A Data Driven Method for Sensor Fault Detection and Isolation in Bilinear Systems

Komi Midzodzi Pekpe, Cyrille Christophe, Vincent Cocquempot

SENSOR FAULT DETECTION AND ISOLATION WITHOUT EXPLICIT MODEL IN BILINEAR SYSTEMS

K. M. Pekpe, C. Christophe , V. Cocquempot

LAGIS - UMR CNRS 8146
Universite des Sciences et Technologies de Lille U.S.T.L.
Cite Scientifique 59655 Villeneuve d'Ascq Cedex
Phone: +33(0)320434565, Fax: +33(0)320337189

midzodzi.pekpe@univ-lille1.fr,
cyrille.christophe@univ-lille1.fr,
vincent.cocquempot@univ-lille1.fr

http://www-lagis.univ-lille1.fr/

Abstract: A method for multi-sensor Fault Detection and Isolation (FDI) is proposed for bilinear systems whose model structure is known but whose model parameters are not known. Thus, the proposed FDI method uses only available data: control signals and measured outputs and do not need an explicit model formulation. System output variables are expressed on a given time window as functions of input variables and initial unknown state. Under stability hypothesis, the influence of initial state may be neglected compared with inputs contribution. Furthermore, considering on-line outputs aggregation and orthogonal projection w.r.t. inputs, computable *model-free residuals* may be obtained. An academic example is provided to show the efficiency of the proposed method.

Keywords: fault detection and isolation (FDI), bilinear systems, model-free residuals

1. INTRODUCTION

On-line Fault Detection and Isolation (FDI) methods consists in verifying the consistency of the known signals (control signals and measured output) with respect to a behavioral model $\mathcal{M}$ of the system. In the literature, methods are classified with respect to the kind of model that is used. $\mathcal{M}$ may be analytical (state space or input/output models), rules-based, learning-based, statistical or qualitative.

When $\mathcal{M}$ is analytical, the first step aims at generating fault indicator signals, namely residuals. Two approaches are classically used to generate these signals: parity space based approach (A.Y.Chow and A.Willsky, 1984; J.Chen *et al.*, 1995; P.M.Frank and X.Ding., 1997; J.Gertler, 1998) and observer-based approaches (P.M.Frank and X.Ding., 1997; R.Patton and J.Chen, 1997).

These methods consider that the system model $\mathcal{M}$ is entirely known. However in some cases only the structure ($\mathcal{S}_{\mathcal{M}}$) of the model is available and the parameters ($\mathcal{P}_{\mathcal{M}}$) are unknown. In that case, the previous cited methods can not be used. To solve the FDI problem in that situation, two ways may be followed. The first one uses as a first step identification techniques (R.Isermann, 1984; R.Isermann, 1993) to determine parameters

$\mathcal{P}_{\mathcal{M}}$ and applies eventually the previous methods as a second step (S.Simani *et al.*, 2003). The second way try to generate residuals without a pre-determination of the model parameters. This paper is concerned with that kind of method when $\mathcal{S}_{\mathcal{M}}$ is a bilinear structure.

Many practical systems can be represented by bilinear models and this kind of system has been the subject of intensive investigations in all fields of control theory ((D. Yu, 1996),(R.Mohler, 1991)). These models are nonlinear ones but it is often possible to extend the known results obtained in the linear case to this special nonlinear case. In this paper, we extend the FDI method developed in (K.M.Pekpe *et al.*, 2004) for systems with linear structure. This kind of method is called data-driven method since it does not use an explicit model $\mathcal{M}$ but only collected inputs/outputs data.

The paper is organized as follows. In section 2 the Fault Detection and Isolation problem is described and the principle and objectives of our data driven approach are presented. The data driven method for systems of bilinear structure is detailed in sections 3 and 4. A simulation example is provided in section 5 to show the effectiveness of our approach.

2. PROBLEM FORMULATION

2.1 General point of view

Residuals are signals that allow to verify the consistency of the known signals (control signals and measured outputs) with respect to a reference model. These residuals equal zero in no fault situation and becomes different from zero when faults occur. Classical model-based methods, namely parity space based and observer based methods, need the perfect knowledge of the model equations i.e. the structure $\mathcal{S}_{\mathcal{M}}$ and the parameters $\mathcal{P}_{\mathcal{M}}$.

The objective of our approach is to generate residuals by emphasizing a link between the known signals on a given time window without knowing the parameters $\mathcal{P}_{\mathcal{M}}$.

