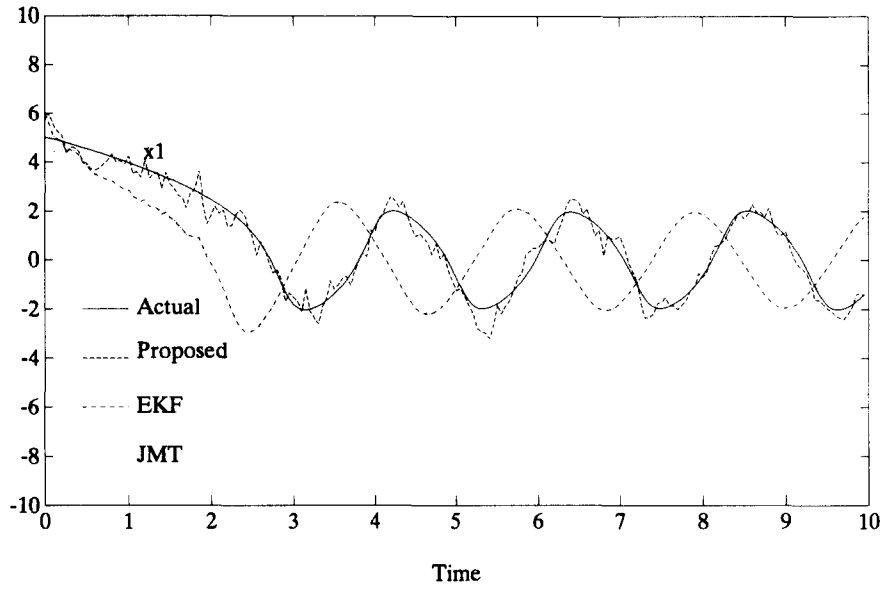FIG. 4. Actual and estimated states (MIV): nonlinear control system.

Brief Papers

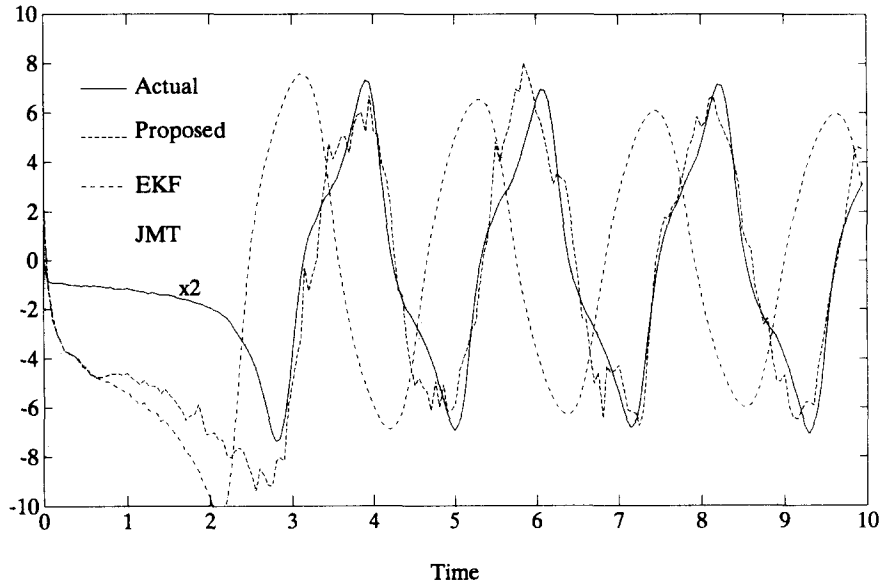FIG. 5. Comparison of the methods; estimation of x_1.

