Copyright © 2002 IFAC
15th Triennial World Congress, Barcelona, Spain

STRUCTURED FAULT DIAGNOSIS IN MILDLY NONLINEAR SYSTEMS : PARITY SPACE AND INPUT-OUTPUT FORMULATIONS

Janos Gertler[1,2] and Marcel Staroswiecki[1]

[1]*LAIL-CNRS, Universitè des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq cedex, France*

[2]*School of Information Technology, George Mason University, Fairfax, VA 22030, USA*

jgertler@gmu.edu marcel.staroswiecki@univ-lille1.fr

Abstract. Concerned is residual generation in systems with mild nonlinearities, that is, nonlinearities which involve the inputs, outputs and faults but not the state. It is shown how structured residuals can be designed by the parity space method, in discrete time, if mild nonlinearities are present. Next a number of fundamental links between the parity space and the input-output representations are explored. Utilizing these results, a new design method is developed which finds structured residuals for the mildly nonlinear system entirely in the input-output framework. It is also shown that the two approaches, parity space and input-output, lead to identical residuals. *Copyright © 2002 IFAC.*

Keywords : Fault diagnosis ; residual generation ; nonlinear systems ; parity sapce ; consistency relations.

1. INTRODUCTION

Model-based methods of fault detection and isolation (diagnosis) rely on the idea of analytical redundancy (Willsky, 1976). The essence of this idea is that measured plant outputs are compared to ones predicted, with the model, from the measured or actuated inputs. Discrepancies, expressed as residuals, are ideally indications of faults, though in reality they are also affected by disturbances, noise and modeling errors. While a single residual may be sufficient to detect any fault in the system, the isolation of faults requires a set of residuals. These are subjected to mathematical manipulations to make them selectively sensitive to subsets of faults. The design of such enhanced residuals, referred to as structured, requires the knowledge of how the various faults act on the system. Decoupling from disturbances, whose effect on the system is known, is also easily included in the design of structured residuals.

There are various methods to generate residuals and to enhance them for fault isolation. These methods include diagnostic observers, parity relations from the state-space model, and input-output consistency relations. The linear theory of these approaches is well developed and their relationship is also rather well understood

Diagnostic observers have been extensively studied by Frank and coworkers (Frank, 1990) . Parity relations were introduced by Chow and Willsky (1984), and extended to structured design by Gertler and Luo (1989) and by Staroswiecki and coworkers (1993). Input-output consistency relations have been investigated by Gertler and coworkers (Gertler and Singer, 1990 ; Gertler, 1998). The equivalence of the various methods has been studied by several authors, see for example (Gertler, 2000).

More recently, interest has been shifted to residual generation in nonlinear systems. Results on nonlinear observers have been published by Frank and coworkers (Alcorta-Garcia and Frank, 1997), Kinnaert (1999) and others. A purely algebraic approach to nonlinear residual generation has been reported by Staroswiecki and Comtet-Varga (2001) . Cocquempot and Christophe (2000) have demonstrated an extension of the parity space method to mild nonlinearities, in continuous time, and showed the relationship of this approach to a class of nonlinear observers.

This work was supported by the French Ministry of Research under the program « sejours de haut niveau » and by NSF under grant #ECS-9906250.

In this paper, we will investigate systems with mild nonlinearities, that is, nonlinearities which involve the inputs, outputs and faults but not the state. We will revisit the parity space method, in discrete time, and show how structured residuals can be designed if mild nonlinearities are present. Next we will investigate a number of fundamental links between the parity space and the input-output representations. Then, utilizing these results, we will develop a design method which finds structured residuals for the mildly nonlinear system entirely in the input-output framework. It will also be shown that the parity space and input-output designs lead to identical residuals also in this case.

2. PROBLEM STATEMENT

Consider a dynamic system with outputs $\mathbf{y}(t)=[y_1(t) \ldots y_m(t)]'$, inputs $\mathbf{u}(t)=[u_1(t) \;\; u_2(t) \;\; \ldots \;]'$ and faults and disturbances $\mathbf{p}(t)=[p_1(t) \;\; \ldots \;\; p_\kappa(t)]'$. Assume that the system can be described by the *n*-th order discrete-time state-space model

$$\mathbf{x}(t+1) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{B}\,\boldsymbol{\varphi}(t) + \boldsymbol{\Psi}(t)\,\mathbf{p}(t) \qquad (1a)$$

$$\mathbf{y}(t) = \mathbf{C}\,\mathbf{x}(t) \qquad (1b)$$

where $\quad \boldsymbol{\varphi}(t) = \boldsymbol{\varphi}\,[\mathbf{y}(t), \mathbf{u}(t)] = [\varphi_1(t) \ldots \varphi_k(t)]' \qquad (2a)$

is a vector of ‘pseudo-inputs’, which are nonlinear functions of the inputs and outputs, and where

$$\boldsymbol{\Psi}(t) = \boldsymbol{\Psi}[\mathbf{y}(t), \mathbf{u}(t)] \qquad (2b)$$

is an $n \cdot \kappa$ nonlinear matrix function of the same. Though this system is nonlinear, the nonlinearities do not involve the state vector. Further, the faults $\mathbf{p}(t)$ appear in a quasy-linear way, in that they are multiplied with a coefficient matrix which depends on the inputs and outputs (and, via these, on time). As it has been shown by Isidori (1995), a broad class of nonlinear systems may be transformed into this form.

