Stability of nonlinear polynomial ARMA models and their inverse

EVELIO HERNÁNDEZ† and Y. ARKUN†

Stability and invertibility of discrete time Polynomial ARMA models are studied in an 'extended neighbourhood' to extend the local results obtained earlier. The analysis combines linear robust control and classical nonlinear system theories. Polynomial ARMA models are used as examples illustrating the theory.

1. Introduction

The main concern of this paper is the analysis of a nonlinear input–output model in terms of its stability and invertibility properties. Proper analysis of input/output models can reveal valuable insight into process dynamics and yield key information about model-based control schemes. Local results have recently been derived by Hernández and Arkun (1993); the thrust of this paper is to extend these local results to a region specified around a particular equilibrium point. This way, more than just a local neighbourhood of unknown size is studied and the nonlinear behaviour of the system is better captured.

This 'extended neighbourhood' analysis is achieved through the construction of conic sectors on the nonlinear input/output map and checking stability by using linear robust control theory. The bounds calculated for the open loop input/output map can also be used for invertibility analysis. Thus, with this framework, one can study the stability of the model's inverse without having to calculate sector bounds for the inverse input/output map. Results on the stability of the inverse are useful for analysing stability of the closed loop system when the model inverse is used as the controller. Stability of more practical extended horizon type controllers, which use multiple step ahead predictions can also be studied using the same sector bounds.

The paper is structured as follows: §2 formulates the problem, introduces the notation and gives a short overview of relevant results that have appeared in the literature. In addition, results from linear robust control theory adopted in this work are summarized; §3 uses conic sector bounds to represent the nonlinear systems as uncertain linear systems. Based on this representation and utilizing known results from linear robust control theory, 'extended neighbourhood' stability conditions for nonlinear input/output systems are derived. The new results are contrasted with others in the literature. Finally, the study is completed with the stability of the model's inverse in an 'extended neighbourhood'. Examples are given throughout the paper.

Received 12 April 1993. Revised 9 March 1994 and 24 September 1994. Communicated by Professor M. Morari.

† School of Chemical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0100, U.S.A.

2. Notation and background

There exist numerous model structures proposed for the identification of nonlinear systems. For a review the reader is referred to Haber and Unbehauen (1990). The model selected for this study has a polynomial ARMA structure as discussed for example by Korenberg *et al.* (1988). Assume that $u(k-\rho-1)$ is the first input to affect the current output $y(k)$. Then we could use a polynomial model to express the next output, $y(k+1)$, as a function of current and previous outputs and inputs prior to and including time $k-\rho$

$$\begin{aligned} y(k+1) = {} & \theta^0 + \sum_{i=0}^{n_y} \theta_i^1 y(k-i) + \sum_{i=\rho}^{n_u} \theta_i^2 u(k-i) \\ & + \sum_{i=0}^{n_y} \sum_{j=0}^{i} \theta_{(i,j)}^3 y(k-i)\, y(k-j) \\ & + \sum_{i=\rho}^{n_u} \sum_{j=\rho}^{i} \theta_{(i,j)}^4 u(k-i)\, u(k-j) \\ & + \sum_{i=0}^{n_y} \sum_{j=\rho}^{n_u} \theta_{(i,j)}^5 y(k-i)\, u(k-j) + \cdots \end{aligned} \tag{1}$$

or simply

$$y(k+1) = f(y(k), \ldots, y(k-n_y), u(k-\rho), \ldots, u(k-n_u)) \tag{2}$$

If the system to be modelled is multivariable, then (1) can be extended in a straightforward fashion by including, on the right-hand side, the appropriate monomials in the additional inputs and outputs. In the remainder of the paper only SISO systems will be considered without any loss of generality. The polynomial ARMA models have been used by several authors for modelling and control of nonlinear systems (e.g. Korenberg *et al.* 1988, Hernández 1992, Hernández and Arkun 1993). Theoretical justifications for this structure are also available (e.g. see Díaz and Descrochers 1988).

For the stability and invertibility analysis it is convenient to construct a state-space realization of (1). Consider first the following definition of the state

$$\begin{aligned} x_1^a(k) &= y(k) & x_1^b(k) &= u(k-\rho-1) \\ &\vdots & &\vdots \\ x_{n_y+1}^a(k) &= y(k-n_y) & x_{n_u-\rho}^b(k) &= u(k-n_u) \\ \underline{x}^a(k) &= [x_1^a(k) \ldots x_{n_y+1}^a(k)]^T & \underline{x}^b(k) &= [x_1^b(k) \ldots x_{n_u-\rho}^b(k)]^T \end{aligned}$$

Note that the state has been clearly divided into the 'output' and 'input' parts: $\underline{x}^a(k)$ and $\underline{x}^b(k)$, respectively. The complete state is then given by

$$\underline{x}(k) = \begin{bmatrix} \underline{x}^a(k) \\ \underline{x}^b(k) \end{bmatrix} \tag{3}$$

Given this definition of state, the state-space realization of (1) is given by

$$\begin{bmatrix} \underline{x}^a(k+1) \\ \underline{x}^b(k+1) \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & 0 & & & & \\ 1 & & & 0 & & & & \\ & \ddots & & \vdots & & & & \\ & & 1 & 0 & & & & \\ & & & & 0 & \cdots & 0 & 0 \\ & & & & 1 & & & 0 \\ & & & & & \ddots & & \vdots \\ & & & & & & 1 & 0 \end{bmatrix} \begin{bmatrix} \underline{x}^a(k) \\ \underline{x}^b(k) \end{bmatrix} + \begin{bmatrix} f(\underline{x}(k), u(k-\rho)) \\ 0 \\ \vdots \\ 0 \\ u(k-\rho) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$y(k) = [1 \quad 0 \quad \cdots \quad 0 \,|\, 0 \quad \cdots \quad 0] \begin{bmatrix} \underline{x}^a(k) \\ \underline{x}^b(k) \end{bmatrix} \tag{4}$$

or

$$\underline{x}(k+1) = \mathbf{F}(\underline{x}(k), u(k-\rho))$$
$$y(k) = h(\underline{x}(k))$$

Note that in defining the above system, the notation has been stretched and allowed for

$$\begin{aligned} y(k+1) &= f(y(k), \ldots, y(k-n_y), u(k-\rho), \ldots, u(k-n_u)) \\ &= f(\underline{x}(k), u(k-\rho)) \end{aligned}$$

Note that the realization used is not necessarily minimal, and introducing more states than minimally necessary adds dynamics that may not be stable or whose inverse may not be stable. This would lead to labelling a stable input–output map as unstable where, if another realization was used, it would be correctly identified as stable. The main reason for using this realization is that the state variables are physical signals that can be measured. Furthermore, its construction is intuitive and is free of the numerical issues associated with other realizations (Constanza *et al.* 1983).

Given the above state-space representation, conditions for local stability and invertibility have been derived based on the linearization of (4) around the equilibrium point of interest (Hernández and Arkun 1993). The purpose of this paper is to go beyond the local results and study stability and invertibility in an extended neighbourhood around the equilibrium point. First, we give an overview of the available result in this context.

2.1. *Results for models with special structures*

The first class of systems to be considered are block-oriented models such as the Hammerstein or Wiener models that are formed by a static nonlinear map, which either precedes or follows a linear dynamical system. Assuming that the nonlinear map is bounded, the open loop stability of block oriented models is determined solely by the linear dynamical system. Thus, the open loop system is globally stable if the poles of the linear dynamical part are inside the unit circle. Similarly, assuming that the inverse of the nonlinear static map exists and is bounded, the open loop stability properties of the inverse system are completely determined by the linear dynamical part. Therefore, if the zeros of the linear system are inside the unit circle, the inverse system is globally stable.

Narendra and Parthasarathy (1990) provided straightforward global stability

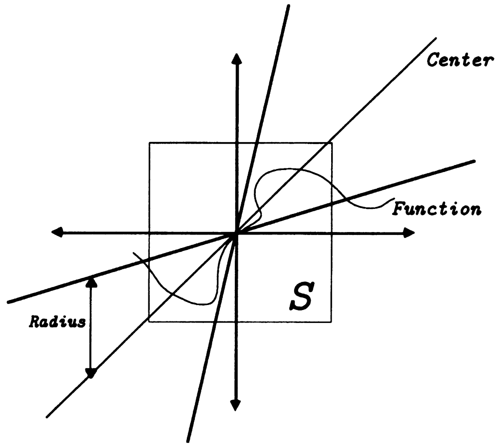

Figure 1. Conic sector bounds.

conditions for systems that are linear in the outputs, and global invertibility conditions for systems linear in the inputs.

