

INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL

*Int. J. Robust Nonlinear Control* 2007; **17**:1294–1319

Published online 14 February 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/rnc.1170

## New results in robust actuator fault reconstruction for linear uncertain systems using sliding mode observers

Kok Yew Ng<sup>1</sup>, Chee Pin Tan<sup>1,\*†</sup>, Christopher Edwards<sup>2</sup> and Ye Chow Kuang<sup>1</sup><sup>1</sup>*School of Engineering, Monash University Malaysia, 2 Jalan Kolej, Petaling Jaya 46150, Malaysia*<sup>2</sup>*Engineering Department, Leicester University, University Road, Leicester LE1 7RH, U.K.*

### SUMMARY

This paper presents a robust actuator fault reconstruction scheme for linear uncertain systems using sliding mode observers. In existing work, fault reconstruction *via* sliding mode observers is limited to either linear certain systems subject to unknown inputs, relative degree one systems or a specific class of relative degree two systems. This paper presents a new method that is applicable to a wider class of systems with relative degree higher than one, and can also be used for systems with more unknown inputs than outputs. The method uses two sliding mode observers in cascade. Signals from the first observer are processed and used to drive the second observer. Overall, this results in actuator fault reconstruction being feasible for a wider class of systems than using existing methods. A simulation example verifies the claims made in this paper. Copyright © 2007 John Wiley & Sons, Ltd.

Received 24 May 2006; Revised 9 August 2006; Accepted 29 November 2006

KEY WORDS: robustness; fault reconstruction; sliding mode observer

### 1. INTRODUCTION

Fault detection and isolation (FDI) is an important area of research activity. A fault is deemed to occur when the system being monitored is subject to an abnormal condition, such as a malfunction [1]. The fundamental purpose of an FDI scheme is to generate an alarm when a fault occurs (detection) and also to identify the nature and location of the fault (isolation). Survey papers that give overviews of work in this area are available in [2–4]. The most commonly used FDI methods are observer based where the measured plant output is compared to the output of an observer designed from a model of the system, and the discrepancy is used to form a residual [5–7]. Using this residual signal, a decision is made

---

\*Correspondence to: Chee Pin Tan, School of Engineering, Monash University Malaysia, 2 Jalan Kolej, Petaling Jaya 46150, Malaysia.

†E-mail: [tan.chee.pin@eng.monash.edu.my](mailto:tan.chee.pin@eng.monash.edu.my)

as to whether a fault condition is present and also an attempt is made to determine its location.

A useful alternative to residual generation is *fault reconstruction* [8–10], which not only detects and isolates the fault, but also provides an estimate of the fault so that its shape and magnitude can be better understood and more precise corrective action can be taken. However, a fault reconstruction scheme is usually designed about a model of the system. This model usually does not perfectly represent the system, as certain dynamics are either unknown or do not fit exactly into the framework of the model. These dynamics are usually represented as a class of disturbances within the model [11]. The disturbances corrupt the reconstruction, and could produce a non-zero reconstruction when there are no faults, or worse, mask the effect of a fault, producing a ‘zero’ reconstruction in the presence of faults. Therefore, the scheme needs to be designed so that the reconstruction is robust to disturbances.

Edwards *et al.* [8, 9] used a sliding mode observer [12] to reconstruct faults, but there was no explicit consideration of the disturbances. Tan and Edwards [13] built on the work in [8, 9] and presented a design algorithm for the observer, using linear matrix inequalities (LMIs) [14], such that the  $\mathcal{L}_2$  gain from the disturbances to the fault reconstruction is minimized. Saif and Guan [10] aggregated the faults and disturbances to form a new ‘fault’ vector and used a linear observer to reconstruct the new ‘fault’ vector. One of the necessary conditions in [8–10, 13] is that the transfer function from the faults to the output has a relative degree of one. This limits the class of systems where the schemes [8–10, 13] are applicable.

Recently, there have been developments in the area of fault reconstruction for systems with relative degree greater than one. Floquet and Barbot [15] transformed the system into an ‘output information’ form such that existing sliding mode observer techniques could be implemented to perfectly estimate the states in finite time and reconstruct faults. However, their algorithm does not consider disturbances (unless as in [10] the unknown inputs (faults) are augmented with the disturbances). Furthermore, the class of systems for which the transformation is feasible is not known, and it is not easy for the designer to immediately recognize whether the algorithm is suitable for the system under consideration. Davila *et al.* [16] developed a second-order sliding mode observer for nonlinear mechanical systems, i.e. second-order differential equations, arising from Newton’s laws where only position (and not velocity) is measured. The work in [16] could be easily extended to the case of robust fault reconstruction for actuator faults occurring in the acceleration equation. However, it is applicable only to a limited class of systems as it requires that *all* position signals are measurable.

This paper presents a robust fault reconstruction method for a class of systems whose relative degree with respect to the fault is higher than one, relaxing the condition required by the previous work [8, 9, 13]. The method in this paper essentially uses two sliding mode observers [12] in cascade. Suitable processing of the equivalent output error injection in the first observer yields the measurable output of a ‘fictitious’ system that is relative degree one. This means the robust fault reconstruction method in [13] is applicable to the fictitious system and a second observer is implemented on the fictitious system to generate a reconstruction of the fault that is robust to the disturbances. This approach is applicable to a wider class of systems for which the methods in [8, 9, 13] are not applicable. Furthermore, this paper considers robustness against disturbances (as opposed to the method in [15]), and the scheme may be feasible for systems for which the method in [16] is not applicable.

This paper is organized as follows: Section 2 introduces the system and states the main result, whilst Section 3 sets up the framework for the proposed method together with existence conditions. An example to demonstrate the effectiveness of the scheme is given in Section 4 and finally Section 5 makes some conclusions. The notation used throughout this paper is quite standard; in particular,  $\|\cdot\|$  represents the Euclidean norm for vectors and the induced spectral norm for matrices, whilst  $\lambda(\cdot)$  denotes the spectrum of a square matrix.

## 2. PRELIMINARIES AND STATEMENT OF THE MAIN RESULT

Consider a system

$$\dot{\tilde{x}}(t) = \tilde{A}\tilde{x}(t) + \tilde{B}u(t) + \tilde{M}f(t) + \tilde{Q}\tilde{\xi}(t) \quad (1)$$

$$y(t) = \tilde{C}\tilde{x}(t) \quad (2)$$

where  $\tilde{x} \in \mathbb{R}^{\tilde{n}}$ ,  $y \in \mathbb{R}^p$ ,  $u \in \mathbb{R}^m$  are the states, outputs and inputs respectively, with  $\tilde{n} \geq p$ . The vector  $f \in \mathbb{R}^q$  is an unknown fault and  $\tilde{\xi} \in \mathbb{R}^h$  is an unknown disturbance, which encapsulates all nonlinearities and unknowns in the system [1]. Assume without loss of generality that  $\text{rank}(\tilde{M}) = q$ ,  $\text{rank}(\tilde{Q}) = h$ ,  $\text{rank}(\tilde{C}) = p$  and suppose that  $\text{rank}(\tilde{C}\tilde{M}) = r < q < p$ . Also assume that  $(\tilde{A}, \tilde{C})$  is observable.