System (1) may arise from a somewhat more general nonlinear system by linearization. Consider

$$\mathbf{x}(t+1) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{B}\,\boldsymbol{\phi}(t) \qquad (3a)$$

$$\mathbf{y}(t) = \mathbf{C}\,\mathbf{x}(t) \qquad (3b)$$

where $\quad \boldsymbol{\phi}(t) = \boldsymbol{\phi}[\mathbf{y}(t), \mathbf{u}(t), \mathbf{p}(t)] = [\phi_1(t) \ldots \phi_k(t)]' \quad (4)$

is set of more general pseudo-inputs with the faults also included in the nonlinear function. Now let us expand $\boldsymbol{\phi}(t)$ with respect to the faults as

$$\boldsymbol{\phi}(t) = \boldsymbol{\varphi}(t) + \boldsymbol{\Gamma}(t)\,\mathbf{p}(t) \qquad (5)$$

where now $\quad \boldsymbol{\varphi}(t) = \boldsymbol{\phi}[\mathbf{y}(t), \mathbf{u}(t), \mathbf{0}] \qquad (6a)$

are the nominal values of the pseudo-inputs $\boldsymbol{\phi}(t)$ and

$$\boldsymbol{\Gamma}(t) = \partial\boldsymbol{\phi}(t)/\partial\mathbf{p}(t)|_{\mathbf{p}(t)=\mathbf{0}} = [\partial\phi_i(t)/\partial p_j(t)]_{\,\mathbf{p}(t)=\mathbf{0}} \qquad (6b)$$

is the fault Jacobian. System (3) is thus converted into system (1), with $\boldsymbol{\Psi}(t) = \mathbf{B}\boldsymbol{\Gamma}(t)$, but now an approximation is involved due to the linearization of the fault effect.

The column-rank properties of the coefficient matrix $\boldsymbol{\Psi}(t)$ and, consequently, of $\boldsymbol{\Gamma}(t)$ play an important role in the isolation of faults. Rank deficiencies may restrict the attainable residual structures while direct linear dependence between a pair of columns renders the concerned faults nonisolable (Gertler, 1998). Rank deficiencies may be due to a number of reasons, including linear relations among pseudo-inputs or among plant inputs and/or outputs.

3. PARITY SPACE FORMULATION

The parity space formulation of residual generation, introduced by Chow and Willsky for linear systems (Chow and Willsky, 1984), is easily extended to the mildly nonlinear system (1), or to (3) with linearized fault-effects. The Chow-Willsky scheme relies on the repeated application of the state-output equation. With the model (1), and for a window-width of σ, the repeated equations lead to the matrix formulation

$$\mathbf{Y}(t) = \mathbf{J}\,\mathbf{x}(t-\sigma) + \mathbf{K}\,\boldsymbol{\Phi}(t) + \boldsymbol{\Lambda}(t)\,\mathbf{P}(t) \qquad (7)$$

where

$$\mathbf{Y}(t)=[\mathbf{y}'(t-\sigma) \;\; \mathbf{y}'(t-\sigma+1) \ldots \mathbf{y}'(t-1) \;\; \mathbf{y}'(t)]' \qquad (8a)$$

$$\boldsymbol{\Phi}(t)=[\boldsymbol{\varphi}'(t-\sigma) \;\; \boldsymbol{\varphi}'(t-\sigma+1) \ldots \boldsymbol{\varphi}'(t-1) \;\; \boldsymbol{\varphi}'(t)]' \qquad (8b)$$

$$\mathbf{P}(t)=[\mathbf{p}'(t-\sigma) \;\; \mathbf{p}'(t-\sigma+1) \ldots \mathbf{p}'(t-1) \; \mathbf{p}'(t)]' \qquad (8c)$$

$$\mathbf{J} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \ldots\ldots \\ \mathbf{CA} \end{bmatrix} \qquad (9a)$$

$$\mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \ldots\ldots\ldots & & & \\ \mathbf{CB} & \mathbf{0} & & & & \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{0} & \ldots\ldots\ldots & & \\ \ldots\ldots\ldots\ldots\ldots\ldots & & & & & \\ \mathbf{CA}^{\sigma-1}\mathbf{B} & \mathbf{CA}^{\sigma-2}\mathbf{B} & \ldots\ldots & \mathbf{CAB} & \mathbf{CB} & \mathbf{0} \end{bmatrix} \qquad (9b)$$

$$\boldsymbol{\Lambda}(t) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \ldots\ldots\ldots & & \\ \mathbf{C}\boldsymbol{\Psi}(t-\sigma) & \mathbf{0} \ldots\ldots\ldots & & \\ \mathbf{CA}\,\boldsymbol{\Psi}(t-\sigma) & \mathbf{C}\boldsymbol{\Psi}(t-\sigma+1) & \mathbf{0} \;\ldots\ldots & \\ \ldots\ldots\ldots\ldots\ldots & & & \\ \mathbf{CA}^{\sigma-1}\,\boldsymbol{\Psi}(t-\sigma) & \mathbf{CA}^{\sigma-2}\boldsymbol{\Psi}(t-\sigma+1) \ldots & \mathbf{C}\boldsymbol{\Psi}(t-1) & \mathbf{0} \end{bmatrix} \qquad (9c)$$