Bilinear systems have also been extensively studied. Priestley (1988) gives conditions for global stability. Explicit conditions for global invertibility of bilinear systems are not available at this time.

2.2. *Analysis of general input–output models*

Stability of nonlinear systems has traditionally been studied via Lyapunov methods (see for example Vidyasagar 1978). However, constructing Lyapunov functions for general nonlinear systems could be very conservative. Usually, some more structural information is needed in order to construct useful Lyapunov functions.

Only a few approaches have been suggested for the stability study of general nonlinear models. One of these approaches is to determine whether the system is a contraction mapping. This approach was taken by Chen and Billings (1989). However, the conditions provided in that paper could never by satisfied due to the nature of the norms and the realization used. A more detailed discussion of this will be given later.

Another general approach is to bound the nonlinear system with conic sectors and use several analysis tools on the bounded system. For example, Zames (1966) used the Circle Theorem, and Molchanov (1987) used Lyapunov's second method to analyse the stability of nonlinear systems bounded by cones. His method is constructive; that is, the explicit computation of the Lyapunov functions is given. Another way to use conic bounds is to reformulate the nonlinear system as an uncertain linear one. Then, the uncertain linear system can be studied by using results from linear robust control theory. This approach has been used by, for example, Kammash and Pearson (1990) and Doyle *et al.* (1989). A problem with this approach is that the analysis only holds in the region where the conic approximation is valid. Thus, the user must guarantee that either the conic approximation is globally valid or that the system cannot leave the region where the conic approximation is valid (as done, for example, by Doyle and Morari 1990). This type of approach will be adopted and improved in this paper.

The stability of the inverse model has already been addressed, at least in principle. If the analytical inverse is constructed, then its stability properties can be studied using one of the methods described above. However, the pitfall of this is that the analytical calculation of the inverse is a cumbersome process and is even impossible for a variety of systems. In those situations, numerical calculations of the inverse may be possible and still useful for control purposes. Also, for these control purposes, it is important to know whether the inverse system (even though no analytical solution may exist) is stable. One of the main contributions of this work is the analysis of the stability of the model's inverse in an extended neighbourhood without having to calculate this inverse analytically.

2.3. *Relevant aspects of linear robust control theory*

This section reviews several standard results from robust control theory that will be used. These are the stability conditions for uncertain, time-varying, linear, difference equations. Consider the linear difference equations

$$\begin{bmatrix} \zeta(k) \\ \hline \underline{x}(k+1) \\ y(k) \end{bmatrix} = \underbrace{\left[\begin{array}{c|cc} E & F & G \\ \hline H & A & B \\ J & C & D \end{array}\right]}_{M^*} \begin{bmatrix} \omega(k) \\ \hline \underline{x}(k) \\ u(k) \end{bmatrix} \tag{5}$$

where $x(k) \in \mathbf{R}^n$, $y(k) \in \mathbf{R}^p$ and $u(k) \in \mathbf{R}^q$ are the state, output and input of a dynamical system at time k. Variables $\zeta(k), \omega(k) \in \mathbf{C}^l$ denote the internal variables used to describe the effects of uncertainty. Now let $\omega(k) = \Delta_1(k)\,\zeta(k)$, with time-varying uncertainty $\Delta_1 \in \mathbf{C}^{l\times l}$, and eliminate the $\zeta(k)$ and $\omega(k)$ variables from (5). The resulting relationship becomes

$$\begin{aligned} \begin{bmatrix} x(k+1) \\ y(k) \end{bmatrix} &= F_u(M^*, \Delta_1) \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \\ &= M^*_{22} + M^*_{21}\,\Delta_1 (I - M^*_{11}\,\Delta_1)^{-1}\, M^*_{12} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \\ &= \begin{bmatrix} A + H\Delta_1(I-E\Delta_1)^{-1}F & B + H\Delta_1(I-E\Delta_1)^{-1}G \\ C + J\Delta_1(I-E\Delta_1)^{-1}F & D + J\Delta_1(I-E\Delta_1)^{-1}G \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \end{aligned} \tag{6}$$

where the upper linear fractional transformation $F_u(M^*, \Delta_1)$ has been implicitly defined. Using the induced 2-norm (the maximum singular value), a stability condition for time varying perturbations reads (Packard 1988)

> *For bounded inputs, $u(k)$, and time varying perturbations, the system in (5) is stable if*
>
> $$\bar{\sigma}(A + H\Delta_1(I-E\Delta_1)^{-1}F) < 1$$
>
> *for all $\Delta_1(k) \in \mathbf{C}^{l\times l}$*

It is true that the structured singular value μ of a matrix with respect to a completely unstructured perturbation set $\Delta_{us} = \{\Delta \mid \Delta \in \mathbf{C}^{n\times n}, \bar{\sigma}(\Delta) \leqslant 1\}$ is equal to the maximum singular value of the matrix. Thus, the singular value stability condition can be replaced by: $\mu_{\Delta_{us}}(A + H\Delta_1(I-E\Delta_1)^{-1}F) < 1$. This condition can be shown to be equivalent to (Packard 1988)

$$\mu_{\tilde{\Delta}}(\tilde{M}_{11}) < 1 \tag{7}$$

where $\breve{\Delta}$ is defined as:

$$\breve{\Delta} = \{\Delta \mid \Delta = \text{diag}(\Delta_1, \Delta_2), \quad \Delta_1 \in \mathbf{C}^{l \times l}, \quad \bar{\sigma}(\Delta_1) \leqslant 1, \quad \Delta_2 \in \mathbf{C}^{n \times n}, \quad \bar{\sigma}(\Delta_2) \leqslant 1\} \tag{8}$$

and

$$\tilde{M}_{11} = \begin{bmatrix} E & F \\ H & A \end{bmatrix}$$

The above result can be too conservative. In fact, given the realization of the input–ouput map used, the condition will always fail, as explained later. An alternative would be to find another realization by defining a different state. Let T be the non-singular transformation between the current state $x(k)$ and the new state $\tilde{x}(k)$. This transformation has the effect of replacing A with TAT^{-1}, H and B with TH and TB and F and C with FT^{-1} and CT^{-1}. Furthermore, note that a coordinate transformation cannot change the stability properties of the system. Thus, a new, less conservative stability condition can be obtained by replacing the μ-condition in (7) with

$$\min_{T \in C^{n \times n}} \mu_{\breve{\Delta}}\left(\begin{pmatrix} I_l & 0 \\ 0 & T \end{pmatrix} \tilde{M}_{11} \begin{pmatrix} I_l & 0 \\ 0 & T^{-1} \end{pmatrix}\right) < 1 \tag{9}$$

Note that condition (9) is not equivalent to $\mu_{\breve{\Delta}}(\tilde{M}_{11}) < 1$ because, in general, T does not commute with Δ_2. A major problem associated with (9) is that it is unclear how to search for the transformation T. Consider instead the upper bound of (9)

$$\min_{d_1, d_2 \in C} \min_{T \in C^{n \times n}} \bar{\sigma}\left(\begin{pmatrix} d_1 I_l & 0 \\ 0 & d_2 I_n \end{pmatrix}\begin{pmatrix} I_l & 0 \\ 0 & T \end{pmatrix} \tilde{M}_{11} \begin{pmatrix} I_l & 0 \\ 0 & T^{-1} \end{pmatrix}\begin{pmatrix} d_1^{-1} I_l & 0 \\ 0 & d_2^{-1} I_n \end{pmatrix}\right) < 1 \tag{10}$$

Finally, note that the transformation T has the effect of changing the set of D over which the optimization is carried. Instead of searching over

$$\tilde{D} = \{D \mid \text{diag}(d_1 I_l, d_2 I_n), \quad d_i \in \mathbf{C}\}$$

the search is over the larger set

$$\breve{D} = \{D \mid \text{diag}(d_1 I_l, D_2), \quad d_1 \in \mathbf{C}, \quad D_2 \in \mathbf{C}^{n \times n}\} \tag{11}$$

and so the conservatism is reduced. Thus, the condition to be used later is

$$\operatorname*{Min}_{D \in \breve{D}} \bar{\sigma}(D \tilde{M}_{11} D^{-1}) < 1 \tag{12}$$

where $\breve{D}$ is defined in (11). Minimization (12) can be formulated as a convex optimization problem and a global solution can be obtained.