The objective is to reconstruct the fault  $f$  whilst being robust to  $\tilde{\xi}$ . Edwards *et al.* [8, 9] have reconstructed the fault  $f$  for the case when  $\tilde{\xi} = 0$ . Tan and Edwards [13] built on this early work and presented a method that minimizes the  $\mathcal{L}_2$  gain from  $\tilde{\xi}$  to the fault reconstruction. In [8, 9, 13], the fault reconstruction scheme is feasible if and only if the following conditions are satisfied:

- A1.  $\text{rank}(\tilde{C}\tilde{M}) = \text{rank}(\tilde{M}) = q$ .
- A2. The invariant zeros of  $(\tilde{A}, \tilde{M}, \tilde{C})$  (if any) are stable.

Condition A1 implies that the system is relative degree one, and condition A2 implies that the system is minimum phase. These conditions are also often assumed if unknown input observers (UIOs) are employed for fault reconstruction [10, 17]. This paper proposes a method to robustly reconstruct the fault when condition A1 is not satisfied.

Assume that the disturbance  $\tilde{\xi}$  is piecewise continuous [10] such that

$$\dot{\tilde{\xi}}(t) = A_\Omega \tilde{\xi}(t) + B_\Omega \zeta(t) \quad (3)$$

where  $\zeta \in \mathbb{R}^h$  and  $A_\Omega \in \mathbb{R}^{h \times h}$  is stable and  $B_\Omega \in \mathbb{R}^{h \times h}$ . This is not an unreasonable assumption provided the frequency content of  $\tilde{\xi}$  is known. If  $\tilde{\xi}$  is known to be a signal in the frequency region  $\omega_1 < \omega < \omega_2$ , then system (3) can be taken to be first-order filters with cut-off frequency  $\omega_2$ .

### Theorem 1

For the case when A1 is not satisfied, i.e.  $r = \text{rank}(\tilde{C}\tilde{M}) < \text{rank}(\tilde{M}) = q$ , then the fault  $f$  can be reconstructed by the scheme given in Figure 1 such that the  $\mathcal{L}_2$  gain from  $\zeta$  to the fault

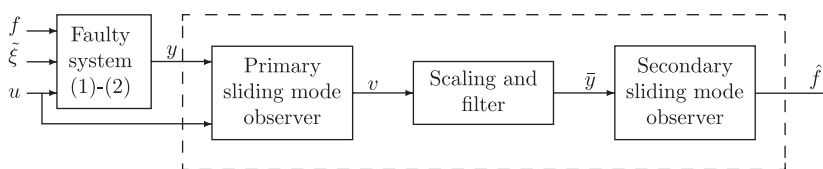


Figure 1. Schematic diagram of the scheme proposed in this paper.

reconstruction will be bounded if and only if

B1. 
$$\text{rank} \begin{bmatrix} \tilde{C}\tilde{A}\tilde{M} & \tilde{C}\tilde{M} \\ \tilde{C}\tilde{M} & 0 \end{bmatrix} = \text{rank}(\tilde{C}\tilde{M}) + \text{rank}(\tilde{M})$$

B2. The invariant zeros of  $(\tilde{A}, \tilde{M}, \tilde{C})$  (if any) must be stable.

In the scheme in Figure 1  $\hat{f}$  is the reconstruction of the fault  $f$ , whilst  $v$  and  $\bar{y}$  are intermediate signals that will be defined later.

Comparing Conditions B1 and A1, it is clear that B1 is less restrictive than A1, i.e. if A1 is satisfied then B1 will also be satisfied, but the converse is not necessarily true. In the case when A1 is satisfied, the method proposed in this paper is also applicable since  $\hat{f}$  can be calculated directly from the output of the primary observer in Figure 1. In practice, however, the simpler scheme in [13] would be used in preference.

The next section will provide a constructive proof of Theorem 1.

### 3. ROBUST FAULT RECONSTRUCTION

Firstly, four lemmas will be introduced to provide a canonical form which underpins the scheme that will be developed. The first is concerned with the system in (1)–(2) and imposes specific structures on the output and fault distribution matrices.

*Lemma 1*

There exist appropriately dimensioned non-singular linear transformations  $\tilde{x} \mapsto T_1\tilde{x}$ ,  $f \mapsto T_2f$  such that the triple  $\tilde{A}, \tilde{M}, \tilde{C}$  from (1) to (2) in the new co-ordinates are given by

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 \\ \tilde{A}_3 & \tilde{A}_4 \end{bmatrix}, \quad \tilde{C} = [0 \quad \tilde{T}], \quad \tilde{M} = \begin{bmatrix} \tilde{M}_1 \\ \tilde{M}_2 \end{bmatrix} \tag{4}$$

where  $\tilde{A}_1 \in \mathbb{R}^{(\tilde{n}-p) \times (\tilde{n}-p)}$ ,  $\tilde{M}_2 \in \mathbb{R}^{p \times q}$  and  $\tilde{T} \in \mathbb{R}^{p \times p}$  is orthogonal. Furthermore, the matrices  $\tilde{M}_1, \tilde{M}_2$  can be partitioned to have the form

$$\tilde{M}_1 = \begin{bmatrix} 0 & 0 \\ M_{11} & 0 \end{bmatrix} \begin{matrix} \overset{q-r}{\leftrightarrow} & \leftrightarrow \\ \downarrow \tilde{n} - p - q + r \\ \downarrow q - r \end{matrix}, \quad \tilde{M}_2 = \begin{bmatrix} 0 & 0 \\ 0 & M_{22} \end{bmatrix} \begin{matrix} \overset{q-r}{\leftrightarrow} & \leftrightarrow \\ \downarrow p - r \\ \downarrow r \end{matrix} \tag{5}$$

where  $M_{11}, M_{22}$  are invertible. In this co-ordinate system,  $f \mapsto T_2f = \text{col}(f_1, f_2)$  where  $f_2 \in \mathbb{R}^r$ .

*Proof*

See Section A.1 in the Appendix.