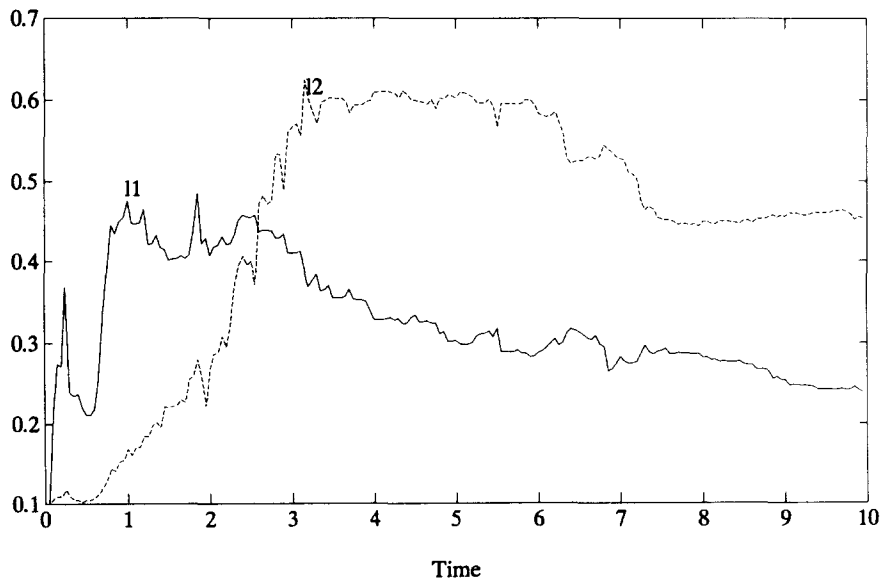FIG. 6. Comparison of the methods; estimation of x_2.

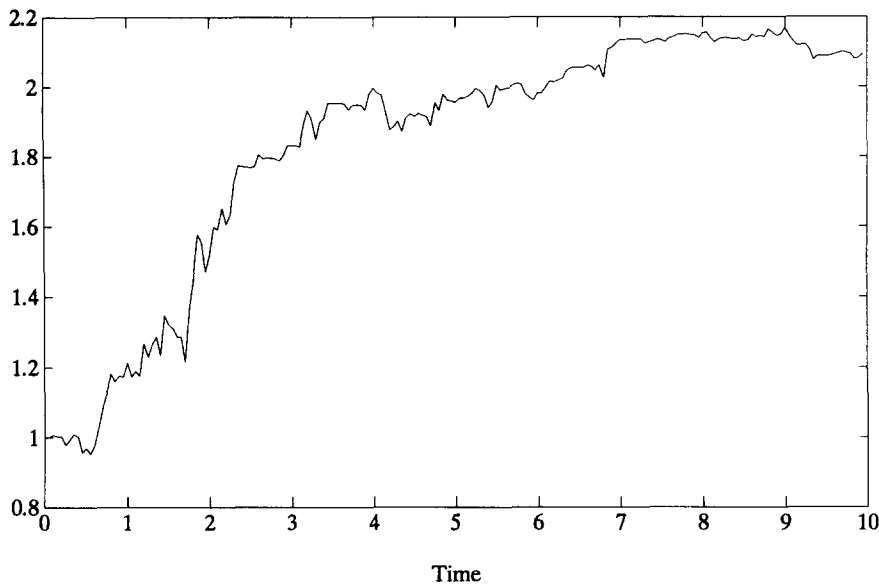FIG. 7. Estimates of the innovation gains (MIV); erroneous model.

FIG. 8. Online estimation of μ (MIV); joint parameter and state estimation.

considered again but μ is assumed as an unknown parameter. In MIV the unknown parameter vector θ, therefore, consists of the elements of L_k as well as μ. $\hat{\mu}_0$ is taken as 1 (the actual value being 2). The variances of $w(t)$ and $v(t)$ are both taken as 1.0. h is taken as 0.05. It is observed that MIV has been able to track the states and as well as estimate the unknown parameter. Figure 8 shows the on-line estimation of the parameter μ. The convergence of $\hat{\mu}$ to the proper value is clearly evident from the figure.

Example 5. State estimation with nonlinearity in the output equation. In this example Van der Pol's equation is modified by introducing nonlinearity in the output equation. The state equation is the same as in Example 1, but the output equation has been modified to

$$y(t) = 0.05[x_1(t)]^3 + v(t). \tag{48}$$

Variances of $w(t)$ and $v(t)$ both have been 1.0. h has been taken as 0.05. Although the innovation representation is approximate in this simulation, the state estimation by MIV has been found to be satisfactory. Further the autocorrelation of the estimated residual sequence revealed that $\hat{e}_k$ has been white with 95% confidence.

5. Conclusion and discussion

In this paper, a robust state and parameter estimation method in nonlinear systems has been proposed through minimization of the innovation variance (MIV). The performance of the algorithm is compared with the extended Kalman filter (EKF) and jump matrix method (JMT).