3. Extended neighbourhood analysis

The 'extended neighbourhood' analysis of nonlinear input–output models begins by using conic sector bounds to recast the system in (4) into a simpler system. This idea has been suggested by Doyle *et al.* (1989) in connection with the control of chemical reactors, and also by Kammash and Pearson (1990) and Molchanov (1987). The simpler system could then be analysed using the theory mentioned above. The idea of conic sector bounds is illustrated in Fig. 1. The nonlinear, real valued map $f(\underline{x})$, ($x \in \mathbf{R}^N$) is bounded in some region S by the cone with centre L and radius R: $C(L, R)$. Mathematically speaking a cone $C(L, R)$ bounds a function $f(\underline{x})$ over the region S if

$$\|f(\underline{x}) - L\underline{x}\| \leqslant \|R\underline{x}\| \quad \forall \underline{x} \in S \tag{13}$$

It is critical to point out that the analysis to follow only focuses in the region of interest (S). The nonlinear function does not need to be bounded by the cone outside this region.

The system in (4) can be divided into a linear part and an 'uncertain' part through the use of conic sectors. For this, note that the only nonlinearity in system (4) appears on the function $f(\underline{x}, u)$, (where the time arguments on $\underline{x}$ and u have been suppressed) which is a scalar for SISO systems. Let a cone with centre $[A^0\, b^0]$ and radius $[R_x\, r_b]$ bound the function $f(\underline{x}, u)$, then

$$\left\| f(\underline{x}, u) - [A^0\, b^0] \begin{bmatrix} x \\ u \end{bmatrix} \right\| \leqslant \left\| [R_x\, r_b] \begin{bmatrix} x \\ u \end{bmatrix} \right\| \quad \forall \begin{bmatrix} x \\ u \end{bmatrix} \in S \tag{14}$$

Note that, since $f(\underline{x}, u)$ is a scalar, for any $\underline{x}$ and u in S there is always a scalar $\Delta(x, u) \in [-1, 1]$ such that

$$f(\underline{x}, u) - [A^0\, b^0] \begin{bmatrix} x \\ u \end{bmatrix} = \Delta(\underline{x}, u)\, [R_x\, r_b] \begin{bmatrix} x \\ u \end{bmatrix}$$

or

$$f(\underline{x}, u) = \underbrace{[A^0\, b^0] \begin{bmatrix} x \\ u \end{bmatrix}}_{\text{Linear part}} + \underbrace{\Delta(\underline{x}, u)\, [R_x\, r_b] \begin{bmatrix} x \\ u \end{bmatrix}}_{\text{'Uncertain' part}} \tag{15}$$

Note that $\Delta(\underline{x}, u)$ is, in general, a nonlinear operator. In order to use the theory developed by Packard (1988), a linear time-varying operator Δ_k is needed. However, the conditions imposed by Packard on a scalar, time-varying operator are that first, it belongs to the set of complex numbers, and second that its magnitude is bounded above by 1. Note that since the nonlinear operator $\Delta(\underline{x}, u)$ is real and restricted to $\{-1, 1\}$, there always exists a sequence $\{\Delta_k | k = 1, \ldots; |\Delta_k| < 1\}$ such that $\Delta_i = \Delta(\underline{x}, u)$, regardless of the trajectory taken by the system. Thus, from now on, we will use a linear time-varying operator Δ_k to represent the nonlinear one $\Delta(\underline{x}, u)$. The time-varying linear system and the nonlinear system are equivalent in the sense that their states follow identical trajectories in response to external signals and initial conditions. In essence, what follows is a study of the stability and invertibility of the time varying system with a state trajectory equivalent to the nonlinear system of interest. If stability is shown for the time varying system, then it is also shown for any system with identical state trajectories to the time varying one.

Using the linear plus 'uncertain' representation in (15) the system in (4) can be rewritten as:

$$
\underline{x}(k+1) = \underbrace{\begin{bmatrix}
\alpha_1^0 & \alpha_2^0 & \cdots & \alpha_{n_y+1}^0 & \alpha_{n_y+2}^0 & \alpha_{n_y+3}^0 & \cdots & \alpha_{N+1}^0 \\
1 & & & 0 & 0 & & x & 0 \\
& & \ddots & \vdots & & \ddots & & \vdots \\
& & 1 & 0 & & & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & & & 0 & 1 & & & 0 \\
& \ddots & & \vdots & & \ddots & & \vdots \\
& & 0 & 0 & & & 1 & 0
\end{bmatrix}}_{\mathscr{A}} \underline{x}(k)
$$

$$
+ \underbrace{\begin{bmatrix} b^0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{b} u(k-\rho) + \begin{bmatrix} \omega(k) \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{16}
$$

where α_i^0 is the ith entry of the A^0 vector, $N = n_y + n_u - \rho$, and $\omega(k)$ accounts for the uncertainty in the system and is given by

$$
\begin{aligned}
\zeta(k) &= R_x \underline{x}(k) + r_b u(k-\rho) \\
\omega(k) &= \Delta_k \zeta
\end{aligned} \tag{17}
$$

Collecting all the information gathered about the system and the uncertainty yields the set of equations in the form of (5)

$$
\begin{aligned}
\begin{bmatrix} \zeta(k) \\ \underline{x}(k+1) \\ y(k+1) \end{bmatrix} &= \underbrace{\begin{bmatrix} 0 & R_x & r_b \\ e_1 & A & b \\ Ce_1 & CA & Cb \end{bmatrix}}_{M} \begin{matrix} \omega(k) \\ \underline{x}(k) \\ u(k-\rho) \end{matrix} \\
\omega(k) &= \Delta_k \zeta(k) \\
\underline{x}(k) &= z^{-1} \underline{x}(k+1)
\end{aligned} \tag{18}
$$

where $e_1 = [1 \quad 0 \quad \cdots \quad 0]^T$.

In cases where one cone is not possible or practical, more than one cone may be constructed in order to represent the system as a linear plus uncertain system. Then it would be necessary to find $R_{1,x}, R_{2,x}, \ldots, R_{s,x}$ such that

$$\left\| f(\underline{x}, u) - [A^0\ b^0]\begin{bmatrix} \underline{x} \\ u \end{bmatrix} \right\| \leqslant \left\| [R_{1,x}\, r_b]\begin{bmatrix} \underline{x} \\ u \end{bmatrix} \right\| + \|R_{2,x}\underline{x}\| + \ldots \|R_{s,x}\underline{x}\|$$

and there exists s scalars: $\Delta_i(\underline{x}, u)$, $i = 1, \ldots, s$ such that

$$f(\underline{x}, u) = [A^0\, b^0]\begin{bmatrix} \underline{x} \\ u \end{bmatrix} + \Delta_1(\underline{x}, u)\,[R_{1,x}\, r_b]\begin{bmatrix} \underline{x} \\ u \end{bmatrix} + \Delta_2(\underline{x}, u)\, R_{2,x}\underline{x} + \ldots + \Delta_s(\underline{x}, u)\, R_{s,x}\underline{x} \quad (19)$$

The result of considering more cones is to augment the dimension of the uncertainties vector and so the original system (4) now becomes

$$\begin{bmatrix} \zeta_1(k) \\ \zeta_2(k) \\ \vdots \\ \zeta_s(k) \\ \underline{x}(k+1) \\ y(k+1) \end{bmatrix} = \begin{bmatrix} \underline{0} & R_{1,x} & r_b \\ \underline{0} & R_{2,x} & 0 \\ \vdots & \vdots & \vdots \\ \underline{0} & R_{s,x} & 0 \\ e_{\underline{1}} & \mathscr{A} & b \\ Ce_{\underline{1}} & \mathscr{C}\mathscr{A} & Cb \end{bmatrix} \begin{bmatrix} \omega_1(k) \\ \omega_2(k) \\ \vdots \\ \omega_s(k) \\ \underline{x}(k) \\ u(k-\rho) \end{bmatrix} \quad (20)$$

$$\omega_i(k) = \Delta_{i,k}\,\zeta_i(k)$$

where $\underline{0}$ is a row vector with s zeros and $e_{\underline{1}}$ is an $n \times s$ matrix where the first row is a vector of ones and the rest of the matrix contains zeros. System (20) is in the form of (5) for which stability conditions were given.