In the co-ordinate system of (4)–(5), further partition  $\tilde{A}$  and  $\tilde{Q}$  (generally) as

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 \\ \tilde{A}_3 & \tilde{A}_4 \end{bmatrix} = \left[ \begin{array}{cc|cc} \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{13} & \tilde{A}_{14} \\ \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} & \tilde{A}_{24} \\ \hline \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} & \tilde{A}_{34} \\ \tilde{A}_{41} & \tilde{A}_{42} & \tilde{A}_{43} & \tilde{A}_{44} \end{array} \right], \quad \tilde{Q} = \begin{bmatrix} \tilde{Q}_{11} \\ \tilde{Q}_{12} \\ \tilde{Q}_{21} \\ \tilde{Q}_{22} \end{bmatrix} \begin{array}{l} \Downarrow \tilde{n} - p - q + r \\ \Downarrow q - r \\ \Downarrow p - r \\ \Downarrow r \end{array} \quad (6)$$

Now the system equations (1)–(2) and the uncertainty/disturbance model from (3) will be augmented to form the system which will be studied in the remainder of the paper. Combine (1)–(2) and (3) to obtain the following augmented system of order  $n := \tilde{n} + h$ :

$$\underbrace{\begin{bmatrix} \dot{\xi} \\ \dot{x} \end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} A_\Omega & 0 \\ \tilde{Q} & \tilde{A} \end{bmatrix}}_A \underbrace{\begin{bmatrix} \xi \\ x \end{bmatrix}}_x + \underbrace{\begin{bmatrix} 0 \\ \tilde{B} \end{bmatrix}}_B u + \underbrace{\begin{bmatrix} 0 \\ \tilde{M} \end{bmatrix}}_M f + \underbrace{\begin{bmatrix} B_\Omega \\ 0 \end{bmatrix}}_Q \xi \quad (7)$$

$$y = \underbrace{\begin{bmatrix} 0 & \tilde{C} \end{bmatrix}}_C \underbrace{\begin{bmatrix} \xi \\ x \end{bmatrix}}_x \quad (8)$$

Now expand the matrices in (7)–(8) as in (4)–(5) to obtain

$$A = \left[ \begin{array}{ccc|cc} A_\Omega & 0 & 0 & 0 & 0 \\ \tilde{Q}_{11} & \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{13} & \tilde{A}_{14} \\ \tilde{Q}_{12} & \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} & \tilde{A}_{24} \\ \hline \tilde{Q}_{21} & \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} & \tilde{A}_{34} \\ \tilde{Q}_{22} & \tilde{A}_{41} & \tilde{A}_{42} & \tilde{A}_{43} & \tilde{A}_{44} \end{array} \right] \begin{array}{l} \Downarrow h \\ \Downarrow n - p - q + r - h \\ \Downarrow q - r \\ \Downarrow p - r \\ \Downarrow r \end{array} \quad (9)$$

$$M = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \hline 0 & 0 \\ 0 & M_{22} \end{bmatrix}, \quad Q = \begin{bmatrix} B_\Omega \\ 0 \\ \hline 0 \\ 0 \end{bmatrix}, \quad C = [0 \quad \tilde{T}] \quad (10)$$

*Lemma 2*

The augmented pair  $(A, C)$  from (7) to (8) is detectable.

*Proof*

See Section A.2 in the Appendix.

*Lemma 3*

Condition B1 from the statement of Theorem 1 is satisfied if and only if  $\tilde{A}_{32}$  from (6) has full column rank  $q - r$ .

*Proof*

See Section A.3 in the Appendix.

Define  $\bar{p} := \text{rank}[\tilde{Q}_{21} \tilde{A}_{31} \tilde{A}_{32}] + r$ . It follows that  $\bar{p} - r \leq \min\{p - r, n - p\}$  and therefore  $\bar{p} \leq p$ . Since condition B1 implies that  $\tilde{A}_{32}$  has full column rank, then  $\bar{p} - r \geq q - r$  which implies that  $\bar{p} \geq q$ .

*Lemma 4*

There exists a non-singular linear change of co-ordinates such that  $x \rightarrow T_5x$  and the matrices  $A, M, Q, C$  from (9) to (10) when partitioned have the structure

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} = \left[ \begin{array}{ccc|cc} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ A_{31} & A_{32} & A_{33} & A_{34} & A_{35} \\ \hline 0 & A_{42} & A_{43} & A_{44} & A_{45} \\ A_{51} & A_{52} & A_{53} & A_{54} & A_{55} \end{array} \right] \begin{array}{l} \Downarrow n - \bar{p} + r - p \\ \Downarrow \bar{p} - q \\ \Downarrow q - r \\ \Downarrow p - r \\ \Downarrow r \end{array} \quad (11)$$

$$M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \left[ \begin{array}{cc} 0 & 0 \\ 0 & 0 \\ \hline M_{11} & 0 \\ 0 & 0 \\ 0 & M_{22} \end{array} \right], \quad Q = \begin{bmatrix} Q_1 \\ 0 \end{bmatrix} = \left[ \begin{array}{c} Q_{11} \\ Q_{12} \\ \hline Q_{13} \\ 0 \\ 0 \end{array} \right], \quad C = [0 \ T] \quad (12)$$

where  $[A_{42} \ A_{43}] \in \mathbb{R}^{(p-r) \times (\bar{p}-r)}$  which can be further partitioned to have the form

$$\begin{array}{cc} \begin{array}{c} \bar{p}-q \\ \leftarrow \\ \leftarrow \end{array} & \begin{array}{c} q-r \\ \leftarrow \\ \leftarrow \end{array} \\ \begin{array}{cc} \bar{p}-q & q-r \\ \leftarrow & \leftarrow \end{array} & = \begin{bmatrix} 0 & 0 \\ \hline A_{42}^0 & A_{43}^0 \end{bmatrix} \begin{array}{l} \Downarrow p - \bar{p} \\ \Downarrow \bar{p} - r \end{array} \end{array} \quad (13)$$

where  $[A_{42}^0 \ A_{43}^0]$  is square and invertible and  $\text{rank}(\tilde{A}_{32}) = \text{rank}(A_{43}^0)$ . Furthermore the matrix  $T \in \mathbb{R}^{p \times p}$  from (12) is orthogonal.

*Proof*

See Section A.4 in the Appendix.

The canonical form in (11)–(12) associated with Lemma 4 is the basis for the proof of Theorem 1 which will be developed in the next section. Also, partition  $A_3 \in \mathbb{R}^{p \times (n-p)}$  from (11) as

$$A_3 = \begin{bmatrix} A_{311} \\ A_{312} \end{bmatrix} \begin{matrix} \Downarrow p - \bar{p} \\ \Downarrow \bar{p} - r \end{matrix} \Rightarrow A_{311} = [0 \ A_{42} \ A_{43}], \quad A_{312} = [A_{51} \ A_{52} \ A_{53}] \quad (14)$$

Assume that the unknown signals  $f(t)$ ,  $\xi(t)$  are norm bounded by known scalars  $\alpha, \beta$  so that

$$\|f(t)\| < \alpha, \quad \|\xi(t)\| < \beta$$

The remainder of this section develops a fault estimation scheme for  $f(t)$  based on a pair of sliding mode observers as illustrated in Figure 1.