Suppose that the stable system under consideration may be modeled using a discretized model $\mathcal{M}$ under the form:

$$\begin{cases} x_{k+1} &= g(x_k, u_k) \\ y_k^* &= h(x_k, u_k) + w_k + f_k \end{cases} \tag{1}$$

where

- g and h are functions whose structure is known but whose parameters are unknown. It is supposed in that paper that these two functions are linear or bilinear ones.
- u_k is the known control vector of dimension m and y_k^* the measured output vector of dimension l.
- w_k represents output noise with zero mean.
- f_k represents the sensor fault vector:

$$f_k = \begin{pmatrix} f_k^1 & f_k^2 & \dots & f_k^\ell \end{pmatrix}^T \tag{2}$$

Let introduce the following notation:

$$y_k = y_k^* - w_k - f_k \tag{3}$$

y_k may be viewed as the "perfect" output since $y_k = y_k^*$ in no fault situation and without considering output noise.

Consider a time window of size i and let $\bar{u}_{k-i,k}$ represent the vector of the input u on this time window: $\bar{u}_{k-i,k} = \begin{pmatrix} u_{k-i}^T & u_{k-i+1}^T & \dots & u_k^T \end{pmatrix}^T$

The output of the system is a function of $\bar{u}_{k-i,k}$ and initial state x_{k-i}

$$y_k = \mathcal{G}(x_{k-i}, \bar{u}_{k-i,k}). \tag{4}$$

Of course, if model parameters $\mathcal{P}_{\mathcal{M}}$ are unknown, $\mathcal{G}$ can not be explicitly written. However, its structure may be deduced from the structure $\mathcal{S}_{\mathcal{M}}$ that is supposed to be known.

If the system is stable and integer i is sufficiently large, the state influence may be neglected in comparison with input and output noise influences. Therefore, the system output may be written as:

$$\begin{aligned} y_k &= \Gamma(\bar{u}_{k-i,k}) + \delta_{k-i} \\ &\simeq \Gamma(\bar{u}_{k-i,k}) = \textstyle\sum_{j=1}^{r} \alpha_j n_j\left(\bar{u}_{k-i,k}\right) \end{aligned} \tag{5}$$

where

- δ_{k-i} represents the initial state influence on y_k value,
- $n_j\left(\bar{u}_{k-i,k}\right)$ is a monomial function. r is the number of the different monomials which compose $\Gamma(\bar{u}_{k-i,k})$.

Because $\mathcal{S}_{\mathcal{M}}$ is a linear or bilinear structure, equation (5) may be rewritten as

$$\begin{aligned} y_k &\simeq \mathcal{H}\mathcal{N}(\bar{u}_{k-i,k}) \\ \mathcal{H} &= \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_r \end{pmatrix} \end{aligned} \tag{6}$$

where $\mathcal{H} \in \mathbb{R}^{\ell \times r}$ depends on $\mathcal{P}_{\mathcal{M}}$ and $\mathcal{S}_{\mathcal{M}}$ and $\mathcal{N}(\bar{u}_{k-i,k}) \in \mathbb{R}^r$ is a function of input vector which is independent of $\mathcal{P}_{\mathcal{M}}$.

Writing the above equation on a sliding time window of size L (L is an integer which satisfied $L > r$, further information on integer L will be given below) leads to

$$\mathcal{Y}_k \simeq \mathcal{H}\mathcal{U}_k \tag{7}$$

where the output matrix $\mathcal{Y}_k$ and the extended input Hankel matrix $\mathcal{U}_k$ are defined as:

$$\mathcal{Y}_k = \left(y_{k-L+1} \; y_{k-L+2} \; \dots \; y_k \right) \in \mathbb{R}^{\ell \times L}$$

$$\begin{aligned} \mathcal{U}_k = [\mathcal{N}(\bar{u}_{k-i-L+1,k-L+1}) \; \mathcal{N}(\bar{u}_{k-i-L+2,k-L+2}) \\ \dots \qquad \mathcal{N}(\bar{u}_{k-i,k})] \end{aligned} \tag{8}$$

Illustration for linear system structure: *Consider a state space model with linear structure:*

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{cases} \tag{9}$$

The output can be expressed as:

$$y_k = CA^i x_{k-i} + \left[CA^{i-1}B \; \cdots \; CB \;\; D \right] \begin{bmatrix} u_{k-i} \\ \vdots \\ u_k \end{bmatrix} \tag{10}$$

Identifying this expression with equation (6) leads to

$$\begin{aligned} \mathcal{H} &= \left[CA^{i-1}B \; \cdots \; CB \;\; D \right] \\ \mathcal{N}(\bar{u}_{k-i,k}) &= \begin{bmatrix} u_{k-i} \\ \vdots \\ u_k \end{bmatrix} \end{aligned} \tag{11}$$