From (7), a primary pre-residual set is obtained as

$$\mathbf{E}(t) = \mathbf{Y}(t) - \mathbf{K}\,\boldsymbol{\Phi}(t) = \mathbf{J}\,\mathbf{x}(t-\sigma) + \boldsymbol{\Lambda}(t)\,\mathbf{P}(t) \qquad (10)$$

A scalar residual may then be generated by the transformation

$$r(t) = \mathbf{W}(t)\, \mathbf{E}(t) \tag{11}$$

where $\mathbf{W}(t) = [\mathbf{w}^{\sigma}(t) \quad \mathbf{w}^{\sigma-1} \quad \mathbf{w}^{\sigma-2}(t) \ldots \mathbf{w}^{0}(t)\,]$ (12a)

$$\mathbf{w}^{q}(t) = [w^{q}_{1}(t) \quad \ldots\ w^{q}_{m}(t)\,] \qquad q=0\ldots\sigma \tag{12b}$$

with the superscripts representing indices.

For r(t) to qualify as a residual, it must be decoupled from $\mathbf{x}(t-\sigma)$, and also from the disturbances in $\mathbf{P}(t)$. For it to serve as an element of a structured set, it must also be decoupled from an appropriate subset of the faults in $\mathbf{P}(t)$, while maintaining sensitivity to the remaining faults. For decoupling from $\mathbf{x}(t-\sigma)$, $\mathbf{W}(t)$ must be orthogonal to all the n columns of $\mathbf{J}$. To decouple from the j-th fault or disturbance, $\mathbf{W}(t)$ must be orthogonal to all columns of the $\mathbf{\Lambda}_j(t)$ matrix, composed of those columns $\boldsymbol{\lambda}_{.}(t)$ of $\mathbf{\Lambda}(t)$ which belong to the various occurances (time-shifts) of the j-th fault (disturbance) :

$$\mathbf{\Lambda}_j(t) = [\ \boldsymbol{\lambda}_j(t) \quad \boldsymbol{\lambda}_{\kappa+j}(t) \quad \ldots \quad \boldsymbol{\lambda}_{(\sigma-1)\kappa+j}(t) \quad \mathbf{0}\ ] \tag{13}$$

(where κ is the number of faults /disturbances in the system). Note that $\mathbf{\Lambda}_j(t)$ follows the exact structure of $\mathbf{\Lambda}(t)$, see (9c), with $\mathbf{\Psi}(t-q)$, $q=1\ldots\sigma$, replaced by its j-th column $\boldsymbol{\psi}_j(t-q)$. For simultaneous decoupling from ρ faults/ disturbances, $\mathbf{W}(t)$ must be orthogonal to a submatrix $\mathbf{\Lambda}^{\#}(t)$ of $\mathbf{\Lambda}(t)$, containing all the concerned columns:

$$\mathbf{\Lambda}^{\#}(t) = [\mathbf{\Lambda}_{j1}(t) \ \ldots \ \mathbf{\Lambda}_{j\rho}(t)] \tag{14}$$

That is, the complete decoupling conditions are

$$\mathbf{W}(t)\,[\ \mathbf{J} \quad \mathbf{\Lambda}^{\#}(t)\ ] = \mathbf{0} \tag{15}$$

This represents a total of $n + \rho.\sigma$ homogeneous conditions for the $m(\sigma + 1)$ elements of the $\mathbf{W}(t)$ vector. For a non-trivial solution

$$m(\sigma + 1) \geq n + \rho\sigma + 1 \tag{16}$$

from which

$$\sigma(m - \rho) \geq n - m + 1 \tag{17}$$

Consider first $m \leq n$. Then with $\rho<m$, the condition is

$$\sigma \geq (n - m + 1)/(m - \rho) \tag{18}$$

With $\rho \geq m$, there is no finite non-negative solution. Now consider $m>n$. In this case, the rank of each group of ρ columns in $\mathbf{\Lambda}(t)$ is $\min(\rho,n)$, so the number of faults from which selective decoupling is possible is $\rho<n$. With this, $\sigma=1$ is the minimum solution.

The solution requires the assignment of one (or more) free paramaters. With only one free parameter, the direction of the transformation $\mathbf{W}(t)$ is defined and the choice of the parameter only scales the solution.