### 3.1. A fault reconstruction scheme (Proof of Theorem 1)

A sliding mode observer [12] for system (7)–(8) (which fulfils the role of the primary observer in Figure 1) is

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) - G_I e_y(t) + G_n v \quad (15)$$

$$\hat{y}(t) = C\hat{x}(t) \quad (16)$$

where  $\hat{x} \in \mathbb{R}^n$  is the estimate of the state  $x$  and  $e_y = \hat{y} - y$  is the output estimation error. The matrices  $G_I, G_n \in \mathbb{R}^{n \times p}$  are observer gains that are to be designed. In particular,  $G_n$  has the structure

$$G_n = \begin{bmatrix} -LT^T \\ T^T \end{bmatrix} P_0^{-1} \quad (17)$$

where  $P_0 \in \mathbb{R}^{p \times p}$  is a symmetric positive definite (s.p.d.) matrix and  $L \in \mathbb{R}^{(n-p) \times p}$  is such that  $A_1 + LA_3$  is stable, where the pair  $(A_1, A_3)$  is defined in (11). The matrix  $T$  is the orthogonal sub-block from (12). The term  $v$  is a nonlinear discontinuous term defined by

$$v = \rho \frac{e_y}{\|e_y\|}, \quad e_y \neq 0 \quad (18)$$

where  $\rho$  is a positive scalar.

Define  $e := \hat{x} - x$  as the state estimation error, and combine (7), (8), (15) and (16) to obtain the error system

$$\dot{e}(t) = (A - G_I C)e(t) + G_n v - Mf(t) - Q\xi(t) \quad (19)$$

Then, it can be shown that for an appropriate choice of  $G_I$  and for a large enough choice of  $\rho$ , an ideal sliding motion takes place on  $\mathcal{S} = \{e : Ce = 0\}$  in finite time. Furthermore, the sliding motion dynamics are governed by the system matrix  $A_1 + LA_3$ . Since from Lemma 2 the pair  $(A, C)$  is detectable, using the Popov–Hautus–Rosenbrock (PHR) test [18], it can be shown that  $(A_1, A_3)$  is detectable and so an  $L$  can always be found to make  $A_1 + LA_3$  stable. Details of the design process will be given in Section 3.2.

Introduce a new change of co-ordinates  $x \mapsto T_L x$  where

$$T_L = \begin{bmatrix} I_{n-p} & L \\ 0 & T \end{bmatrix}$$

Then, the matrices in (11)–(12) are transformed to be

$$T_L A T_L^{-1} = \begin{bmatrix} A_1 + L A_3 & * \\ T A_3 & * \end{bmatrix}, \quad T_L M = \begin{bmatrix} M_1 + L M_2 \\ T M_2 \end{bmatrix}, \quad C T_L^{-1} = [0 \ I_p] \quad (20)$$

$$T_L Q = \begin{bmatrix} Q_1 \\ 0 \end{bmatrix}, \quad T_L G_n = \begin{bmatrix} 0 \\ P_0^{-1} \end{bmatrix}, \quad T_L x = \begin{bmatrix} x_1 \\ y \end{bmatrix} \quad (21)$$

where  $x_1 \in \mathbb{R}^{n-p}$  are the ‘non-output’ states, and (\*) are matrices that play no role in the analysis that follows. Partition the error system (19) according to (20) and (21), and let  $e_1$  be the estimation error of  $x_1$ . Assume that an ideal sliding motion has taken place on  $\mathcal{S}$  so that  $e_y = \dot{e}_y = 0$  [8, 9, 12], then the error system (19) can be partitioned and re-arranged as

$$\dot{e}_1(t) = (A_1 + L A_3)e_1(t) - (M_1 + L M_2)f(t) - Q_1 \xi(t) \quad (22)$$

$$T^T P_0^{-1} v_{\text{eq}} = -A_3 e_1(t) + M_2 f(t) \quad (23)$$

where  $v_{\text{eq}}$  is the equivalent output error injection required to maintain a sliding motion [8, 9] and can be approximated to any degree of accuracy [8] by replacing  $v$  with

$$v = -\rho \frac{e_y}{\|e_y\| + \delta} \quad (24)$$

where  $\delta$  is a small positive scalar. As the term  $e_y$  is a measurable signal, the signal  $v_{\text{eq}}$  is computable online. For full details, see [8, 9].

Define  $v$  in Figure 1 as  $v := T^T P_0^{-1} v_{\text{eq}}$  and partition  $v = \text{col}(v_1, v_2)$  where  $v_2 \in \mathbb{R}^r$ . The contents of the scaling and filter block from Figure 1 will now be described to extract the signal  $\bar{y}$  from  $v$ . Partition (23) conformably according to (14) as

$$v_1(t) = -A_{311} e_1(t) \quad (25)$$

$$v_2(t) = -A_{312} e_1(t) + M_{22} f_2(t) \quad (26)$$

where  $f_2$  is a partition of  $f$  in Lemma 1. Then, define a matrix  $Z \in \mathbb{R}^{(\bar{p}-r) \times (p-r)}$  as  $Z = [0 \ I_{\bar{p}-r}]$  and multiply (25) by  $Z$  to get

$$\bar{v}_1(t) := Z v_1(t) = -Z A_{311} e_1(t) \quad (27)$$

From (13) and the partitions of  $A_{311}$  in (14) it is clear that  $Z A_{311} = [0 \ A_{42}^0 \ A_{43}^0]$  which has full row rank  $\bar{p} - r$  as deduced from Lemma 4.

Now low-pass filter  $v_2$  to produce  $v_f$  according to

$$\dot{v}_f(t) = -A_f v_f(t) + A_f v_2(t) = -A_f v_f(t) - A_f A_{312} e_1(t) + A_f M_{22} f_2(t) \quad (28)$$



where  $-A_f \in \mathbb{R}^{r \times r}$  is a stable design matrix, and combine (22), (28) and (27) to get the following system of order  $\bar{n} := n - p + r$

$$\begin{bmatrix} \dot{e}_1(t) \\ \dot{v}_f(t) \end{bmatrix} = \underbrace{\begin{bmatrix} A_1 + LA_3 & 0 \\ -A_f A_{312} & -A_f \end{bmatrix}}_{\mathcal{A}} \underbrace{\begin{bmatrix} e_1(t) \\ v_f(t) \end{bmatrix}}_z + \underbrace{\begin{bmatrix} -(M_1 + LM_2) \\ 0 \quad A_f M_{22} \end{bmatrix}}_{\mathcal{M}} f(t) + \underbrace{\begin{bmatrix} -Q_1 \\ 0 \end{bmatrix}}_{\mathcal{Q}} \zeta(t) \quad (29)$$

$$\underbrace{\begin{bmatrix} \bar{v}_1(t) \\ v_f(t) \end{bmatrix}}_{\bar{y}(t)} = \underbrace{\begin{bmatrix} -ZA_{311} & 0 \\ 0 & I_r \end{bmatrix}}_{\mathcal{C}} \underbrace{\begin{bmatrix} e_1(t) \\ v_f(t) \end{bmatrix}}_z \quad (30)$$