Finally the input Hankel matrix is obtained

$$\mathcal{U}_k = \begin{pmatrix} u_{k-L-i+1} & u_{k-L-i+2} & \dots & u_{k-i} \\ \vdots & \vdots & \dots & \vdots \\ u_{k-L} & u_{k-L+1} & \dots & u_{k-1} \\ u_{k-L+1} & u_{k-L+2} & \dots & u_k \end{pmatrix} \in \mathbb{R}^{m(i+1) \times L} \tag{12}$$

2.2 Model-free Residual Generation

Consider now the projection on the right kernel of the extended input Hankel matrix $\mathcal{U}_k$. Note Π_k the projection matrix. This matrix is such that

$$\mathcal{U}_k \Pi_k = 0 \tag{13}$$

Right multiplying eq. (7) by Π_k leads to

$$\mathcal{Y}_k \Pi_k = \Delta_k^i \simeq 0 \tag{14}$$

where Δ_k^i corresponds to the contribution of the initial states.

Note that, the dimension of the kernel of $\mathcal{U}_k$ is $L - r$, that implies $L - r \geq \ell$ since ℓ sensor faults have to be detected and isolated.

Because additive noise and faults are considered in the output equation, we have:

$$\mathcal{Y}_k = \mathcal{Y}^*{}_k - \mathcal{F}_k - \mathcal{W}_k \tag{15}$$

where

$$\mathcal{F}_k = \left(f_{k-L+1} \; f_{k-L+2} \; \dots \; f_k \right) \in \mathbb{R}^{\ell \times L} \tag{16}$$

and

$$\mathcal{W}_k = \left(w_{k-L+1} \; w_{k-L+2} \; \dots \; w_k \right) \in \mathbb{R}^{\ell \times L} \tag{17}$$

Define ϵ_k as follows:

$$\epsilon_k = \mathcal{Y}^*{}_k \Pi_k \tag{18}$$

ϵ_k is a computable signal since $\mathcal{Y}^*{}_k$ contains actual outputs of the system.

From eq. (15) and (14), the expression (called evaluation form) of ϵ_k may be obtained:

$$\epsilon_k = \mathcal{F}_k \Pi_k + \mathcal{W}_k \Pi_k + \Delta_k^i \Pi_k \tag{19}$$

For actual system, Π_k may be computed on-line using actual inputs values. ϵ_k may also be deduced considering the actual measured outputs i.e. y_k^*. In normal ideal situation, it means when actual system works exactly as model (eq. 1), with no output noise and no fault, ϵ_k equals zero. However, in practical situation, outputs are corrupted by noise w_k. The mean value, or mathematical expectation, of ϵ_k: $\mathbf{E}[\epsilon_k]$ has thus to be considered. To be a residual signal, $\mathbf{E}[\epsilon_k]$ must equal zero in no fault situation ($f_k = 0$) and must be different from zero when a fault occurs ($f_k \neq 0$). These two properties have been proved for linear systems ((K.M.Pekpe *et al.*, 2004)). This result is extended in the rest of the paper for system with bilinear structure.

3. BILINEAR MODEL STRUCTURE

Consider a bilinear model as follows

$$\begin{aligned} x_{k+1} &= Ax_k + Gx_k \otimes u_k + Bu_k \\ y_k &= Cx_k + Du_k = y_k^* - w_k - f_k \end{aligned} \tag{20}$$

where $\otimes$ represents the Kronecker product.

It is assumed that the system is strongly stable and thus:

- $\|max(eig(A))\| < 1 \Rightarrow$ linear part is stable
- $\|max(eig(G \otimes I))\| < 1 \Rightarrow$ all the bilinear parts are stable

By applying the same methodology as explained previously, equation (7) may be obtained (see appendix for detailed computation). This leads to the following expression of $\mathcal{U}_k$:

$$\mathcal{U}_k = \begin{pmatrix} \widetilde{u}_{k-L+1,i} & \widetilde{u}_{k-L+2,i} & \dots & \widetilde{u}_{k,i} \\ \bar{u}_{k-L+1,i} & \bar{u}_{k-L+2,i} & \dots & \bar{u}_{k,i} \end{pmatrix} \tag{21}$$

where

$$\bar{u}_{k,i} = \left(u_{k-i+1}^T \; u_{k-i+2}^T \; \dots \; u_k^T \right)^T \in \mathbb{R}^{mi} \tag{22}$$

and

$$\widetilde{u}_{k,i} = \begin{pmatrix} u_{k-i} \otimes u_{k-i+1} \otimes ... \otimes u_{k-1} \\ u_{k-i+1} \otimes u_{k-i+2} \otimes ... \otimes u_{k-2} \\ \vdots \\ u_{k-2} \otimes u_{k-1} \end{pmatrix} \in \mathbb{R}^{S_{m,i}} \tag{23}$$

where

$$S_{m,i} = \sum_{j=0}^{i-2} C_i^{i-j} m^{i-j} \tag{24}$$

and where

$$C_n^p = \frac{n!}{(n-p)!p!} \tag{25}$$

Expression of $\mathcal{H}$ may also be obtained but is omitted here since it will not be used in the following developments.