All the examples provided in this paper revealed the effectiveness of MIV to estimate the system parameters and the states of nonlinear systems. In the presence of stationary state and output measurement noise, and with a good model of the system, EKF is generally superior. With identical conditions but with small values of sampling time h, JMT outperforms EKF at high noise level. However, in the presence of nonstationary state and measurement noise, when the statistics of the state and measurement noise are not known properly, and/or, when the system model is imprecise the proposed MIV algorithm clearly outperforms the EKF-type algorithms. MIV always attempts to minimize the prediction error variance by adjusting the innovation gain matrix and therefore has a superior model mismatch tolerance.

The major weakness of MIV is the lack of a rigorous stability proof as outlined in Appendix 2. Computational experience, however, indicated that the algorithm in most

cases remains stable without a stability check. In all other cases, the guidelines given in the Appendix are adequate to keep the algorithm stable. In our simulations, all but two cases remained stable without any stability check. All simulations remained stable when only stability on $[I - \hat{L}_k H_k^x]$ is checked, where stability has been violated and the new estimates had to be discarded only 0–4% times. The algorithm never became stagnant.

References

Ahmed, M. S. and N. Sait (1989). State-space adaptive control through a modified bootstrap algorithm for parameter and state estimation. *IEEE Proc. D, Control Theory and Appl.*, **136**, 215–224.

Astrom, A. J. and B. Wittenmark (1990). *Computer Controlled System: Theory and Design.* Prentice-Hall, Englewood Cliffs, NJ.

Dhingra, J. S., R. L. Moose, H. Vanlandingham and T. A. Lauzon (1992). A computationally efficient technique for state estimation of nonlinear systems. *Automatica*, **28**, 395–399.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory.* Academic Press, New York.

Kautsky, J., N. K. Nichols and P. Van Dooren (1985). Robust pole assignment in linear state feedback. *Int. J. Control*, **41**, 1129–1155.

Kuo, B. C. (1970). *Discrete Data Control Systems.* Prentice-Hall, Englewood Cliffs, NJ.

Ljung, L. and T. Soderström (1983). *Theory and Practice of Recursive Identification.* MIT, Cambridge, MA.

Moore, J. B. and H. Weiss (1979). Recursive prediction methods for adaptive estimation. *IEEE Trans. on Syst. Man and Cyb.*, **SMC-9**, 197–205.

Omani, F. K. and N. K. Sinha (1987). A modified state space approach to multivariable self-tuning control with pole assignment. *IEE Proc. D, Control Theory and Appl.*, **134**, 31–36.

Papoulis, A. (1991). *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, New York.

Soderström, T. and P. Stoica (1989). *System Identification.* Prentice-Hall, Englewood Cliffs, NJ.

Appendix

In this appendix the convergence and stability properties of MIV are analyzed.

1. *Algorithm convergence.* The parameters of the gain matrix L_k in equation (8) generally are time varying. In a RPE algorithm, this time varying nature is accommodated by proper choice of the variable λ. Nevertheless, it is of interest to investigate the convergence of the algorithm when the true parameter set assumed constant values. In the following we investigate the possible convergence points of the proposed algorithm assuming that the true parameter vector θ is constant.

Assuming that the sequences $\hat{e}_k(\hat{\theta})$ and $\Psi_k(\hat{\theta})$ are bounded, the convergence of the recursive algorithm presented in Section 3 can be analyzed by considering the associated differential equation (de) (Ljung and Söderström, 1983)

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\hat{\theta}_\tau = \mathfrak{f}(\hat{\theta}_\tau)\mathcal{R}_\tau^{-1} \tag{A.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathcal{R}_\tau = \mathcal{G}(\hat{\theta}_\tau) - \mathcal{R}_\tau \tag{A.2}$$

where

$$\mathfrak{f}(\hat{\theta}) = \mathcal{E}\{\hat{e}_k^T(\hat{\theta})\hat{\Sigma}_e^{-1}\Psi_k(\hat{\theta})\}$$
$$\mathcal{G}(\hat{\theta}) = \mathcal{E}\{\Psi_k^T(\hat{\theta})\hat{\Sigma}_e^{-1}\Psi_k(\hat{\theta})\}$$
$$\tau \approx \log k, \quad \mathcal{R}(k) = \gamma_k P_k^{-1}.$$

where γ_k is a positive quantity, $\hat{\Sigma}_e$ is the estimated value of Σ_e which will become approximately constant for large k and the subscript τ indicates the change of independent variable from k to τ. The possible limit points of $\hat{\theta}_k$ as k tends to infinity are the stable stationary points of the differential equations given by equations (A.1) and (A.2).