Define a non-singular transformation matrix  $\mathcal{T} \in \mathbb{R}^{\bar{n} \times \bar{n}}$  so that  $\bar{x} = \mathcal{T}z$  where

$$\mathcal{T} = \begin{bmatrix} I_{\bar{n}-r} & \tilde{L}A_f^{-1} \\ 0 & I_r \end{bmatrix}$$

where  $\tilde{L}$  represents the last  $r$  columns of  $L$ . Then, the matrices  $\mathcal{A}, \mathcal{M}, \mathcal{C}, \mathcal{Q}$  from (29) to (30) are transformed ( $\bar{A} = \mathcal{T}\mathcal{A}\mathcal{T}^{-1}, \bar{M} = \mathcal{T}\mathcal{M}, \bar{C} = \mathcal{C}\mathcal{T}^{-1}, \bar{Q} = \mathcal{T}\mathcal{Q}$ ) to be

$$\bar{A} = \left[ \begin{array}{c|ccc} A_{11} & * & * & * \\ \hline A_{21} & * & * & * \\ A_{31} & * & * & * \\ -A_f A_{51} & * & * & * \end{array} \right] = \begin{matrix} \bar{n} \leftrightarrow \bar{p} & \bar{q} \\ \left[ \begin{array}{cc} \bar{A}_1 & \bar{A}_2 \\ \bar{A}_3 & \bar{A}_4 \end{array} \right] & \begin{matrix} \downarrow \bar{n} - \bar{p} \\ \downarrow \bar{p} \end{matrix} \end{matrix}, \quad \bar{M} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -M_{11} & 0 \\ 0 & A_f M_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{M}_2 \end{bmatrix} \quad (31)$$

$$\bar{C} = \left[ \begin{array}{c|ccc} 0 & A_{42}^0 & A_{43}^0 & * \\ \hline 0 & 0 & 0 & I_r \end{array} \right] = [0 \quad \bar{T}], \quad \bar{Q} = \begin{bmatrix} Q_{11} \\ Q_{12} \\ Q_{13} \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{Q}_1 \\ \bar{Q}_2 \end{bmatrix} \quad (32)$$

where (\*) are terms that play no role in the subsequent analysis. Clearly, the matrix  $\bar{T} \in \mathbb{R}^{\bar{p} \times \bar{p}}$  is invertible since  $[A_{42}^0 \quad A_{43}^0]$  is square and invertible. Define  $\bar{M}_0$  to be the bottom  $q$  rows of  $\bar{M}_2$ , therefore,  $\bar{M}_0$  is square and invertible. From (31) and (32) it is easy to verify

$$\bar{C}\bar{M} = \begin{bmatrix} -A_{43}^0 M_{11} & * \\ 0 & A_f M_{22} \end{bmatrix}$$

By construction,  $M_{11}$  and  $M_{22}$  are invertible, and from Lemma 4,  $\text{rank}(A_{43}^0) = \text{rank}(\tilde{A}_{32})$ . It is shown in Lemma 3 that condition B1 implies that  $\text{rank}(\tilde{A}_{32}) = q - r$  and hence  $\bar{C}\bar{M}$  is full rank. It is shown in Lemma 5 in Section A.5 in the Appendix that the invariant zeros of  $(\bar{A}, \bar{M}, \bar{C})$  are given by the invariant zeros of the original system  $(\tilde{A}, \tilde{M}, \tilde{C})$  from (1) to (2) together with  $\lambda(A_\Omega)$ . Since from B2 it is assumed that  $(\tilde{A}, \tilde{M}, \tilde{C})$  has stable invariant zeros, the system  $(\bar{A}, \bar{M}, \bar{C})$  has

stable invariant zeros. So by construction the system in (29)–(30) meets the necessary and sufficient conditions of the reconstruction method in [13], if and only if conditions B1 and B2 are satisfied.

Since the signal  $\bar{y}$  defined in (30) is measurable, the approach from [13] will be used to design the secondary sliding mode observer from Figure 1 based on (29)–(30) to reconstruct the fault  $f$  whilst being robust to  $\zeta$ . From (31) and (32), note that  $\bar{M}_{22}, \bar{T}$  are both invertible. Therefore, the triple  $(\bar{A}, \bar{M}, \bar{C})$  is already in the co-ordinates in which the robustness analysis in [13] is carried out, hence no further co-ordinate transformations are required.<sup>‡</sup> The proposed observer for system (29)–(30) (the secondary sliding mode observer in Figure 1) in the co-ordinates of (31)–(32) is

$$\dot{\hat{x}}(t) = \bar{A}\hat{x}(t) - \bar{G}_l\bar{e}_y(t) + \bar{G}_n\bar{v} \tag{33}$$

$$\hat{y}(t) = \bar{C}\hat{x}(t) \tag{34}$$

where  $\bar{e}_y := \hat{y} - \bar{y}$ . As before, the matrices  $\bar{G}_l, \bar{G}_n \in \mathbb{R}^{\bar{n} \times \bar{p}}$  are observer gains to be designed, and in particular,  $\bar{G}_n$  has the structure (in the co-ordinates of (31) and (32))

$$\bar{G}_n = \begin{bmatrix} -\bar{L}\bar{T}^{-1} \\ \bar{T}^{-1} \end{bmatrix} \bar{P}_0^{-1}, \quad \bar{L} = [\bar{L}_1 \quad 0] \tag{35}$$

where  $\bar{P}_0 \in \mathbb{R}^{\bar{p} \times \bar{p}}$  is a s.p.d. matrix,  $\bar{L} \in \mathbb{R}^{(\bar{n}-\bar{p}) \times \bar{p}}$  and  $\bar{L}_1 \in \mathbb{R}^{(\bar{n}-\bar{p}) \times (\bar{p}-q)}$ . The term  $\bar{v}$  is a nonlinear discontinuous term defined by

$$\bar{v} = -\bar{\rho} \frac{\bar{e}_y}{\|\bar{e}_y\|} \tag{36}$$

where  $\bar{\rho}$  is a positive scalar.

For an appropriate choice of  $\bar{G}_l$  and a large enough choice of  $\bar{\rho}$ , it can be shown that an ideal sliding motion takes place on  $\bar{\mathcal{S}} = \{\bar{e} : \bar{C}\bar{e} = 0\}$  in finite time where  $\bar{e} := \hat{x} - \bar{x}$ . A detailed discussion on the design aspects is given in Section 3.2.