The size of the sliding window L is chosen such that $L > S_{m,i}+mi+\ell$. This choice guarantees that the right orthogonal space Π_k of $\mathcal{U}_k \in \mathbb{R}^{m(2^i-1)\times L}$ exists. The signal ϵ_k can be generated as:

$$\epsilon_k = \mathcal{Y}^*{}_k \Pi_k \tag{26}$$

In order to consider ϵ_k as a residual for FDI, it must be proved that

- The mean value of ϵ_k is close to zero when no sensor fault occurs even if outputs are corrupted by noise w_k.
- The mean value of ϵ_k is different from zero when a sensor fault occurs

The expression of the mean value of ϵ_k is:

$$\mathbf{E}[\epsilon_k] = \mathbf{E}[\mathcal{F}_k\Pi_k] + \mathbf{E}[\mathcal{W}_k\Pi_k] + \mathbf{E}[\Delta_k^i\Pi_k] \tag{27}$$

Because deterministic faults are considered, and Δ_k^i is also a deterministic vector we have

$$\mathbf{E}[\epsilon_k] = \mathcal{F}_k\Pi_k + \mathbf{E}[\mathcal{W}_k]\Pi_k + \Delta_k^i\Pi_k \tag{28}$$

Since output noise is of zero mean, $\mathbf{E}[\mathcal{W}_k] = 0$
On the other hand, $\Delta_k^i\Pi_k$ represents the contribution of the initial state vector, i.e. the neglected dynamics. From a theoretical point of view, it is not equal to zero. However, as it will be discussed in the next section, its value may be made very small with an appropriate choice of the time windows sizes. In the no fault situation, the mean value of ϵ_k is thus close to zero and when a fault occurs, we have:

$$\mathbf{E}[\epsilon_k] \simeq \mathcal{F}_k\Pi_k. \tag{29}$$

$\mathbf{E}[\epsilon_k]$ will be sensitive to sensor faults if the following condition holds

$$span(\mathcal{F}_k) \not\subset span(\Pi_k) \Rightarrow \mathcal{F}_k\Pi_k \neq 0 \tag{30}$$

This condition is not restrictive and corresponds to a very special case. Under this hypothesis on the fault vector, ϵ_k may be considered as a model-free residual for FDI.

4. FAULT DETECTION AND ISOLATION

In this section, it is shown that the model-free residual ϵ_k is already structured with respect to sensor faults. Fault isolation is thus directly performed. The value of ϵ_k depends on the sizes i and L of the two time windows that are needed to compute this value. The decision threshold has to be chosen w.r.t. these sizes. This choice is also discussed in the second subsection.

4.1 Sensor fault isolation

It has been proved in the previous section that the mean value of ϵ_k may be approximated as:

$$\mathbf{E}[\epsilon_k] \simeq \mathcal{F}_k\Pi_k. \tag{31}$$

ϵ_k is a vector of size l: $\epsilon_k = \begin{pmatrix} \epsilon_k^1 \\ \epsilon_k^2 \\ \vdots \\ \epsilon_k^\ell \end{pmatrix}$.

From the definition of $\mathcal{F}_k$, it is evident that each component $\epsilon_k^j, \forall j \in 1, 2, ..l$ depends only on the j^th component of the fault vector f_k. As a consequence, the residual vector ϵ_k is structured w.r.t. the sensor faults and these faults may be directly isolated.

4.2 Computation windows determination

Two time windows are used to compute a residual vector ϵ_k. The first one, named $\bar{I}$, of size i, allows to express each output at time k in function of the inputs on this time window and the initial state i.e. x_{k-i}. The size i must be sufficiently large to be able to neglect the state influence. The second window of size L allows to construct the extended input Hankel matrix. L must be chosen to guarantee the existence of the projection matrix Π_k. Of course, L is connected to i and must be as small as possible in order to guaranty not only a minimal detection delay but also an acceptable computational complexity.

The choice of an optimal size i may be done using a set of output data y_k^* collected in a no fault situation. Consider the following criterion:

$$J(i) = \|\mathcal{Y}_k^*\Pi_k\| \tag{32}$$

This criterion is also equal to

$$J(i) = \left\|\Delta_k^i\Pi_k + \mathcal{W}\Pi_k\right\| \tag{33}$$

The general evolution of this criterion is plotted on figure (1). $J(i)$ is of course a decreasing function of i but two parts, separated by a value i_0

may be distinguished: when i is small ($i < i_0$), the value of $J(i)$ is very large and its value is highly dependent on i. When $i > i_0$, the value of $J(i)$ is small and do not evolve significantly w.r.t. i.