The convergence analysis of the proposed RPE algorithm directly parallels the analysis of the RPE algorithm in linear models reported in (Ljung and Söderström) 1983 and Söderström and Stoica (1989). It can be shown that the algorithm converges (locally and globally) to the stationary points of the asymptotic loss function

$$V_\infty(\hat{\theta}) = \mathcal{E}\{e_k^T(\hat{\theta})\hat{\Sigma}_e^{-1}e_k(\hat{\theta})\} \tag{A.3}$$

among which the true parameter is a candidate. If the true parameter set is the only stationary point, convergence to the true parameter set is ensured.

2. *Stability analysis.* The above analysis is based on the assumption that e_k and Ψ_k are bounded. This in turn requires that $\hat{x}_k$ and Λ_k remains bounded [see equations (14)–(17) and (25)–(27), respectively]. In the following, the condition under which they remain stable is investigated.

Combining equations (25)–(27), one obtains

$$\Lambda_{k+1} = C_k\Lambda_k + [I + F_k^x][L_k^\theta - \hat{L}_k H_k^\theta] + F_k^\theta \tag{A.4}$$

where

$$C_k = [I + F_k^x][I - \hat{L}_k H_k^x].$$

Similarly equations (14)–(17) may be combined to obtain

$$\hat{x}_{k+1} = x_k^{**} + f(g, x_k^{**} + \hat{L}_k y_k, u_k, t_k) + \hat{L}_k y_k \tag{A.5}$$

where $x_k^{**} = \hat{x} - \hat{L}_k h(\hat{x}_k, u_k, t_k, \hat{\theta}_k)\hat{x}_k$.
A linearized version of equation (A.5) can be obtained as

$$\Delta\hat{x}_{k+1} \cong C_k \Delta\hat{x}_k + F_k^u \Delta u_k + \hat{L}_k \Delta y_k, \tag{A.6}$$

where Δx, Δu and Δy are the incremental values. In order to gain insight into the stability property of equation (A.5), this linearized equation can be used. The analysis therefore remains valid for small values of the incremental variables.

The stability of equations (A.4) and (A.6) is ensured if the eigenvalues of (i) $[I + F_k^x]$ and (ii) $[I - \hat{L}_k H_k^x]$ lie within the unit circle at least infinitely often. When all the system parameters are known, the eigenvalue requirement of $[I + F_k^x]$ translates into the stability of the linearized system at the operating point. Since a system should be either open loop stable or stabilized in closed loop, the above stability requirement is always met.

The stability of the algorithm therefore depends on the eigenvalues of $[I - \hat{L}_k H_k^x]$. Since the matrices L_k and H_k^x are available in each iteration Jury stability criteria (Aström and Wittenmark, 1990) can be incorporated to detect the instability. The stability of the algorithm can be maintained in three different ways.

(i) Discard the new parameter estimates and keep the old ones when the stability condition is violated.

(ii) The gain M_k in equation (30) is iteratively halved until the stability condition is satisfied (Ljung and Söderström, 1983). This approach, called the interval halving, however, may increase the computational requirements considerably. Further, both of the above approaches may make the algorithm stagnant as mentioned in Ahmed and Sait (1989).

(iii) Recompute L_k to obtain an equivalent minimum phase description. Whenever one or more eigenvalues of $I - \hat{L}_k H_k^x$ move outside the unit circle, they are substituted by their image values inside the unit circle. Then a new value of $\hat{L}_k$ is computed such that $I - \hat{L}_k H_k^x$ has the modified eigenvalues as obtained above. This is a typical problem of closed loop pole placement in control theory. Many techniques can be adopted (Kautsky et al., 1985). This approach is capable of keeping the algorithm stable without making it stagnant.

The latter approach, however, is only applicable when all system parameters are known. If some of the system parameters are unknown, the eigenvalues of $[I + F_k^x]$ may also contribute to the instability, and a combination of (i) and (iii) may be attempted. Unlike the linear case, F_k^x depends on $\hat{x}_k^*$. Consequently, a particular $\hat{\theta}$ although satisfying the stability requirement for $[I + F_k^x]$ in one iteration, may violate the stability requirement in the subsequent iterations. This poses an open problem, which is the development of a suitable method to ensure stability of the algorithm in such cases.