Let the fault reconstruction  $\hat{f}$  in Figure 1 be defined as

$$\hat{f}(t) := \bar{W}\bar{T}^{-1}\bar{P}_0^{-1}\bar{v}_{\text{eq}} \tag{37}$$

where  $\bar{W} := [\bar{W}_1 \quad \bar{M}_0^{-1}]$  with  $\bar{W}_1 \in \mathbb{R}^{q \times (\bar{p}-q)}$  and  $\bar{v}_{\text{eq}}$  is the equivalent output error injection required to maintain the sliding motion. The term  $\bar{v}_{\text{eq}}$  can be calculated online in the same way that  $v_{\text{eq}}$  in (24) is computed. When a sliding mode motion has taken place on  $\bar{\mathcal{S}}$ , from [13] and the definition of  $\hat{f}$  in (37), it can be shown that  $\hat{f}$  can be expressed as

$$\hat{f}(t) = f(t) + G(s)\xi(t) \quad \text{where } G(s) := \bar{W}\bar{A}_3(sI - (\bar{A}_1 + \bar{L}\bar{A}_3))^{-1}(\bar{Q}_1 + \bar{L}\bar{Q}_2) + \bar{W}\bar{Q}_2 \tag{38}$$

Therefore, it is clear that  $\hat{f}$  will capture  $f$  as well as a dynamic function of  $\xi$ . The approach in [13] seeks to minimize a measure of  $G(s)$  so that the effect of  $\xi$  on  $\hat{f}$  will be minimized. If there is no uncertainty then  $\bar{Q}_1 = \bar{Q}_2 = 0$  and so  $G(s) = 0$  and perfect reconstruction of  $f$  by  $\hat{f}$  is obtained.

<sup>‡</sup>However, there is a slight difference in that  $\bar{T}$  is invertible but not necessarily orthogonal as in [13]. This is of no major consequence as will be shown in the proceeding analysis.

3.2. Design of observers

This section discusses the design of the gains of the two observers from Figure 1. In this paper, the observers will be designed using the LMI control toolbox [19], which is a commonly available commercial software package.

For the design of the primary observer (15)–(16),  $G_l$  needs to be calculated such that the following inequality is satisfied:

$$P(A - G_l C) + (A - G_l C)^T P < 0 \tag{39}$$

where  $P$  has the structure

$$P = \begin{bmatrix} P_1 & P_1 L \\ L^T P_1 & T^T P_0 T + L^T P_1 L \end{bmatrix}, \quad P_1 \in \mathbb{R}^{(n-p) \times (n-p)} \text{ is s.p.d.} \tag{40}$$

so that a stable sliding motion can take place on  $\mathcal{S}$ . Then, the matrices  $L$  and  $P_0$  can be calculated from  $P$ , and subsequently  $G_n$  can be calculated from (17).

*Key observation:* Note that  $G(s)$  in (38) is not affected by the elements of  $L$  because of the structures of the partitions in (31)–(32). This means the transfer function  $G(s)$  is unaffected by the design parameters of the primary observer, which therefore can be designed using any method as long as  $P$  and  $G_l$  satisfy (39) and (40).

In this paper, the primary observer will be designed using the method in [20]. Define the following decision variable:

$$P_{\text{LMI}} = \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} > 0$$

where  $P_{11} \in \mathbb{R}^{(n-p) \times (n-p)}$  and  $P_{22} \in \mathbb{R}^{p \times p}$  are s.p.d. matrices. Note that  $P_{\text{LMI}}$  has the same structure as  $P$  in (40). Define another symmetric decision variable  $X \in \mathbb{R}^{n \times n}$ . The algorithm in [20] can be summarized as: minimize  $\text{trace}(X)$  subject to the following inequalities:

$$\begin{bmatrix} P_{\text{LMI}} A + A^T P_{\text{LMI}} - C^T V_2^{-1} C & P_{\text{LMI}} \\ P_{\text{LMI}} & -V_1^{-1} \end{bmatrix} < 0 \tag{41}$$

$$\begin{bmatrix} -P_{\text{LMI}} & I_n \\ I_n & -X \end{bmatrix} < 0 \tag{42}$$

where  $V_1 \in \mathbb{R}^{n \times n}$ ,  $V_2 \in \mathbb{R}^{p \times p}$  are s.p.d. weighting matrices to be chosen by the designer to tune the observer gains. The LMI toolbox will return values for the decision variables  $P_{\text{LMI}}$  and  $X$ , and the following observer parameters can be calculated:

$$G_l = P_{\text{LMI}}^{-1} C^T V_2^{-1}, \quad L = P_{11}^{-1} P_{12}, \quad P_0 = T^T (P_{22} - P_{12}^T P_{11}^{-1} P_{12}) T \tag{43}$$

and  $G_n$  can be calculated as in (17). The choice of  $G_l$  in (43) together with (41) ensures (39) is satisfied. Then, choosing the scalar  $\rho$  from (18) as

$$\rho \geq 2 \|P_0 T A_3\| \mu_1 (\alpha + \beta) / \mu_0$$

where

$$\mu_0 = -\lambda_{\max}(P(A - G_l C) + (A - G_l C)^T P), \quad \mu_1 = \|P[Q \ M]\|$$

ensures the sliding surface  $\mathcal{S}$  is reached in finite time and that a stable sliding mode is maintained [13]. For full details about the observer design and its motivation, see [20].

*Remark*

The observer (15)–(17) is slightly different to the one in [12, 13], in that here the matrix  $L$  is unconstrained. In [12, 13], the matrix  $L$  is forced to have a special structure. The observer (15)–(16) in this paper treats all the unknown signals  $\text{col}(\xi, f)$  as an ‘unmatched’ disturbance, because in general the distribution matrices  $M$  and  $Q$  are not matched to  $G_n$ , i.e.  $\text{rank}[G_n \ M \ Q] > \text{rank}(G_n)$ . As  $L$  is unconstrained, the observer (15)–(17) can be considered to be a modified Utkin observer [21] with the additional term  $G_l e_y$ .

The secondary observer in (33)–(34) is designed to satisfy

$$\bar{P}(\bar{A} - \bar{G}_l \bar{C}) + (\bar{A} - \bar{G}_l \bar{C})^T \bar{P} < 0 \tag{44}$$

where  $\bar{P}$  is s.p.d. and has the structure

$$\bar{P} = \begin{bmatrix} \bar{P}_1 & \bar{P}_1 \bar{L} \\ \bar{L}^T \bar{P}_1 & \bar{T}^T \bar{P}_0 \bar{T} + \bar{L}^T \bar{P}_1 \bar{L} \end{bmatrix} \tag{45}$$

where  $\bar{L}$  is given in (35) in order to achieve a sliding motion on  $\bar{\mathcal{S}}$ . In particular, the design algorithm in [13] will be used, where the objective is to minimize the  $\mathcal{L}_2$  gain of  $G(s)$ . *The design of the secondary observer is crucial to the quality of the reconstruction.* Again an LMI method will be used.