From the figure that may be drawn using actual data in no fault situation, a correct choice i may be done, $i \geq i_0$.

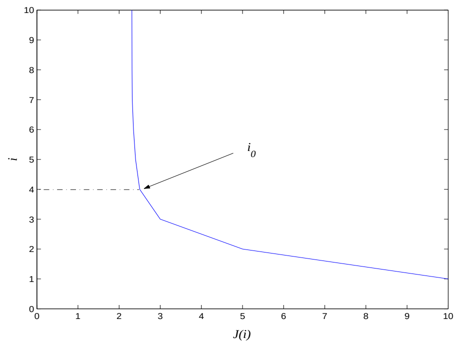

Fig. 1. Illustration of the criterion $J(i)$

5. ILLUSTRATIVE EXAMPLE

Consider the following bilinear system:

$$A = \begin{pmatrix} 0.2 & -0.3 & 0.3 & 0 \\ 0 & 0.2 & -0.3 & 0 \\ -1 & 0 & 0 & 0.4 \\ 0 & 0 & 0.3 & 0.4 \end{pmatrix}, \; B = \begin{pmatrix} 0.45 & 0.5 \\ 0.3 & 0.55 \\ 0.2 & 0.6 \\ 0.4 & 0.3 \end{pmatrix}$$

$$\bar{G} = \begin{pmatrix} 0.3 & 0.4 \\ -0.4 & 0.1 \\ 0.35 & 0.45 \\ 0.5 & -0.2 \end{pmatrix}, \; G = \left(\bar{G} \; \bar{G} \; \bar{G} \; \bar{G} \right)$$

$$C = \begin{pmatrix} 0.5 & 0.35 & 0.3 & 0.4 \\ 0.8 & 0.2 & 0.26 & 0.45 \\ 0.37 & 0.5 & 0.4 & 0.6 \end{pmatrix}, \; D = \begin{pmatrix} 0.35 & 0.85 \\ 0.65 & 0.75 \\ 0.45 & 0.69 \end{pmatrix}$$

The system input u_k is a Pseudo-Random Binary Sequence (PRBS). The signal to noise ratio of the outputs with respect to the measurement noises is 20dB. A process noise v_k has also been added with a covariance $var(v_k) \simeq 7 \times 10^{-4} I_3$. The system inputs are plotted in figure (2) and the system outputs in figure (3).
The size i of the time window $\bar{I}$ is determined by drawing criterion $J(i)$ in healthy situation like in figure (1). The figure shows that i_0 may be chosen equal to 3 since the criterion does not decrease noticeably when i is greater. Integer i must be chosen greater or equal to i_0, here this integer is fixed at 4. The size of the sliding window L is 160 ; ($L > S_{m,i} + mi = 80$, see (24)).

The residuals and the fault are plotted in figure (4) to (6). All faults are clearly detected using these residuals.

6. CONCLUSION

A sensor fault detection for bilinear systems has been proposed in the paper. The method is based only on the knowledge of the input-output data and the model structure. Model parameters are not required. An academic example was provided to illustrate the performance of the method. Future works will focus on process faults and more general nonlinear model structures.

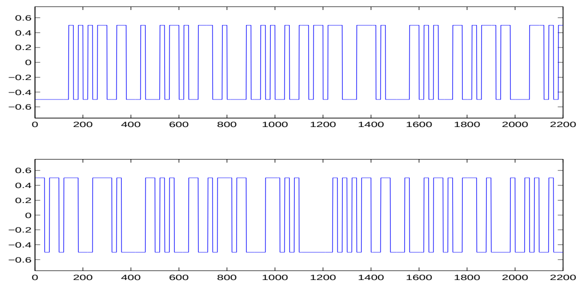

Fig. 2. System input

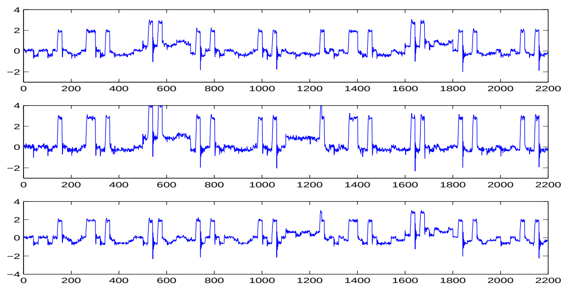

Fig. 3. System output

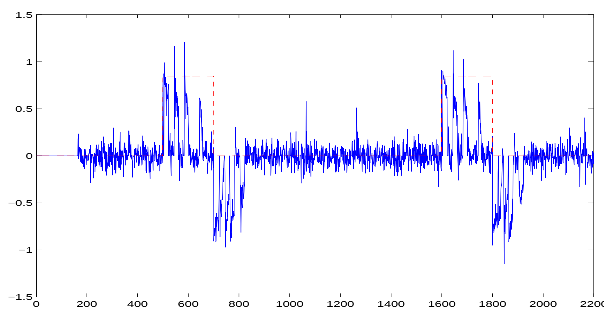

Fig. 4. First row residual ϵ_k^1 and fault occurrence (solid: residual, dashed: fault signal)