A particular residual structure is attainable if, while decoupled from the selected faults and disturbances, the residual remains sensitive to all the other faults. This requires that

$$\text{Rank}\,[\mathbf{\Lambda}^{\#}(t) \quad \mathbf{\Lambda}_g(t)\,] > \text{Rank}\,\mathbf{\Lambda}^{\#}(t) \tag{19}$$

for any fault-related $\mathbf{\Lambda}_g(t)$ outside $\mathbf{\Lambda}^{\#}(t)$. Because of the way the $\mathbf{\Lambda}(t)$ matrix is constructed, see (9c), any linear dependence among the columns of $\mathbf{\Psi}(t-q)$ is transferred to $\mathbf{\Lambda}(t)$. Define

$$\mathbf{\Psi}^{\#}(t-q) = [\boldsymbol{\psi}_{j1}(t-q) \ \ldots \ \boldsymbol{\psi}_{j\rho}(t-q)\,] \tag{20}$$

along the lines of (15) . Then (20) requires that

$$\text{Rank}[\mathbf{\Psi}^{\#}(t-q) \quad \boldsymbol{\psi}_g(t-q)] > \text{Rank}\,\mathbf{\Psi}^{\#}(t-q) \tag{21}$$

for any fault-related column $\boldsymbol{\psi}_g(t-q)$ outside $\mathbf{\Psi}^{\#}(t-q)$, for at least one q between 1 and σ. This implies $\rho < n$.

Note that r(t), as defined here, is a perfect residual, in the sense that its value is zero in the absence of faults and disturbances, as long as the nonlinear model (1) is a correct description of the system. That is, any linearization of the fault-effects does not influence the accuracy of the residual, only its decoupling from the selected faults and disturbances.

4. INPUT-OUTPUT FORMULATION

The input-output relationship, derived from (1), is

$$\mathbf{y}(t) = [\mathbf{G}(z)/h(z)]\ \boldsymbol{\varphi}(t) + [\mathbf{F}(z)/h(z)]\ \mathbf{\Psi}(t)\ \mathbf{p}(t) \tag{22}$$

where

$$\mathbf{F}(z) = \mathbf{F}^1 z^{-1} + \ldots + \mathbf{F}^n z^{-n} = \mathbf{C}\text{Adj}(\mathbf{I} - z^{-1}\mathbf{A})z^{-1}$$

$$\mathbf{G}(z) = \mathbf{G}^1 z^{-1} + \ldots + \mathbf{G}^n z^{-n} = \mathbf{C}\text{Adj}(\mathbf{I} - z^{-1}\mathbf{A})z^{-1}\mathbf{B} = \mathbf{F}(z)\mathbf{B}$$

$$h(z) = 1 + h^1 z^{-1} + \ldots + h^n z^{-n} = \text{Det}(\mathbf{I} - z^{-1}\mathbf{A}) \tag{23}$$

Note that we interpret the shift operator applied to an expression with a time-varying coefficient as

$$z^{-1}\,\mathbf{\Psi}(t)\,\mathbf{p}(t) = z^{-1}\,[\mathbf{\Psi}(t)\,\mathbf{p}(t)] = \mathbf{\Psi}(t-1)\,\mathbf{p}(t-1) \tag{24}$$

Eq. (26) gives rise to the primary residual

$$\mathbf{e}(t) = \mathbf{y}(t) - [\mathbf{G}(z)/h(z)]\boldsymbol{\varphi}(t) = [\mathbf{F}(z)/h(z)]\mathbf{\Psi}(t)\mathbf{p}(t) \tag{25}$$

where the mid-section is the computational form and the right-hand side is the fault-effect form of the residual. Note that $\mathbf{F}(z)$, $\mathbf{G}(z)$ and $h(z)$, as defined in (23) are, in general, not relative prime; they may contain an excess pole/zero polynomial $\vartheta(z)$.

We will seek structured residuals by the transformation

$$r(t) = \mathbf{w}(z, t)\ \mathbf{e}(t) \tag{26}$$

where $\mathbf{w}(z, t) = \mathbf{w}^0(t) + \mathbf{w}^1(t)z^{-1} + \ldots + \mathbf{w}^{\sigma}(t)z^{-\sigma}$ (27)

The transformation $\mathbf{w}(z, t)$ will be so designed that the residual r(t) is decoupled from the selected faults and disturbances. First, however, several links between the parity-space and the input-output description will be explored.

5. RELATIONSHIP BETWEEN PARITY SPACE AND INPUT-OUTPUT DESCRIPTIONS

The Chow-Willsky scheme and the input-output model are two descriptions of the same system. Naturally, they are related in several ways. Two of those relationships will be explored below. One concerns the generation of the input-output model by a transformation applied to the Chow-Willsky model. The other reveals an important property of the de-coupling transformation, designed in the parity space framework, in relation to the input-output model.