Define the following symmetric decision variable:

$$\bar{P}_{\text{LMI}} = \begin{bmatrix} \bar{P}_{11} & \bar{P}_{12} \\ \bar{P}_{12}^T & \bar{P}_{22} \end{bmatrix}, \quad \bar{P}_{12} = [\bar{P}_{121} \ 0]$$

where  $\bar{P}_{11} \in \mathbb{R}^{(\bar{n}-\bar{p}) \times (\bar{n}-\bar{p})}$ ,  $\bar{P}_{22} \in \mathbb{R}^{\bar{p} \times \bar{p}}$ ,  $\bar{P}_{121} \in \mathbb{R}^{(\bar{n}-\bar{p}) \times (\bar{p}-q)}$ . Also, define other decision variables  $\bar{\gamma} \in \mathbb{R}$  and  $\bar{W}_1 \in \mathbb{R}^{q \times (\bar{p}-q)}$ . Note that the structure of  $\bar{P}_{12}$  causes  $\bar{P}_{\text{LMI}}$  to have the same structure as  $\bar{P}$  in (45).

The design in [13] can be summarized as follows: minimize  $\bar{\gamma}$  subject to the following inequalities:

$$\begin{bmatrix} \bar{P}_{11} \bar{A}_1 + \bar{A}_1^T \bar{P}_{11} + \bar{P}_{12} \bar{A}_3 + \bar{A}_3^T \bar{P}_{12}^T & * & * \\ -(\bar{P}_{11} \bar{Q}_1 + \bar{P}_{12} \bar{Q}_2)^T & -\bar{\gamma} I_h & * \\ -\bar{W}_1 \bar{A}_3 & \bar{W}_1 \bar{Q}_2 & -\bar{\gamma} I_q \end{bmatrix} < 0 \tag{46}$$

$$\begin{bmatrix} \bar{P}_{\text{LMI}} \bar{A} + \bar{A}^T \bar{P}_{\text{LMI}} - \bar{\gamma}_0 \bar{C}^T (\bar{D}_d \bar{D}_d^T)^{-1} \bar{C} & * & * \\ -\bar{B}_d^T \bar{P} & -\bar{\gamma}_0 I_{\bar{p}+h} & * \\ \bar{E} & \bar{H} & -\bar{\gamma}_0 I_q \end{bmatrix} < 0 \tag{47}$$

$$\bar{P}_{\text{LMI}} > 0 \tag{48}$$

where (\*) are terms that make (46)–(47) symmetric. The fixed matrices are  $\bar{B}_d := [0 \ \bar{Q}]$ ,  $\bar{D}_d := [\bar{D}_1 \ 0]$  whilst  $\bar{H} = [0 \ \bar{W} \ \bar{Q}_2]$  where  $\bar{D}_1 \in \mathbb{R}^{\bar{p} \times \bar{p}}$  and  $\bar{\gamma}_0$  are user-specified parameters to tune the gains  $\bar{G}_l, \bar{G}_n$ . After the LMI solver returns the values of  $\bar{W}$  and  $\bar{P}$ , the gain  $\bar{G}_l$  can be calculated as  $\bar{G}_l = \bar{\gamma}_0^{-1} \bar{P}_{\text{LMI}}^{-1} \bar{C}^T (\bar{D}_d \bar{D}_d^T)^{-1}$  and  $\bar{G}_n$  as in (35). This algorithm ensures inequality (44) is satisfied and the  $\mathcal{L}_2$  gain from  $\xi$  to  $\hat{f}$  is bounded by  $\bar{\gamma}$ . Then choosing the gain from (36) as

$$\bar{\rho} > 2\|\bar{P}_0 \bar{T} \bar{A}_3\| \bar{\mu}_1 \beta / \bar{\mu}_0 + \|\bar{P}_0 \bar{T} \bar{Q}_2\| \beta + \|\bar{P}_0 \bar{T} \bar{M}_2\| \alpha$$

where

$$\bar{\mu}_0 = -\lambda_{\max}(\bar{P}(\bar{A} - \bar{G}_l \bar{C}) + (\bar{A} - \bar{G}_l \bar{C})^T \bar{P}), \quad \bar{\mu}_1 = \|\bar{P} \bar{Q}\|$$

ensures a sliding motion of the secondary observer on  $\mathcal{S}$ . For full details, see [13]. The secondary observer now treats the fault vector  $f$  as the matched fault (in the sense that its distribution matrix is ‘matched’ to  $\bar{G}_n$ , i.e.  $\text{rank}[\bar{G}_n \ \bar{M}] = \text{rank}(\bar{G}_n)$ ) and  $\xi$  as the unmatched disturbance.

*Remark*

The matrix in condition B1 is formed from Markov parameters and is system realization independent. (It is also a sub-block of the Hankel matrix [22].) Intuitively, it is related to the system  $(\tilde{A}, \tilde{M}, \tilde{C})$  having relative degree two since for example if

$$\tilde{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \tilde{M} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \tilde{C} = [1 \ 0]$$

(i.e. a double integrator realization) then B1 is satisfied although  $\tilde{C} \tilde{M} = 0$ .

3.3. Design algorithm summary

The design algorithm for the method in this paper can be summarized as follows:

1. Preliminary checks and co-ordinate transformations:
  - (a) Check that  $\text{rank}(\tilde{C} \tilde{M}) < \text{rank}(\tilde{M})$ . Otherwise, stop, and the existing methods of [13] for example can be used. Then, ensure that Conditions B1 and B2 are satisfied. If not, the method in this paper is not applicable.
  - (b) From the knowledge of the frequency of  $\tilde{\xi}$ , select the matrices  $A_\Omega$  and  $B_\Omega$ .
  - (c) Find the co-ordinate transformations  $T_1, T_2$ , following the steps shown in the proof of Lemma 1 in Section A.1 in the Appendix. Then, apply the co-ordinate transformation to  $\tilde{A}, \tilde{M}, \tilde{C}$  and  $\tilde{Q}$ . The matrices  $\tilde{C}$  and  $\tilde{M}$  will then have the special structures shown in (4) and (5) so that the analysis of the observer in (22)–(23) and (25)–(26) is simplified without loss of generality.
  - (d) Use the matrices obtained in the previous step to form the matrices  $A, M, Q, C$  of the augmented system in (7) and (8).
  - (e) Find the co-ordinate transformation  $T_5$ , following the steps shown in the proof of Lemma 4 in Section A.4 in the Appendix. Then, apply the co-ordinate transformation to the matrices  $A, M, C$  and  $Q$ , which then have the structures in (11)–(13). Obtain the partitions in Lemma 4 and Equation (14). The purpose of the co-ordinate transformation  $T_5$  is to give the matrix  $A$  the special structure in (11) and (13) so

that the analysis of the secondary observer in (31)–(32) can be simplified without loss of generality.