3.1. '*Extended neighbourhood*' *stability*

We have, so far laid the framework and the results needed to characterize the 'extended neighbourhood' stability of input–output models. In particular, the nonlinear system was reformulated as a linear plus (time-varying) uncertain system. Conditions for stability of such systems were given in §2. Therefore, all that is left to do is to state formally the conditions for the 'extended neighbourhood' stability of input–output models. This is done now in the form of a theorem.

Theorem 3.1: *Let the system in* (4) *be bounded in a region S by conic sectors with centre* $[A^0\, b^0]$ *and radii* $R_{1,x}, \ldots, R_{s,x}$, *and* r_b. *Using these bounds, let the system be represented by* (20). *Then, assuming that the system does not leave S, its equilibria in S will be asymptotically stable provided that*

$$\min_{D \in \check{D}} \bar{\sigma}(DM_{11}D^{-1}) < 1 \quad (21)$$

where this time M_{11} *is given by*

$$M_{11} = \begin{bmatrix} \underline{0} & R_{1,x} \\ \vdots & \vdots \\ \underline{0} & R_{s,x} \\ e_{\underline{1}} & \mathscr{A} \end{bmatrix}$$

$\mathscr{A}$ *is defined in* (16) *and* $\check{D}$ *is defined in* (11)

Remarks

(*a*) An excellent discussion on the computation of conic sector bounds can be found in Doyle and Morari (1990)

(*b*) There is nothing in the theorem that limits the results to SISO systems. In fact, if conic bounds are obtained for MIMO systems, then the result is also valid for multivariable systems, this is shown in detail in Hernández (1992).

(*c*) The framework being used for analysis is conservative in the sense that it will provide conditions for stability of more systems than just (4). In fact, any one of the systems that is bounded by the cone in (20) is analysed. This is important since we know that the model will only be an approximation of the actual plant.

(*d*) Other conditions for stability can be derived given that conic sector bounds are available. Notably, Molchanov (1987) derives Lyapunov functions for nonlinear systems bounded by sectors. However, this approach is not pursued any further because it cannot be used for the study of the stability of the inverse of the input–output map.

(*e*) It must be stressed that Theorem 3.1 assumes that the system does not leave the region S. The reason being that the conic bounds are only valid in S. If the system leaves this region, then the representation of (20) is not valid. The next section addresses this issue by providing restrictions on the construction of S and by bounding the jacobian of the $\mathsf{F}(\underline{x}, u)$ map instead of the function itself.

3.2. *Improvement to the result of Chen and Billings (1989)*

By using the ideas presented so far, it is possible to strengthen a result obtained by Chen and Billings (1989). Interestingly, this will, in turn, strengthen the general result given in Theorem 3.1. Since the result of Chen and Billings is based on the concept of a 'contraction' and the Contraction Mapping Theorem, these ideas will begin the discussion in this section.

Definition 3.1: A mapping $\mathsf{F}(\underline{x}, u^*)$ is said to be *Lipschitz continuous* in S if there exists a constant $\mathscr{L}$ such that for all $x', x'' \in S$

$$\|\mathsf{F}(\underline{x}', u^*) - \mathsf{F}(\underline{x}'', u^*)\| \leqslant \mathscr{L}\,\|x' - \underline{x}''\| \tag{22}$$

Note that the input, u^*, is assumed constant for now.

Definition 3.2: A mapping $\mathsf{F}(\underline{x}, u^*)$ is said to be a *contraction mapping* in S if: (i) $\mathsf{F}(\underline{x}, u^*)$ maps S into itself; and (ii) $\mathscr{L}$ in (22) is less than 1.

Theorem 3.2—Contraction mapping theorem: *If a mapping, $\mathsf{F}(\underline{x}, u^*)$ is a contraction in S then*

(i) $\mathsf{F}(\underline{x}, u^*)$ *has an equilibrium point* $(\underline{x}^*, u^*)$ *in S.*

(ii) *The equilibrium point is unique in S.*

(iii) *The equilibrium can be reached by successive iterations of the* F *mapping. That is, the system is asymptotically stable.*

The reader should note that one of the defining points of a contraction in a set S is that the mapping must map S into itself. This is equivalent to the requirement in the general Theorem 3.1 that the system must not leave the region S. Unfortunately, it is difficult to confirm this condition *a priori* and thus it is necessary to strengthen the general result of Theorem 3.1. An alternative form of Theorem 3.2 is obtained if the region S is known *a priori* to contain an equilibrium point. This will likely be the situation in the type of analysis sought because S would usually be an 'extended neighbourhood' of an equilibrium point. If this is the case, then the following result can be stated (Economou 1985).

Theorem 3.3: *If a mapping $\mathbf{F}(\underline{x}, u^*)$ is Lipschitz continuous in S, the Lipschitz constant is less than 1 and S is known to contain an equilibrium point, $(\underline{x}^*, u^*)$ of $\mathbf{F}(\underline{x}, u^*)$ then:*

(*a*) *the equilibrium point is unique in S.*

(*a*) *the equilibrium can be reached by successive iterations of the F mapping. That is, the system is asymptotically stable.*

Finally, if the input is not fixed, then the results need only slight modification. First of all, the region S would include the variation in the input. Furthermore, for every possible value of the input u, it is necessary to ensure that S contains an equilibrium point for this value of u. This is needed so that Theorem 3.3 can be applied to every different value of u considered.

From the definitions and the theorems displayed above, it is clear that the computation of $\mathscr{L}$ is of major importance. One way to compute this quantity is through the mean value theorem, which states the following

Mean value theorem: *Let $S \subseteq \mathbf{R}^p$ be an open set and let $\mathbf{F}: S \to \mathbf{R}^q$. Suppose that S contains the points a, b and the line segment joining these points, and that $\mathbf{F}$ is differentiable at every point on this segment. Then there exists a point on the segment, c, such that*

$$\|\mathbf{F}(b) - \mathbf{F}(a)\| = \|\nabla \mathbf{F}(c)(b-a)\| \tag{23}$$

where $\nabla \mathbf{F}(c)$ denotes the jacobian of $\mathbf{F}$ evaluated at the point c

It follows from (23) that

$$\|\mathbf{F}(b) - \mathbf{F}(a)\| \leqslant \|\nabla \mathbf{F}(c)\| \cdot \|b-a\|$$

Thus, one $\mathscr{L}$ can be found through the following optimization

$$\mathscr{L} = \max_{\{\underline{x}, u\} \in S} \|\nabla \mathbf{F}(\underline{x}, u)\| \tag{24}$$

Cheng and Billings use the contraction mapping ideas presented above to study the stability of the general input–output map.

$$y(k+1) = f(y(k), \ldots, y(k-n_y), u(k-\rho), \ldots, u(k-n_u))$$

through the state-space realization in (4). They obtain an upper bound of $\mathscr{L}$ that is simpler to compute by 'breaking up' the optimization in (24) into smaller problems. Specifically, they define

$$\tilde{\alpha}_i = \max_{\{\underline{x}, u\} \in S} \left| \frac{\partial f(\underline{x}, u)}{\partial x_i^a} \right| \quad i = 1, \ldots, n_y + 1$$

$$\tilde{\beta}_j = \max_{\{\underline{x}, u\} \in S} \left| \frac{\partial f(\underline{x}, u)}{\partial x_j^b} \right| \quad j = 1, \ldots, n_u$$

Then, $\|\tilde{J}\|$ can be shown to be an upper bound of $\mathscr{L}$ in (24), where $\tilde{J}$ is defined as

$$\tilde{J} = \begin{bmatrix} \tilde{\alpha}_1 & \cdots & \tilde{\alpha}_{n_y} & \tilde{\alpha}_{n_y+1} & \tilde{\beta}_1 & \cdots & \tilde{\beta}_{n_u-1} & \tilde{\beta}_{n_u} \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ & \ddots & & \vdots & \vdots & & & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & & \cdots & 0 & 0 & & \cdots & 0 \\ 0 & & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \vdots & & \ddots & & \vdots \\ 0 & & \cdots & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Since $\|\tilde{J}\|$ is an upper bound to $\mathscr{L}$, then if $\|\tilde{J}\|$ is less than 1, $\mathscr{L}$ is less than one and from Theorem 3.3, the system is asymptotically stable in S. The particular norm used by Chen and Billings was the induced 2-norm, or the maximum singular value of $\tilde{J}$. This is precisely the problem with their result. Recall that the maximum singular value of a companion matrix such as $\tilde{J}$ is always greater than or equal to 1. To see this, first note that

$$\bar{\sigma}(A\tilde{J}) \leqslant \bar{\sigma}(A)\,\bar{\sigma}(\tilde{J}) \tag{25}$$

Now let

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

and $\tilde{J}$ as defined above. Then note that

$$A\tilde{J} = \begin{bmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \vdots \\ & & 1 & 0 \end{bmatrix}$$

Finally, observe that $\bar{\sigma}(A) = \bar{\sigma}(A\tilde{J}) = 1$. Substituting this back to (25), the following is obtained

$$\bar{\sigma}(\tilde{J}) \geqslant 1$$

So the conditions given by Chen and Billings can never be met. Furthermore, other simple-to-compute matrix norms such as the induced 1 and ∞ norms (maximum column and row sums, respectively) are also guaranteed to be greater than or equal to 1 due to the 1s in $\tilde{J}$.