Lemma 1. The Chow-Willsky model (7), with the transformation

$$[h^n \mathbf{I}_m \dots h^1 \mathbf{I}_m \quad \mathbf{I}_m]\mathbf{Y}(t) = \qquad (28)$$

$$[h^n \mathbf{I}_m \dots h^1 \mathbf{I}_m \quad \mathbf{I}_m][\mathbf{J}\,\mathbf{x}(t\text{-}n) + \mathbf{K}\,\mathbf{\Phi}(t) + \mathbf{\Lambda}(t)\,\mathbf{P}(t)]$$

where $\mathbf{I}_m$ is an *m.m* unit matrix, yields the input-output model (22) in the polynomial form

$$h(z)\mathbf{y}(t) = \mathbf{G}(z)\,\boldsymbol{\varphi}(t) + \mathbf{F}(z)\,\mathbf{\Psi}(t)\,\mathbf{p}(t) \qquad (29)$$

The proof, which utilizes the Cayley-Hamilton theorem, will not be presented here.

This Lemma also provides an alternative expression for the $\mathbf{F}^q$ and $\mathbf{G}^q$ matrices (compare to (23)) :

$$\mathbf{F}^q = \mathbf{C} \sum_{\mu=0}^{q-1} h^{\mu}\,\mathbf{A}^{q-1-\mu} \qquad \mathbf{G}^q = \mathbf{F}^q\,\mathbf{B} \qquad q=1\dots n \qquad (30)$$

Lemma 2. Consider a transforming vector $\mathbf{W}(t) = [\mathbf{w}^{\sigma}(t) \quad \mathbf{w}^{\sigma-1}(t) \quad \dots \quad \mathbf{w}^1(t) \quad \mathbf{w}^0(t)]$, with any window-width $\sigma>0$, designed to be orthogonal to the $\mathbf{J}$ matrix of appropriate size. Form a vector polynomial $\mathbf{w}(z, t) = \mathbf{w}^0(t)+\mathbf{w}^1(t)z^{-1}+\dots+\mathbf{w}^{\sigma-1}(t)z^{-\sigma+1}+\mathbf{w}^{\sigma}(t)z^{-\sigma}$. Then the following hold :

a. $\mathbf{w}(z, t)\,\mathbf{F}(z) = \boldsymbol{\alpha}(z, t)\,h(z)$ (31)

where $\boldsymbol{\alpha}(z, t) = \boldsymbol{\alpha}^1(t)z^{-1} + \dots + \boldsymbol{\alpha}^{\sigma}(t)z^{-\sigma}$ (32a)

$$\boldsymbol{\alpha}^q(t) = [\alpha^q{}_1(t) \dots \alpha^q{}_n(t)] \qquad q=1\dots\sigma \qquad (32b)$$

b. If $\mathbf{w}(z, t)$ is applied to the input-output primary residual (25) then the transformed residual is

$$r(t)=\mathbf{w}(z, t)\mathbf{y}(t) - \boldsymbol{\beta}(z, t)\boldsymbol{\varphi}(t) = \boldsymbol{\alpha}(z, t)\mathbf{\Psi}(t)\mathbf{p}(t) \qquad (33)$$

where $\boldsymbol{\beta}(z, t)= \boldsymbol{\alpha}(z, t)\mathbf{B}$, and this is identical with the residual obtained by applying $\mathbf{W}(t)$ to the parity space pre-residual (10) .

Proof. The proof is constructive. Write the $\mathbf{w}(z, t)\mathbf{F}(z)$ product expressing $\mathbf{F}^q$ by (30) and then simplify it utilizing the orthogonality equation $\mathbf{W}(t)\mathbf{J} = \mathbf{0}$ and the Cayley-Hamilton relation. Simple manipulations yield

$$\mathbf{w}(z, t)\,\mathbf{F}(z) = [\mathbf{w}^0\,\mathbf{C}\,z^{-1} + (\mathbf{w}^1\,\mathbf{C} + \mathbf{w}^0\,\mathbf{CA})\,z^{-2} + \dots$$
$$+ (\mathbf{w}^{\sigma-1}\mathbf{C}+\mathbf{w}^{\sigma-2}\,\mathbf{CA}+\dots+\mathbf{w}^0\mathbf{CA}^{\sigma-1})z^{-\sigma}]\,h(z) \qquad (34)$$

This proves (31). Then (33) follows by

$$r(t)=\mathbf{w}(z, t)\mathbf{e}(t)=\mathbf{w}(z, t)\mathbf{y}(t) - \mathbf{w}(z, t)[\mathbf{G}(z)/h(z)]\boldsymbol{\varphi}(t)$$
$$= \mathbf{w}(z, t)\mathbf{y}(t) - \boldsymbol{\alpha}(z, t)\mathbf{B}\boldsymbol{\varphi}(t) \qquad (35)$$

Finally, from (34)

$$\boldsymbol{\alpha}^q(t) = \sum_{\mu=0}^{q-1} \mathbf{w}^{\mu}(t)\,\mathbf{CA}^{q-1-\mu} \qquad q=1\dots\sigma \qquad (36)$$

Recall that $\boldsymbol{\alpha}^q(t)$ is the coefficient of $\mathbf{\Psi}(t\text{-}q)\mathbf{p}(t\text{-}q)$ and $\boldsymbol{\alpha}^q(t)\mathbf{B}$ is that of $\boldsymbol{\varphi}(t\text{-}q)$. A comparison with the Chow-Willsky scheme, (9)-(12), reveals that (36) yields exactly the same as the respective coefficients in that scheme. Also, the coefficient of $\mathbf{y}(t\text{-}q)$ is clearly $\mathbf{w}^q$ in both schemes. Thus the two residuals are, indeed, identical. ♣

Lemma 2 has two important consequences :

1. The transformation $\mathbf{W}(t)$, computed in the parity space framework, can be used, after reformatting into $\mathbf{w}(z, t)$, to obtain the residual in the input-output framework.
2. Equations (31) and (33) will serve as the basis for a design algorithm to compute the transformation $\mathbf{w}(z, t)$ directly in the input-output framework, without even obtaining the state-space model of the system. This will be done in the next section.