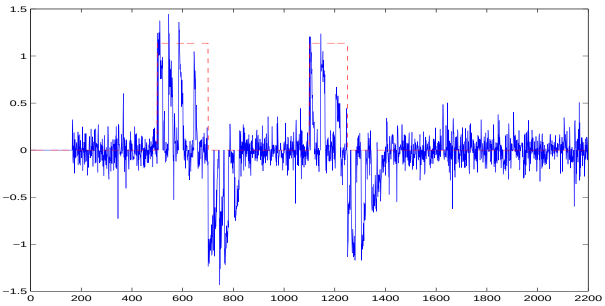

Fig. 5. Second row residual ϵ_k^2 and fault occurrence (solid: residual, dashed: fault signal)

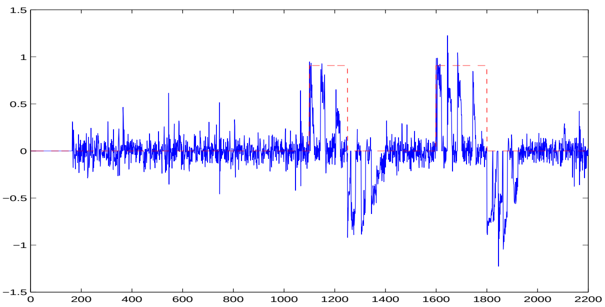

Fig. 6. Third row residual ϵ_k^3 and fault occurrence (solid: residual, dashed: fault signal)

REFERENCES

A.Y.Chow and A.Willsky (1984). Analyticaly redundancy and the design of robust failure detection systems. *IEEE Transactions on Automatic Control* **7**(29), 603–614.

D. Yu, D.N. Shields (1996). *Automatica* **32**(11), 1597–1602.

J.Chen, R.Patton and H.Zhang (1995). Design of robust structured and directional residuals for fault isolation via unknown input observers. European Control Conference. pp. 348–353. Rome, Italy.

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

K.M.Pekpe, G.Mourot and J.Ragot (2004). Subspace method for sensor fault detection and isolation-application to grinding circuit monitoring. 11th IFAC Symposium on automation in Mining, Mineral and Metal processing. Nancy, France.

P.M.Frank and X.Ding. (1997). Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control* **7**(6), 403–424.

R.Isermann (1984). Process fault detection based on modelling and estimation methods: a survey. *Automatica* **20**(4), 387404.

R.Isermann (1993). Fault diagnosis of machines via parameter estimation and knowledge processing - tutorial paper. *Automatica* **29**(4), 815–836.

R.Mohler (1991). *Nonlinear Systems: Applications to Bilinear Control*. Prentice Hall.

R.Patton and J.Chen (1997). Observer-based fault detection and isolation: Robustness and applications. *Control Engineering Practice* **5**(5), 671–682.

S.Simani, C.Fantuzzi and R.J.Patton (2003). *Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. Springer-Verlag.

7. APPENDIX

Here relation (5) is established for bilinear systems (20). By using model equations (20) the system output can be expressed as:

$$y_k = CAx_{k-1} + CGx_{k-1} \otimes u_{k-1} + CBu_{k-1} + Du_k \tag{34}$$

Replacing x_{k-1} in function of x_{k-2} leads to

$$\begin{aligned} y_k = & CA^2x_{k-2} + CAGx_{k-2} \otimes u_{k-2} + CGA \otimes Ix_{k-2} \otimes u_{k-1} + CG \otimes IGx_{k-2} \otimes u_{k-2} \otimes u_{k-1} \\ & +CGB \otimes Iu_{k-2} \otimes u_{k-1} + CABu_{k-2} + CBu_{k-1} + Du_k \end{aligned} \tag{35}$$

where I is the identity matrix of size n.