One way to correct the result would be to use an invertible state transformation, T to define a new state $\underline{x}'$

$$\underline{x}' = T\underline{x}$$

which, in turn, defines a new system

$$\underline{x}'(k+1) = T\underline{x}(k+1) = T\mathbf{F}(T^{-1}\underline{x}'(k), u(k-\rho)) = \mathbf{F}'(\underline{x}'(k), u(k-\rho)) \tag{26}$$

It is crucial to stress that this invertible transformation cannot change the stability properties of a system. Thus, if the system in (26) is stable, so is the one in (4). A Lipschitz constant $\mathscr{L}$ for the new system can be computed as before through (24). The jacobian of $\mathbf{F}'(\cdot,\cdot)$ with respect to $\underline{x}'(k)$ can be calculated through the chain rule as

$$\nabla_{\underline{x}'}\mathbf{F}'(\underline{x}', u) = T\nabla_{\underline{x}}\mathbf{F}(\underline{x}, u)\,T^{-1}$$

The goal is now to find a coordinate transformation such that

$$\max_{\{\underline{x},u\}\in S} \min_T \bar{\sigma}(T\nabla_{\underline{x}} \mathbf{F}(\underline{x},u)\, T^{-1}) < 1 \tag{27}$$

The nonlinear program posed in (27) could be difficult to solve. However, an upper bound of this quantity can be cast into a convex optimization problem through the use of the linear robust control theory presented in §2. The plan undertaken is as follows. First recast the matrix $\nabla_{\underline{x}} \mathbf{F}(\underline{x},u)$ for $\{\underline{x},y\}\in S$ with a constant, nominal matrix plus an uncertainty. This representation will be then used to calculate an upper bound to (27), which can be computed through a convex optimization problem. If this upper bound is less than one, it will then follow that the nonlinear system is stable.

First, observe that the partial derivatives of $f(\underline{x},u)$ with respect to each of the state variables must be bounded in S in order for the optimal value of (27) to be bounded. If this is the case, then there exists 'means': α_i^0, β_j^0 and 'ranges': α_i^R, β_j^R; $i=1,\dots,n_y+1$; $j=1,\dots,n_u$ such that for any $(\underline{x},u)\in S$

$$\frac{\partial f(\underline{x},u)}{\partial x_i^a} = \frac{\partial x_1^a(k+1)}{\partial x_i^a(k)} = \alpha_i^0 + \delta_i\,\alpha_i^R \quad \text{and} \quad \frac{\partial f(\underline{x},u)}{\partial x_j^b} = \frac{\partial x_1^a(k+1)}{\partial x_j^b(k)} = \beta_j^0 + \delta_{n_y+1+j}\,\beta_j^R$$

for some $\delta_i \in [-1,1]$; $i=1,\dots,N$. Then define F_x^δ as

$$F_x^\delta \equiv \begin{bmatrix} \alpha_1^0+\delta_1\alpha_1^R & \cdots & \alpha_{n_y+1}^0+\delta_{n_y+1}\alpha_{n_y+1}^R & \beta_1^0+\delta_{n_y+2}\beta_1^R & \cdots & \beta_{n_u}^0+\delta_N\beta_{n_u}^R \\ 1 & 0 & \cdots\ 0 & 0 & \cdots & 0 \\ & \ddots & \vdots & \vdots & & \\ & 1 & 0 & & & \vdots \\ & & & 0 & \cdots & 0 \\ & & & 1 & & \\ & & & & \ddots & \vdots \\ & & & & 1 & \\ & & & & & 0 \end{bmatrix} \tag{28}$$

and thus $\nabla_{\underline{x}} \mathbf{F}(\underline{x},u) = F_x^\delta$ for some $\delta_i \in [-1,1]$, $i=1,\dots,N$. Therefore, it follows that for every $(\underline{x}',u')\in S$ there exists a $\delta' = [\delta_1',\dots,\delta_N']$ such that

$$\bar{\sigma}(\nabla_{\underline{x}} F(\underline{x}',u') = \bar{\sigma}(F_x^{\delta'}) \tag{29}$$

and

$$\min_T \bar{\sigma}(T\nabla_{\underline{x}} F(\underline{x}',u')\,T^{-1}) = \min_T \bar{\sigma}(TF_x^{\delta'}T^{-1}) \tag{30}$$

Since (30) holds for every $(\underline{x}',u')\in S$, then it is also true that

$$\max_{(\underline{x},u)\in S} \min_T \bar{\sigma}(T\nabla_{\underline{x}} F(\underline{x},u)\,T^{-1}) \leqslant \max_\delta \min_T \bar{\sigma}(TF_x^\delta\,T^{-1}) \tag{31}$$

Therefore $\max_\delta \min_T \bar{\sigma}(TF_x^\delta\,T^{-1})$ is an upper bound to the quantity in (27). This upper bound, however, is easier to solve because it can be casted as a convex optimization as shown next.

In what follows, it will prove convenient to write F_x^δ as an upper linear fractional transformation. For this, define

(i) $A = F_x^\delta$ 'evaluated' at $\delta_i = 0$ for $i=1,\dots,N$.

(ii) $H = e_{\underline{1}}$ (the matrix with ones in the first row and zeros elsewhere).

(iii) $G = \begin{bmatrix} \alpha_1^R & & \\ & \ddots & \\ & & \beta_{n_u}^R \end{bmatrix}$

(iv) E is an $N \times N$ matrix of zeros.

(v) $\hat{M}_{11} = \begin{bmatrix} E & G \\ H & A \end{bmatrix}$

(vi) $\Delta_1 = \begin{bmatrix} \delta_1 & & \\ & \ddots & \\ & & \delta_N \end{bmatrix}$

The $F_x^\delta = A + H\Delta_1(I - E\Delta_1)^{-1}G = F_u(\hat{M}_{11}, \Delta_1)$ and, as discussed in §2, the following problem:

$$\min_{D \in \mathscr{D}} \bar{\sigma}(D\hat{M}_{11}D^{-1}) \tag{32}$$

with

$$\mathscr{D} = \left\{ \begin{pmatrix} D & 0 \\ 0 & T \end{pmatrix} \middle| D = \operatorname{diag}(d_1, \ldots, d_N), d_i \in \mathbf{C}, d_i \neq 0; T \in \mathbf{C}^{N \times N}, T^{-1} \text{ exists} \right\}$$

serves as an upper bound to

$$\max_{\delta} \min_{\mathrm{T}} \bar{\sigma}(T\{F_x^\delta\}T^{-1})$$

The theory presented above can be formally summarized in the following theorem.