6. DIRECT INPUT-OUTPUT TRANSFORMATION

In this section, we develop an algorithm to compute the transformation $\mathbf{w}(z, t)$ directly in the input-output framework. The algorithm rests on Eqs. (31) and (33) and consists of two steps

1. First we design, at least partially, a set of $\boldsymbol{\alpha}^1(t) \dots \boldsymbol{\alpha}^{\sigma}(t)$ vectors, so that they provide decoupling from all occurances of the fault and disturbance subset $\mathbf{p}^{\#}(t)$, that is, from $\mathbf{p}^{\#}(t\text{-}1) \dots \mathbf{p}^{\#}(t\text{-}\sigma)$, by being orthogonal to the $\mathbf{\Psi}^{\#}(t\text{-}1) \dots \mathbf{\Psi}^{\#}(t\text{-}\sigma)$ submatrices, as suggested by (33) ;

2. Then we design the $\mathbf{w}(z, t)$ transformation so that (31) is satisfied, that is, $\mathbf{w}(z, t)\mathbf{F}(z)$ is divisible by h(z) and the quotient is $\boldsymbol{\alpha}(z, t)$ computed in the first step.

Step 1 : Design of the $\alpha(z, t)$ vector. Expand the right-hand side of (33) as

$$r(t)=\boldsymbol{\alpha}^1(t)\mathbf{\Psi}(t\text{-}1)\mathbf{p}(t\text{-}1)+\dots+\boldsymbol{\alpha}^{\sigma}(t)\mathbf{\Psi}(t\text{-}\sigma)\mathbf{p}(t\text{-}\sigma) \qquad (37)$$

To decouple this residual from $\mathbf{p}^{\#}(t\text{-}1) \dots \mathbf{p}^{\#}(t\text{-}\sigma)$, the $\boldsymbol{\alpha}^1(t) \dots \boldsymbol{\alpha}^{\sigma}(t)$ vectors have to be orthogonal to the respective coefficient matrices $\mathbf{\Psi}^{\#}(t\text{-}1) \dots \mathbf{\Psi}^{\#}(t\text{-}\sigma)$, the latter containing the columns which belong to the selected fault/disturbance subset. That is,

$$\boldsymbol{\alpha}^q(t)\,\boldsymbol{\Psi}^{\#}(t-q) = \mathbf{0} \qquad q=1\ldots\sigma \tag{38}$$

Let the number of faults/disturbances in $\mathbf{p}^{\#}(t)$ be ρ (as before). The vectors $\boldsymbol{\alpha}^q(t)$ contain n elements each. Thus, provided that at least one of the matrices $\boldsymbol{\Psi}^{\#}(t-q)$ has full rank, the homogeneous conditions (38) can only be satisfied in a nontrivial way if $\rho < n$. (Note that this is the same condition found for the Chow-Willsky scheme.) Eq. (38) leaves $n-\rho$ elements of each of the $\boldsymbol{\alpha}^q(t)$ vectors free; we are keeping these undefined until the second step of the design. Here we only express the rest of the elements in terms of the free ones. Decompose $\boldsymbol{\alpha}^q(t)$ and $\boldsymbol{\Psi}^{\#}(t-q)$ as

$$\boldsymbol{\alpha}^q(t)=[\boldsymbol{\alpha}^q_{I}(t)\quad \boldsymbol{\alpha}^q_{II}(t)] \qquad \boldsymbol{\Psi}^{\#}(t-q)=\begin{bmatrix}\boldsymbol{\Psi}^{\#}_{I}(t-q)\\ \boldsymbol{\Psi}^{\#}_{II}(t-q)\end{bmatrix} \tag{39}$$

where $\boldsymbol{\alpha}^q_{I}(t)$ contains the free parameters and $\boldsymbol{\alpha}^q_{II}(t)$ the rest, and $\boldsymbol{\Psi}^{\#}(t-q)$ is partitioned according-ly. Then $\boldsymbol{\alpha}^q_{II}(t)$ can be expressed in terms of $\boldsymbol{\alpha}^q_{I}(t)$ as

$$\boldsymbol{\alpha}^q_{II}(t)= -\boldsymbol{\alpha}^q_{I}(t)\boldsymbol{\Psi}^{\#}_{I}(t-q)[\boldsymbol{\Psi}^{\#}_{II}(t-q)]^{-1} \qquad q=1\ldots\sigma \tag{40}$$