Replacing x_{k-2} in function of x_{k-3} leads to

$$\begin{aligned} y_k = & CA^3x_{k-3} + CA^2Gx_{k-3} \otimes u_{k-3} + CAGA \otimes Ix_{k-3} \otimes u_{k-2} + CAGG \otimes Ix_{k-3} \otimes u_{k-3} \otimes u_{k-2} \\ & +CGA \otimes IA \otimes Ix_{k-3} \otimes u_{k-1} + CGA \otimes IG \otimes Ix_{k-3} \otimes u_{k-3} \otimes u_{k-1} \\ & +CG \otimes IGA \otimes Ix_{k-3} \otimes u_{k-2} \otimes u_{k-1} + CG \otimes IG \otimes IGx_{k-3} \otimes u_{k-3} \otimes u_{k-2} \otimes u_{k-1} \\ & +CG \otimes IGB \otimes Iu_{k-3} \otimes u_{k-2} \otimes u_{k-1} + CGA \otimes IB \otimes Iu_{k-3} \otimes u_{k-1} + CAGB \otimes Iu_{k-3} \otimes u_{k-2} \\ & +CGB \otimes Iu_{k-2} \otimes u_{k-1} + CA^2Bu_{k-3} + CABu_{k-2} + CBu_{k-1} + Du_k \end{aligned} \tag{36}$$

Replacing x_{k-3} in function of x_{k-4} leads to

$$\begin{aligned} y_k = & CA^4x_{k-4} + CA^3Gx_{k-4} \otimes u_{k-4} + CA^2GA \otimes Ix_{k-4} \otimes u_{k-3} + CA^2GG \otimes Ix_{k-4} \otimes u_{k-4} \otimes u_{k-3} \\ & +CAGA \otimes IA \otimes Ix_{k-4} \otimes u_{k-2} + CAGA \otimes IG \otimes Ix_{k-4} \otimes u_{k-4} \otimes u_{k-2} + CAGG \otimes IA \otimes Ix_{k-4} \otimes u_{k-3} \\ & +CAGG \otimes IG \otimes Ix_{k-4} \otimes u_{k-4} \otimes u_{k-3} + CGA \otimes IA \otimes IA \otimes Ix_{k-4} \otimes u_{k-1} \\ & +CGA \otimes IA \otimes IG \otimes Ix_{k-4} \otimes u_{k-4} \otimes u_{k-1} + CGA \otimes IG \otimes IA \otimes Ix_{k-4} \otimes u_{k-3} \otimes u_{k-1} \\ & +CGA \otimes IG \otimes IG \otimes Ix_{k-4} \otimes u_{k-4} \otimes u_{k-3} \otimes u_{k-1} + CG \otimes IGA \otimes IA \otimes Ix_{k-4} \otimes u_{k-2} \otimes u_{k-1} \\ & +CG \otimes IGA \otimes IG \otimes Ix_{k-4} \otimes u_{k-4} \otimes u_{k-2} \otimes u_{k-1} + CG \otimes IG \otimes IGA \otimes Ix_{k-4} \otimes u_{k-3} \otimes u_{k-2} \otimes u_{k-1} \\ & +CG \otimes IG \otimes IGG \otimes Ix_{k-4} \otimes u_{k-4} \otimes u_{k-3} \otimes u_{k-2} \otimes u_{k-1} \\ & +CG \otimes IG \otimes IGB \otimes Iu_{k-4} \otimes u_{k-3} \otimes u_{k-2} \otimes u_{k-1} + CG \otimes IGA \otimes IB \otimes Iu_{k-4} \otimes u_{k-2} \otimes u_{k-1} \\ & +CGA \otimes IG \otimes IB \otimes Iu_{k-4} \otimes u_{k-3} \otimes u_{k-1} + CAGG \otimes IB \otimes Iu_{k-4} \otimes u_{k-3} \otimes u_{k-2} \\ & +CG \otimes IGB \otimes Iu_{k-3} \otimes u_{k-2} \otimes u_{k-1} + CGA \otimes IA \otimes IB \otimes Iu_{k-4} \otimes u_{k-1} + CAGA \otimes IB \otimes Iu_{k-4} \otimes u_{k-2} \\ & +CA^2GB \otimes Iu_{k-4} \otimes u_{k-3} + CGA \otimes IB \otimes Iu_{k-3} \otimes u_{k-1} + CAGB \otimes Iu_{k-3} \otimes u_{k-2} \\ & +CGB \otimes Iu_{k-2} \otimes u_{k-1} + CA^3Bu_{k-4} + CA^2Bu_{k-3} + CABu_{k-2} + CBu_{k-1} + Du_k \end{aligned} \tag{37}$$