Theorem 3.4: *If a region S is known to contain an equilibrium point* $(\underline{x}^*, \underline{u}^*)$ *for an input* u^* *in S, and if*

$$\min_{\mathrm{D} \in \mathscr{D}} \bar{\sigma}(D\hat{M}_{11}D^{-1}) < 1$$

then the equilibrium point $(\underline{x}^*, u^*)$ *of* $\underline{x}(k+1) = \mathbf{F}(\underline{x}, u)$ *in S is asymptotically stable.*

Proof: The theorem has been essentially proven from the arguments in this section since it has been shown that if (32) is less than one, then (27) is less than one. In turn, this means that the Lipschitz constant for the transformed nonlinear system (using T as a transformation) is less than one and stability follows from Theorem 3.3. Since an invertible transformation cannot change the stability properties of a system then $\mathbf{F}(\underline{x}, u)$ is asymptotically stable. □

At this point we must comment on the conservatism of Theorem 3.4. Note that this theorem is derived from a series of conservative tests. First of all, the cones approximation is conservative as mentioned earlier. Afterwards, the contraction mapping theorem is used, which is a sufficient condition. The Lipschitz constant was calculated using the mean value theorem, thus it is conservative. Finally, the quantity computed in Theorem 3.4 is an upper bound to (27). Thus, there is considerable incentive in reducing the conservatism of this result. The following example, however, shows that the theorem is nevertheless useful.

Example 1: The following model has been identified as a continuous Stirred Tank Reactor from data collected around the lower stable branch of the equilibrium curve (see Hernández 1992 for details)

$$y(k+1) = \theta_0 + \theta_1 y(k) + \theta_2 u(k) + \theta_3 u(k-1) + \theta_4 y(k)\, u(k-1)\, u(k-2) + \theta_5 u^2(k-2) + \theta_6 y(k-1)\, u(k)\, u(k-1) \quad (33)$$

where

$$\theta_0 = 0{\cdot}041 \quad \theta_1 = 0{\cdot}719 \quad \theta_2 = 0{\cdot}013 \quad \theta_3 = 0{\cdot}012$$
$$\theta_4 = 0{\cdot}039 \quad \theta_5 = 0{\cdot}020 \quad \theta_6 = 0{\cdot}031$$

By defining the following state,

$$x_1(k) = y(k)$$
$$x_2(k) = y(k-1)$$
$$x_3(k) = u(k-1)$$
$$x_4(k) = u(k-2)$$

a realization of (33) is obtained as

$$\begin{bmatrix} x_1(k+1 \\ x_2(k+1) \\ x_2(k+1) \\ x_4(k+1) \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} \theta_0 + \theta_1 x_1(k) + \theta_2 u(k) + \theta_3 x_3(k) + \theta_4 z_1(k)\, x_3(k)\, x_4(k) \\ + \theta_5 (x_4(k))^3 + \theta_6 x_2(k)\, x_3(k)\, u(k) \end{pmatrix} \\ x_1(k) \\ u(k) \\ x_3(k) \end{bmatrix} \quad (34)$$

Direct calculations show that, for every input in $(-0{\cdot}5,\ 0{\cdot}5)$, there exists a steady-state output in $(0{\cdot}1,\ 0{\cdot}21)$. Thus, the region S is selected as

$$S = \{(x_1, x_2, x_3, x_4, u) \mid x_1, x_2 \in [0{\cdot}1, 0{\cdot}21],\ x_3, x_4, u \in [-0{\cdot}5, 0{\cdot}5]\}$$

The jacobian of (34) is given by

$$\nabla_x F = \begin{bmatrix} \theta_1 + \theta_4 x_3 x_4 & \theta_6 x_3 u & \theta_3 + \theta_4 x_1 x_4 + \theta_6 x_2 u & \theta_4 x_1 x_3 + 3\theta_5 (x_4)^2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (35)$$

which in S can be represented as

$$F_x^\delta = \begin{bmatrix} 0{\cdot}72 + \delta_1 0{\cdot}01 & 0 + \delta_2 0{\cdot}0326 & 0{\cdot}012 + \delta_3 0{\cdot}00947 & 0{\cdot}015 + \delta_4 0{\cdot}004 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

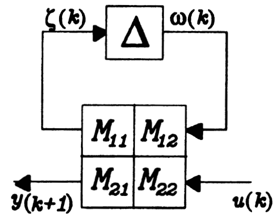

Figure 2. General $M-\Delta$ configuration.

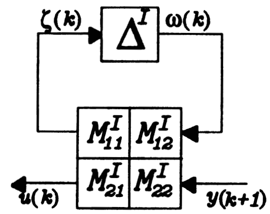

Figure 3. $M-\Delta$ configuration of uncertain difference equations.

for some $\delta_i \in [-1, 1]$, $i = 1, \ldots, 4$. For this example, $\hat{M}_{11}$ is given by

$$\hat{M}_{11} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0{\cdot}01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0{\cdot}01 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0{\cdot}008 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{\cdot}02 \\ 1 & 1 & 1 & 1 & 0{\cdot}72 & 0 & 0{\cdot}012 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{36}$$

Performing the optimization required in condition (32) results in

$$\bar{\sigma}(D^* \hat{M}_{11} (D^*)^{-1}) = 0{\cdot}7382$$

for

$$D^* = \begin{bmatrix} 0{\cdot}8547 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1{\cdot}2781 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 28{\cdot}748 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 28{\cdot}748 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0{\cdot}0695 & 0{\cdot}0063 & 0{\cdot}0233 & -0{\cdot}014 \\ 0 & 0 & 0 & 0 & -0{\cdot}0551 & 0{\cdot}159 & -0{\cdot}0039 & 0{\cdot}0051 \\ 0 & 0 & 0 & 0 & 0{\cdot}0581 & -0{\cdot}043 & 4{\cdot}1104 & -0{\cdot}0635 \\ 0 & 0 & 0 & 0 & -0{\cdot}0035 & 0{\cdot}0082 & 0{\cdot}0088 & 0{\cdot}1878 \end{bmatrix}$$

Therefore, the model obtained predicts that the equilibria on the chosen region of operation is stable. This prediction is correct since we chose the region to be part of the stable branch. □

3.3. '*Extended neighbourhood*' *invertibility*

This section concentrates on the study of the invertibility and stability of the model's inverse in an extended neighbourhood. By the inverse of a system Σ, we mean a system, Σ^{-1}, which takes as its input a signal r, and produces as an output a signal u which, if provided as an input to system Σ, the output of Σ is equal to the signal r delayed by the deadtime of system Σ. Note that the previous section studied the stability of general systems; thus the stability of the inverse of a model has been studied in principle. However, in order to use those results for the study of 'Extended Neighbourhood' invertibility, one would need to:

(1) construct an analytical expression for the inverse;

(2) reformulate the nonlinear inverse constructed as a linear plus (time-varying) uncertain one by using conic sector bounds on the nonlinear inverse.

However, Step 1 of this procedure is impractical in general since explicit sector bounds of the inverse are needed. The calculation of these bounds would require the analytical expression of the inverse.

The following result presents the first step of an alternative route for the determination of stability of the inverse. It studies the stability of the inverse of the linear time-varying system. This is justified since the state and output trajectories of the linear time varying system and the nonlinear system of interest are identical by construction, the inverse to one system is the inverse to the other. The invertibility result obtained here is important because it presents explicit formulae for the $M-\Delta$ representation of the inverse of a system given the $M-\Delta$ representation of the system.

Theorem 3.5: *Consider a system characterized by an $M-\Delta$ configuration as shown in Fig. 2. Then, if M_{22} and $(I-M_{11}\Delta)$ are invertible, explicit formulae for M^I and Δ^I of the inverse system shown in Fig. 3 are given by*

$$\Delta^I = \Delta$$

$$M_{11}^I = M_{11} - M_{12}M_{22}^{-1}M_{21}$$

$$M_{12}^I = M_{12}M_{22}^{-1}$$

$$M_{21}^I = -M_{22}^{-1}M_{21}$$

$$M_{22}^I = M_{22}^{-1}$$

Proof: The result follows from applying the matrix inversion lemma to $F_u(M,\Delta)$. □

Note that the assumptions of Theorem 3.5 are that: (*a*) $(I-M_{11}\Delta)$ and (*b*) M_{22} are invertible matrices. These are reasonable assumptions and are not restrictive. The first assumption is necessary for the relationship between y and u (namely $F_u(M,\Delta)$) to be well defined. The second assumption is necessary so that the map between $y(k+1)$ and $u(k-\rho)$ is invertible. This is an implicit assumption made in the very definition of the input–output map 2.

Theorem 3.5 can now be used to obtain 'extended neighbourhood' stability conditions for the inverse model.