With (39) and (40), $\boldsymbol{\alpha}^q(t)$ may be written as

$$\boldsymbol{\alpha}^q(t)=\boldsymbol{\alpha}^q_{I}(t)[\mathbf{I}_{n-\rho}\quad -\boldsymbol{\Psi}^{\#}_{I}(t-q)[\boldsymbol{\Psi}^{\#}_{II}(t-q)]^{-1}]=\boldsymbol{\alpha}^q_{I}(t)\,\Xi(t-q) \qquad q = 1\ldots\sigma \tag{41}$$

where $\mathbf{I}_{n-\rho}$ is a unit matrix of size $n-\rho$ and $\Xi(t-q)$ is as defined in the equation. Numerically the same matrix $\Xi(t-q)$ is used in multiple samples, first as $\Xi(t-1)$, then as $\Xi(t-2)$, etc. until $\Xi(t-\sigma)$, so that only one such expression needs to be newly computed at each sample.

Step 2 : Design of the w(z, t) vector. The second step of the design relies on (31). The excess pole-zero factor $\vartheta(z)$ can be canceled out from the equation. Further, to avoid difficulties with zero-valued poles, positive-power polynomials will be used here. Thus

$$\mathbf{w}^+(z,t)\,\mathbf{F}^+(z) = \boldsymbol{\alpha}^+(z,t)\,h^+(z) \tag{42}$$

where $\mathbf{F}^+(z)=z^{\nu}\mathbf{F}(z)/\vartheta(z) \qquad h^+(z)=z^{\nu}h(z)/\vartheta(z)$

$$\mathbf{w}^+(z,t)=z^{\sigma}\mathbf{w}(z,t) \qquad \boldsymbol{\alpha}^+(z,t)=z^{\sigma}\boldsymbol{\alpha}(z,t) \tag{43}$$

and $(n-\nu)$ is the order of the excess pole/zero polynomial $\vartheta(z)$. Eq . (42) will be solved by assigning numerical values to the variable z. The polynomials on either side of (42) are of degree $\nu+\sigma-1$, thus the equation must be satisfied at $\nu+\sigma$ different values of z . We will pick the ν roots of $h^+(z)$, $\zeta_1 \ldots \zeta_\nu$, and σ additional values, $z_1 \ldots z_\sigma$.

At a pole ζ_i of $h^+(z)$, (42) becomes

$$\mathbf{w}^+(\zeta_i,t)\,\mathbf{F}^+(\zeta_i) = \mathbf{0} \tag{44}$$

Here we make the following observations:

Observation 3. a. For any single pole ζ_i of $h^+(z)$, Rank$\mathbf{F}^+(\zeta_i)$ does not exceed the total multiplicity of ζ_i in $h(z)$. This follows from the way how excess poles/zeros are created, see Gilbert's method, e.g. in (Kailath, 1980).

b. For any pole ζ_i which has a multiplicity of κ in $h^+(z)$, (44) is supplemented with the derivative equations

$$d^s[\mathbf{w}^+(z,t)\,\mathbf{F}^+(z)]/dz^s\Big|_{z=\zeta_i} = 0 \qquad s=1\ldots\kappa-1 \tag{45}$$

This follows from the fact that, at a multiple pole, the first κ derivatives of the righ-hand side in (44) are zero.

Cases a. and b. may occur in combination. Thus the number of scalar equations (44) represents for any pole ζ_i , does not exceed the total multiplicity of ζ_i in $h(z)$, and the total number of scalar conditions from (44) does not exceed n. ♣

The remaining solution equations arise from (42) by substituting $z=z_i$, $i=1\ldots\sigma$:

$$\mathbf{w}^+(z_i,t)\mathbf{F}^+(z_i) - \boldsymbol{\alpha}^+(z_i,t)h^+(z_i) = \mathbf{0} \qquad i=1\ldots\sigma \tag{46}$$

where the z_i values are arbitrary but distinct and different from any pole ζ_i. With (41), this may be rewritten in a more direct form as

$$\sum_{q=0}^{\sigma} z_i^{\sigma-q}\mathbf{w}^q(t)\mathbf{F}^+(z_i) - \sum_{q=1}^{\sigma} z_i^{\sigma-q}\,\boldsymbol{\alpha}^q_{I}(t)\Xi(t-q)h^+(z_i) = 0 \qquad i = 1\ldots\sigma \tag{47}$$

Eq . (44) (alone or together with (45)) represents up to n scalar equations while (47), with n columns in $\mathbf{F}^+(z_i)$, represents $n.\sigma$ scalar equations. Notice that both (44) and (47) are homogeneous; the latter since $\boldsymbol{\alpha}^+(z_i,t)$ is completely expressed in terms of the free parameters $\boldsymbol{\alpha}^q_{I}(t)$. There are $m(\sigma+1)$ unknowns in $\mathbf{w}^+(z,t)$ and $\sigma(n-\rho)$ free parameters in $\boldsymbol{\alpha}^+(z,t)$. Thus the condition governing the window-width is

$$m(\sigma+1) + \sigma(n-\rho) \geq n + n\sigma + 1 \tag{48}$$

yielding

$$\boldsymbol{\sigma} \geq (n - m + 1)/(m - \rho) \tag{49}$$

This is identical with the window-width obtained for the Chow-Willsky scheme, see (18).