Continuing this procedure up to order i, leads to the system output expression:

$$\begin{aligned} y_k = & CA^ix_{k-i} + C\sum_{\alpha=1}^{i} q(A^\alpha, G, (A \otimes I)^\gamma, (G \otimes I)^{i-1-\alpha-\gamma}, x_{k-i}, P(u^{i-1-\alpha-\gamma})) \\ & + \sum_{s=2}^{i} \sum_{i \geq r_s > \ldots > r_1 \geq 1} \rho_{r_s,\ldots,r_1} u_{k-r_s} \otimes \ldots \otimes u_{k-r_1} + \textstyle\sum_{s=0}^{i} h_{s-1}u_{k-s} \end{aligned} \tag{38}$$

$$\Delta_k^i = CA^ix_{k-i} + C\sum_{\alpha=1}^{i-1} \sum_{\gamma < i-\alpha} q(A^\alpha, G, (A \otimes I)^\gamma, (G \otimes I)^{i-1-\alpha-\gamma}, x_{k-i}, P(u^{i-1-\alpha-\gamma})) \tag{39}$$

where $q(A^\alpha, G, (A \otimes I)^\gamma, (G \otimes I)^{i-2-\alpha-\gamma}, x_{k-i}, P(u^{i-2-\alpha-\gamma}))$ is a product of A^α, G, $(A \otimes I)^\gamma$, $(G \otimes I)^{i-2-\alpha-\gamma}$, x_{k-i}, and a Kronecker product of $i-2-\alpha-\gamma$ inputs.

By using the triangular inequality one can obtain:

$$\begin{aligned} & \left\| q(A^\alpha, G, (A \otimes I)^\gamma, (G \otimes I)^{i-1-\alpha-\gamma}, x_{k-i}, P(u^{i-1-\alpha-\gamma})) \right\| \leq \\ & \|A\|^\alpha \|G\| \|A \otimes I\|^\gamma \|G \otimes I\|^{i-1-\alpha-\gamma} \|max(x)\| \underbrace{\|max(u) \otimes max(u) \otimes \ldots \otimes max(u)\|}_{i-2-\alpha-\gamma \text{ Kronecker product}} \end{aligned} \tag{40}$$

where $max(x) = max(x_k)$ and $max(u) = max(|u_k|)$.

$$\begin{aligned} & \left\| q(A^\alpha, G, (A \otimes I)^\gamma, (G \otimes I)^{i-2-\alpha-\gamma}, x_{k-i}, P(u^{i-2-\alpha-\gamma})) \right\| \leq \\ & \|M\|^{i-1} \|G\| \underbrace{\|max(u) \otimes max(u) \otimes \ldots \otimes max(u)\|}_{i-2-\alpha-\gamma \text{ Kronecker product}} \end{aligned} \tag{41}$$

where

$$\|M\| = max(\|A\|, \|A \otimes I\|, \|G \otimes I\|) \tag{42}$$

Since A is stable, $A \otimes I = \begin{pmatrix} A & 0 & ... & 0 \\ 0 & A & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & A \end{pmatrix}$ is also stable (A and $A \otimes I =$ have the same eigenvalues). Then, Δ_k^i tends to zero when integer i increases. Therefore, for a suitable choice of integer i equation (38) becomes:

$$y_k = \sum_{s=2}^{i} \sum_{i \geq r_s > ... > r_1 \geq 1} \rho_{r_s,...,r_1} u_{k-r_s} \otimes ... \otimes u_{k-r_1} + \sum_{s=0}^{i} h_{s-1} u_{k-s} \tag{43}$$

$$y_k = \widetilde{\mathcal{H}} \widetilde{u}_{k,i} + \bar{\mathcal{H}} \bar{u}_{k,i} \tag{44}$$

with

$$\widetilde{\mathcal{H}} = \left(\rho_{r_i,...,r_1} \; \rho_{r_{i-1},...,r_1} \; ... \; \rho_{r_1} \right), \bar{\mathcal{H}} = \left(h_{i-1} \; h_{i-2} \; ... \; h_{-1} \right) \tag{45}$$

and where $\widetilde{u}_{k,i}$ and $\bar{u}_{k,i}$ are respectively defined in (23) and (22). The above equation can be rewritten as:

$$y_k = \underbrace{\left(\widetilde{\mathcal{H}} \; \bar{\mathcal{H}} \right)}_{\mathcal{H}} \begin{pmatrix} \widetilde{u}_{k,i} \\ \bar{u}_{k,i} \end{pmatrix} \tag{46}$$

Finally, this equation can be stacked on time window of size L, which leads to

$$\mathcal{Y}_k = \mathcal{H} \begin{pmatrix} \widetilde{u}_{k-L+1,i} & \widetilde{u}_{k-L+2,i} & ... & \widetilde{u}_{k,i} \\ \bar{u}_{k-L+1,i} & \bar{u}_{k-L+2,i} & ... & \bar{u}_{k,i} \end{pmatrix} \tag{47}$$

where $\mathcal{Y}_k$ is defined by (8).