Theorem 3.6: *Let the system in* (4) *be bounded in a region S by conic sectors with centre* $[A^0\, b^0]$ *and radii* $R_{1,x}, \ldots, R_{s,x}$, *and* r_b. *Using these bounds, let the system be represented by* (20). *Then, assuming that the system does not leave S, the equilibria of the inverse in S will be asymptotically stable provided that*

(*a*) $(I - M_{11}\Delta)$ *and* M_{22} *are invertible*

(*b*) $\min_{D \in \hat{D}} \bar{\sigma}(D M_{11}^I D^{-1}) < 1$

where M_{11}^I *is given in Theorem* 3.5.

Proof: The proof follows from the stability condition (12) for time-varying uncertain systems and that M_{11}^I is the correct M_{11} for the inverse system as proved in Theorem 3.5. □

Again, nothing in Theorem 3.6 limits the results to SISO systems: the result is valid for multivariable systems although care must be taken in the definition of the inverse for a multivariable system. A more detailed discussion can be found in Hernández (1992).

A parallel result to Theorem 3.4 can be formulated for the inverse system. The difference in the analysis of the open loop and inverse systems is that the input, $u(k)$, in the inverse system is a function of the state, $\underline{x}(k)$. Therefore, this dependence should be accounted for in the derivative of $\mathsf{F}(\underline{x}, u(\underline{x}))$ with respect to $\underline{x}$. This total derivative is now given by

$$\frac{\mathrm{d}F(\underline{x}, u)}{\mathrm{d}\underline{x}} = \frac{\partial F(\underline{x}, u)}{\partial \underline{x}} + \frac{\partial F(\underline{x}, u)}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}\underline{x}}$$

The derivation begins as before, noting that the partial derivative of $\mathsf{F}(\underline{x}, u)$ with respect to the state can be written as a constant matrix plus an uncertainty (29). In the same spirit, $(\partial F(\underline{x}, u)/\partial u)\,(\mathrm{d}u/\mathrm{d}\underline{x})$ can be represented as

$$\frac{\partial F(\underline{x}, u)}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}\underline{x}} = F_u^\delta \frac{\mathrm{d}u}{\mathrm{d}\underline{x}} = \begin{bmatrix} \gamma_u^0 + \delta_u \gamma_u^R \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \frac{\mathrm{d}u}{\mathrm{d}\underline{x}}$$

Thus, the total derivative of $\mathsf{F}(\underline{x}, u)$ with respect to the state is given by

$$\frac{\mathrm{d}F(\underline{x}, u)}{\mathrm{d}x} = F_x^\delta + F_u^\delta \frac{\mathrm{d}u}{\mathrm{d}\underline{x}}$$

which can be expressed as

where F_x^δ is given by (28) and

$$F_x^0 \equiv F_x^\delta(\delta = 0)$$
$$F_u^0 \equiv F_u^\delta(\delta = 0)$$
$$\omega = \Delta\zeta$$
$$\zeta = R_x + r_b \frac{\mathrm{d}u}{\mathrm{d}\underline{x}}$$

$$R_x = \begin{bmatrix} \alpha_1^R & 0 & & & 0 & 0 \\ & \ddots & & \cdots & & \\ & & \alpha_{n_y+1}^R & & & \\ & & & \beta_1^R & & \\ & & & & \ddots & \\ & & & & \cdots & \beta_{n_u}^R \\ 0 & & & & 0 & 0 \end{bmatrix}; \quad r_b = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \gamma_u^R \end{bmatrix}; \quad \Delta_1 = \begin{bmatrix} \delta_1 & & & & \\ & \delta_2 & & & \\ & & \ddots & & \\ & & & \delta_N & \\ & & & & \delta_u \end{bmatrix}$$

The total derivative of $y(\cdot)$ with respect to $\underline{x}$ is given by

$$\frac{\mathrm{d}y}{\mathrm{d}\underline{x}} = C\frac{\partial F}{\partial \underline{x}} + C\frac{\partial F}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}\underline{x}} \tag{37}$$

Using the linear-uncertainty representations, the following equations are derived

$$\begin{bmatrix} \zeta \\ \frac{\mathrm{d}F}{\mathrm{d}\underline{x}} \\ \frac{\mathrm{d}y}{\mathrm{d}\underline{x}} \end{bmatrix} = \underbrace{\left[\begin{array}{cc|c} \underline{0} & R_x & r_b \\ e_{\underline{1}} & F_{\underline{x}}^0 & F_u^0 \\ \hline Ce_{\underline{1}} & CF_{\underline{x}}^0 & CF_u^0 \end{array}\right]}_{\equiv \hat{M}} \begin{bmatrix} \omega \\ I \\ \frac{\mathrm{d}u}{\mathrm{d}\underline{x}} \end{bmatrix} \tag{38}$$

Since $u(\cdot)$ is obtained by equating $y(k+1)$ to a constant r^*, then $\mathrm{d}y/\mathrm{d}\underline{x}$ is zero and

$$\frac{\mathrm{d}u}{\mathrm{d}\underline{x}} = -\hat{M}_{22}^{-1}\hat{M}_{21}\begin{bmatrix} \omega \\ I \end{bmatrix}$$

Therefore

$$\begin{bmatrix} \zeta \\ \frac{\mathrm{d}F}{\mathrm{d}\underline{x}} \end{bmatrix} = \{\underbrace{\hat{M}_{11} - \hat{M}_{12}\hat{M}_{22}^{-1}\hat{M}_{21}}_{\equiv \hat{M}_{11}^I}\}\begin{bmatrix} \omega \\ I \end{bmatrix}$$

and

$$\frac{\mathrm{d}F}{\mathrm{d}\underline{x}} = F_u(\hat{M}_{11}^I, \Delta_1)$$

Using the same arguments that led to Theorem 3.4, one obtains the following sufficient condition for the stability of the nonlinear inverse in S

$$\min_{D_1 \in \hat{D}} \bar{\sigma}\left(\begin{pmatrix} D_1 & 0 \\ 0 & T \end{pmatrix} \hat{M}^I_{11} \begin{pmatrix} D_1^{-1} & 0 \\ 0 & T^{-1} \end{pmatrix}\right) < 1 \tag{39}$$

$$T \in \mathbf{C}^{N \times N}$$

$$\hat{D} = \{D \mid D = \operatorname{diag}(d_1, \ldots, d_{N+1}), d_i \in \mathbf{C}, \quad i = 1, \ldots\}$$

Formally, the result can be stated in the form of a Theorem.

Theorem 3.7: *If a region S is known to contain an equilibrium point for each desired output value, r^*, in S, and if condition (39) is satisfied, then all of the equilibrium points of the inverse system in S are asymptotically stable.*

Example 2: This example investigates the effects of large sampling times on the zero dynamics of an identified input–output model. The system to be studied is the Van de Vusse reaction in a CSTR. This system is known to contain a region of unstable zero dynamics. Therefore, if the system is sampled arbitrarily fast, the discretized system will also display zeros. A polynomial ARMA model for this system has been identified by Hernández (1992)

$$y(k+1) = \theta_0 + \theta_1 u_s(k) + \theta_2 y(k) + \theta_3 u_s^3(k) + \theta_4 y(k-1) u_s(k-1) u_s(k)$$

where the sampling time is 0·04 hours and u_s denotes the scaled inputs in [0, 1]. The parameters are given by

$$\theta_0 = 0{\cdot}558 \qquad \theta_1 = 0{\cdot}538 \quad \theta_2 = 0{\cdot}116$$
$$\theta_3 = -0{\cdot}127 \quad \theta_4 = -0{\cdot}034$$

By defining the state as

$$x_1(k) = y(k)$$
$$x_2(k) = y(k-1)$$
$$x_3(k) = u(k-1)$$

The following realization is obtained:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1 \end{bmatrix} = \begin{bmatrix} \theta_0 + \theta_1 u_s(k) + \theta_2 x_1(k) + \theta_3 u_s^3(k) + \theta_4 x_2(k) x_3(k) u_s(k) \\ x_1(k) \\ u_s(k) \end{bmatrix}$$

Direct calculations show that for every scaled input in [0, 1] there is a steady-state output in (0·6, 0·105). Thus, define the region S as

$$S = \{(x_1, x_2, x_3, u) \mid x_1, x_2 \in [0{\cdot}6, 1{\cdot}05]; x_3, u_s \in [0, 1]\}$$