The equations being homogeneous, the solution requires the assignment of the parameters which stay free in Step 2. This is a single parameter if (49) is satisfied in equality and more if it is met in inequality.

Computing the residual. The algorithm described above yields the polynomials $\mathbf{w}(z, t)$ and $\boldsymbol{\alpha}(z, t)$. The most straightforward computation of the residual would require the polynomial $\boldsymbol{\beta}(z, t) = \boldsymbol{\alpha}(z, t)\ \mathbf{B}$ but $\mathbf{B}$ is not known in the input-output framework. There are two possible paths :

a) First the primary residual is computed as

$$\mathbf{e}(t) = \mathbf{y}(t) - [\mathbf{G}(z)/h(z)]\boldsymbol{\varphi}(t) \tag{50}$$

then the transformed residual is obtained as

$$r(t) = \mathbf{w}(z, t)\mathbf{e}(t) \tag{51}$$

By this approach, the cancellation of the h(z) polynomial takes place numerically. This poses no problem if the plant is stable but may lead to numerical instability if the plant is unstable.

b) First the $\boldsymbol{\beta}(z, t)$ polynomial is computed by polynomial division from (31), with $\mathbf{G}(z)$ as

$$\boldsymbol{\beta}(z, t) = \boldsymbol{\beta}^1(t)z^{-1}+\ldots+\boldsymbol{\beta}^{\sigma}(t)z^{-\sigma} = \mathbf{w}(z, t)\mathbf{G}(z)/h(z) \tag{52}$$

and then the residual is obtained as

$$r(t) = \mathbf{w}(z, t)\mathbf{y}(t) - \boldsymbol{\beta}(z, t)\boldsymbol{\varphi}(t) \tag{53}$$

This is computationally more demanding than approach a) but is safe, even if the plant is unstable, because h(z) is cancelled algebraically.

7. CONCLUSION

In this paper, we have studied discrete systems with mild nonlinearities, that is, nonlinearities implying the inputs, outputs and faults but not the states. Our objective has been to design structured residuals, ones that are selectively sensitive to subsets of faults and are decoupled from disturbances. For such mildly nonlinear systems, the fault- and disturbance-free residuals are always computed exactly. Further, if the fault and disturbance effect appears linearly, with a nonlinear coefficient, then fault isolation and disturbance decoupling is exact, and the design procedure is linear but needs to be repeated at every sample. If the faults and disturbances appear nonlinearly then, by linearization, the same linear procedure may be applied but decoupling and isolation become approximate.

We first applied the Chow-Willsky scheme (parity space method) to such systems. Then we exposed two important links between the linear part of the Chow-Willsky scheme and that of the input-output representation. Based on these results, we developed an algorithm which generates the residual transformation directly in the input-output framework. The two approaches produce identical residuals. The input-output design is slightly more complex but it does not require the state-space representation of the system.

REFERENCES

Alcorta-Garcia, E. and Frank, P.M. (1997). Deterministic nonlinear observer-based approaches to fault diagnosis : A survey. *Control Engineering Practice,* **5**. pp. 663-670.

Chow, E.Y. and Willsky, A.S. (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Tr. Auto. Contr.* **29***,* pp. 603-614.

Cocquempot, V. and Christophe, C. (2000). On the equivalence between observer based and parity space approaches for FDI in nonlinear systems. *IFAC Safeprocess Symp.* (Budapest, Hungary), pp. 232-237.

Frank, P.M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. *Automatica*, **26***,* pp. 459-474.

Gertler, J. (1998). *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, New York.

Gertler, J. (2000). All linear methods are equal – and extendible to nonlinearities. *IFAC Safeprocess Symp.* (Budapest, Hungary), Preprints **1**, pp. 52-63.

Gertler, J . and Luo, Q. (1989). Robust isolable models for fault diagnosis. *AIChE Journal*, **35,** pp. 1856-1868.

Gertler, J. and Singer, D. (1990). A new structural framework for parity equation based fault detection and isolation. *Automatica,* **26,** pp. 381-388.

Isidori, A . (1995). *Nonlinear Control Systems*. Third Edition. Springer Verlag, New York.

Kailath, T. (1980). *Linear Systems*. Prentice Hall, Englewood Cliffs, NJ.

Kinnaert, M. (1999). Robust fault detection based on observers for bilinear systems. *Automatica*, **35,** pp. 387-404.

Staroswiecki, M., Cassar, J.P. and Cocquempot, V. (1993). Generation of optimal structured residuals in the parity space. *IFAC World Congress* (Sydney, Australia) , **8**, pp. 299-305.

Staroswiecki, M. and Comtet-Varga, G. (2001). Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica,* **37***,* pp.687-699

Willsky, A.S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica***, 12***,* pp. 601-611.