In the region, F_x^δ can be written as

$$F_x^\delta = \begin{bmatrix} 0{\cdot}116 & -0{\cdot}017 + \delta_1 0{\cdot}017 & -0{\cdot}018 + \delta_2 0{\cdot}018 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and F_u^δ as

$$F_u^\delta = \begin{bmatrix} 0{\cdot}330 + \delta_3 0{\cdot}208 \\ 0 \\ 1 \end{bmatrix}$$

Thus, $\hat{M}$ is given by

$$\hat{M} = \left[\begin{array}{ccccccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0{\cdot}017 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0{\cdot}018 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{\cdot}208 \\ 1 & 1 & 1 & 1 & 0{\cdot}116 & -0{\cdot}017 & -0{\cdot}018 & 0{\cdot}330 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 1 & 1 & 1 & 0{\cdot}116 & -0{\cdot}017 & -0{\cdot}018 & 0{\cdot}330 \end{array}\right]$$

Therefore, $\hat{M}^I_{11}$ is given by

$$\hat{M}^I_{11} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0{\cdot}017 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0{\cdot}018 \\ -0{\cdot}632 & -0{\cdot}632 & -0{\cdot}632 & -0{\cdot}632 & -0{\cdot}073 & 0{\cdot}019 & 0{\cdot}011 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -3{\cdot}034 & -3{\cdot}034 & -3{\cdot}034 & -3{\cdot}034 & -0{\cdot}351 & 0{\cdot}052 & 0{\cdot}055 \end{bmatrix}$$

Finally, performing the optimization in (39), it is found that for

$$D^* = \begin{array}{ccccccc} \text{anything} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 17{\cdot}1158 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0{\cdot}7458 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0{\cdot}2521 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2{\cdot}0079 & 0{\cdot}744 & -0{\cdot}006 \\ 0 & 0 & 0 & 0 & 0{\cdot}0369 & 0{\cdot}3868 & 0{\cdot}0007 \\ 0 & 0 & 0 & 0 & -0{\cdot}0063 & -0{\cdot}001 & 0{\cdot}0235 \end{array}$$

Then

$$\bar{\sigma}(D^*\hat{M}^I_{11}(D^*)^{-1}) = 0{\cdot}7587$$

and thus, the equilibria of the inverse system in S are asymptotically stable when the sampling time is chosen very large.

4. Conclusions

In this paper the stability of nonlinear input/output dynamical models and the stability of their inverses were studied in an extended neighbourhood around an equilibrium point. The first type of result assumed that it is known *a priori* that the dynamical system does not leave the 'extended neighbourhood'. This would be the case, for example, if the system is part of a constrained control scheme where the controller guarantees that the system is kept in some region. Also, in some situations it can be guaranteed from physical laws that the system will not leave a region (e.g. mole fractions must be between 0 and 1). However, in more general settings, this first type of result is difficult to apply. To remedy this situation, the second type of result obtained combines the 'Contraction Mapping Theorem' of operator theory with linear robust control results to provide conditions that guarantee that the system does not leave the pre-specified region of analysis and that the system's equilibria in this region are asymptotically stable.

Acknowledgements

The financial support of IBM and International Paper are greatly acknowledged.

References

Chen, S. and Billings, S. A., 1989, Modelling and analysis of nonlinear time series. *International Journal of Control*, **50**, 2151.

Constanza, V., Dickinson, B. W. and Johnson, E. F., 1983, Universal approximations for discrete-time control systems over finite time. *IEEE Transactions on Automatic Control*, **28**, 439.

Díaz, H. and Desrochers, A., 1988, Modelling of nonlinear discrete-time systems for input/output data. *Automatica*, **24**, 629.

Doyle, F. J. and Morari, M., 1990, A conic sector based methodology for nonlinear controller design. *Proceedings of the 1990 American Control Conference*, p. 2946.

Doyle, F. J., Packard, A. K. and Morari, M., 1989, Robust controller design for a nonlinear CSTR *Chemical Engineering Science*, **44**, 1929.

Economou, C. G., 1988, An operator theory approach to nonlinear controller design. Ph.D. dissertation, Department of Chemical Engineering, California Institute of Technology.

Haber, R. and Unbehauen, H., 1990, Structure identification of nonlinear dynamical systems—a survey on input–output approaches. *Automatica*, **26**, 651.

Hernández, E., 1992, Control of nonlinear systems using input–output information. Ph.D. dissertation, School of Chemical Engineering, Georgia Institute of Technology.

Hernández, E. and Arkun, Y., 1993, Control of nonlinear systems using polynomial ARMA models. *AIChE*, **39**, 446.

Kammash, M. and Pearson, J. B., 1990, Performance robustness of discrete-time systems with structured uncertainty. *Proceedings of the 29th Conference on Decision and Control*, p. 414.

Korenberg, M., Billings, S. A., Liu, Y. P. and McIlroy, P. J., 1988, Orthogonal parameter estimation algorithm for nonlinear stochastic systems. *International Journal of Control*, **48**, 193.

Molchanov, A. P., 1987, Lyapunov functions for nonlinear discrete-time control systems. *Automation and Remote Control*, **48**, 728.

Narendra, K. S. and Parthasarathy, K., 1990, Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, **1**, 4.

Packard, A. K., 1988, What is new with μ: Structured uncertainty in multivariable control. Ph.D. dissertation, University of California, Berkeley, Electrical Engineering.

Priestley, M. B., 1988, *Nonlinear and Nonstationary Time Series Analysis* (San Diego, Calif.; Academic Press).

Vidyasagar, M., 1978, *Nonlinear Systems Analysis* (Englewood Cliffs, NJ: Prentice Hall).

Zames, G., 1966, On the input–output stability of time-varying nonlinear feedback systems—Part II: Conditions involving circles in the frequency plane and sector nonlinearities. *IEEE Transactions on Automatic Control*, **11**, 465.

The authors

Evelio Hernández *Control Engineering Department, Motiva Norco Refining Company, P.O. Box 10, 15536 River Road, Norco, Louisiana 70079 (EH725981@MSXSOPC.SHELL.COM).* Dr. Hernández is a Process Control Specialist at the Motiva Refinery in Norco, Louisiana. He received his Ph.D. degree from the Georgia Institute of Technology in 1992. He has published twelve papers in the area of process control. This paper was written while he was at the Georgia Institute of Technology. Dr. Hernández received the 1995 Ted Peterson Award given by the CAST Division of the American Institute of Chemical Engineers to an individual for outstanding published work in the application of computing and systems technology to chemical engineering. In 1996 he was named an Outstanding Young Engineering Alumnus of the Georgia Institute of Technology. He joined the Shell Development Company shortly after graduating from Georgia Tech to work on developing methods to assist control engineers in the design and implementation of process control systems. Currently, his work at Motiva is on the design and implementation of advanced control systems for refining processes. Dr. Hernández is a member of the American Institute of Chemical Engineers.

Yaman Arkun *College of Engineering, Koç University, Istiny e 80860, Istanbul, Turkey (yarkun@ku.edu.tr).* Dr. Arkun is the Dean of the College of Engineering at Koç University. He received his B.S. degree from the University of Bosphorous, Istanbul, Turkey, and his M.S. and Ph.D. degrees from the University of Minnesota. He was a member of the RPI faculty from 1979 to 1986. From 1986 to 1999, before moving to Turkey, he was a Professor of Chemical Engineering at the Georgia Institute of Technology. This paper was written during his tenure there. Dr. Arkun has held industrial visiting positions at the Tennessee Eastman, Du Pont, and Weyerhaeuser companies. He has served as the North American Editor of the *Journal of Process Control,* as the Editor of *Automatica,* as an editorial board member of the *International Journal of Control,* as Trustee and Secretary of CACHE, and as Director of the CAST Division of the American Institute of Chemical Engineers (AIChE). He is the 1986 recipient of the Donald P. Eckman Award given by the American Automatic Control Council. He has chaired the Systems and Control Area 10b of CAST (1988–1990) and served as the AIChE Director of the American Automatic Control Council (1989–1991). His research interests are in process dynamics, modeling, and control. In particular, he is interested in robust, nonlinear, and predictive control, and synthesis of control systems for large-scale plants.

Publication of this paper

This paper was originally published on pages 885–906 of the *International Journal of Control,* Volume 63, Number 5 (1996) [Copyright © 1996 by Taylor and Francis, Ltd. All rights reserved.]. It was produced by scanning the original version and contains added biographical sketches of its authors. We gratefully acknowledge permission to include it in this issue.