## 2. Design observers

- (a) Design the primary observer using the method in Section 3.2 (although in fact any method of choice can be used). Full details of this design method are available in [20]. The design method will yield the gains  $G_l, G_n, L$  and the matrix  $P_0$ . Note that other methods can be used to design the primary observer, as long as the designed parameters satisfy (39).
- (b) Define a stable matrix  $-A_f$  as in (28). Then, obtain the matrices for the system in (29)–(30), in particular  $\mathcal{A}, \mathcal{M}, \mathcal{Q}$  and  $\mathcal{C}$ .
- (c) Perform the co-ordinate transformation  $\mathcal{T}$  on  $\mathcal{A}, \mathcal{M}, \mathcal{Q}, \mathcal{C}$  to obtain  $\bar{A}, \bar{M}, \bar{Q}, \bar{C}$  as in (31)–(32). The purpose of the co-ordinate transformation  $\mathcal{T}$  is to make the system in (29)–(30) have the canonical observer structure from [13] as described in (31)–(32).
- (d) Design the secondary observer using the method in Section 3.2, in particular from (46) to (48). Full details of this design method are available in [13]. This design method will synthesize the gains  $\bar{G}_l, \bar{G}_n, \bar{W}$  and also minimize the  $\mathcal{L}_2$  gain from  $\xi$  to  $\hat{f}$ .

## 3. Implement the observers and generate the fault reconstruction as shown in Figure 1.

- (a) Implement the primary observer (15)–(16) on the original system (1)–(2). The observer will generate the signal  $v$  as in (24).
- (b) Process the signal  $v$  as in (27)–(28) to generate the output  $\bar{y}$ .
- (c) Implement the secondary observer (33)–(34) on the fictitious system (29)–(30), using the (measurable) signal  $\bar{y}$  as the driving signal.
- (d) Generate the fault reconstruction as in (37) from  $\bar{v}_{\text{eq}}$ .

## 4. AN EXAMPLE

The method proposed in this paper will now be demonstrated by an example, which is a seventh-order model of an aircraft [23]. In the notation of (1)–(2), the matrices that describe the system are as follows:

$$\tilde{A} = \begin{bmatrix} 0 & 0 & 1.0000 & 0 & 0 & 0 & 0 \\ 0 & -0.1540 & -0.0042 & 1.5400 & 0 & -0.7440 & -0.0320 \\ 0 & 0.2490 & -1.0000 & -5.2000 & 0 & 0.3370 & -1.1200 \\ 0.0386 & -0.9960 & -0.0003 & -2.1170 & 0 & 0.0200 & 0 \\ 0 & 0.5000 & 0 & 0 & -4.0000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -20.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -25.0000 \end{bmatrix}$$

where the states are the bank angle, yaw rate, roll rate, sideslip angle, washed-out filter state, rudder deflection, aileron deflection, and the inputs are the rudder command and the aileron

command. Assume that the bank angle, yaw rate and sideslip angle are measurable, and that the first actuator is faulty. Therefore, the matrices  $\tilde{C}$  and  $\tilde{M}$  are

$$\tilde{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \quad \tilde{M} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 20 \\ 0 \end{bmatrix}$$

Suppose that the matrix  $\tilde{A}$  is imprecisely known and that there exists parametric uncertainty. Therefore, the state equation of the system becomes

$$\dot{\tilde{x}} = (\tilde{A} + \Delta\tilde{A})\tilde{x} + \tilde{B}u + \tilde{M}f \tag{49}$$

where  $\Delta\tilde{A}$  is the discrepancy between the known matrix  $\tilde{A}$  and its actual value. For simplicity let  $u \equiv 0$ . Note that the first, fifth, sixth and seventh rows of the matrix  $\tilde{A}$  do not contain any uncertainty due to the nature of the state equations. Hence, any parametric uncertainty will appear in the second, third and fourth rows of  $\tilde{A}$ . Let the actual value of the system matrix be

$$\tilde{A} + \Delta\tilde{A} = \begin{bmatrix} 0 & 0 & 1.0000 & 0 & 0 & 0 & 0 \\ 0 & -0.1600 & -0.0042 & 1.6600 & 0 & -0.7440 & -0.0500 \\ 0 & 0.2490 & -1.0000 & -5.1600 & 0 & 0.4000 & -1.2400 \\ 0.0386 & -0.9960 & -0.0003 & -2.2300 & 0 & 0.0230 & 0 \\ 0 & 0.5000 & 0 & 0 & -4.0000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -20.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -25.0000 \end{bmatrix}$$

Equation (49) can be placed in the same framework as (1) by writing

$$\underbrace{\Delta\tilde{A}\tilde{x}}_{\tilde{Q}} = \underbrace{\tilde{Q}\tilde{\xi}}_{\tilde{\xi}} = \begin{bmatrix} 0 & -0.0060 & 0 & 0.1200 & 0 & 0 & -0.0180 \\ 0 & 0 & 0 & 0.0400 & 0 & 0.0630 & -0.1200 \\ 0 & 0 & 0 & -0.1130 & 0 & 0.0030 & 0 \end{bmatrix} \tilde{x}$$

The disturbance  $\tilde{\xi}$  will be generated by the states  $\tilde{x}$ , which is in turn generated by the fault  $f$ . Assuming  $f$  is bounded, then  $\tilde{x}$  and  $\tilde{\xi}$  will also be bounded since  $\tilde{A} + \Delta\tilde{A}$  is stable. Note that the method in [15] cannot be used on this system as there is no consideration of the disturbance  $\tilde{\xi}$ . If the signals  $f$  and  $\tilde{\xi}$  are augmented to form a new ‘fault’ vector, as in [10], this would result in the new ‘fault’ having four components. The number of outputs in this system is only three, resulting in a ‘more faults than outputs’ scenario, and hence the method in [15] would still be not applicable.

The FDI literature based on UIOs is also not applicable here (because assumptions A1 and A2 are typically required [10, 24–26]). Note that all faults and disturbances appear in states 2, 3, 4 and 6. For the method in [16] to be applicable, the integral of the states 2, 3, 4 and 6 would need to be measurable, which means that the yaw and roll angles need to be measurable, together with the integrals of the sideslip angle and rudder deflection. In this system, of the four states, only the sideslip angle is measured. The rest are not measured and hence the method in [16] is not applicable for this system. Also note that  $\tilde{C}\tilde{M} = 0 \Rightarrow r = 0 < q$ , and hence the existing sliding mode methods [8, 13] cannot be used to reconstruct the fault.

#### 4.1. Observer design

It can be easily verified that conditions B1 and B2 are satisfied. Hence, the method proposed in this paper can be used.

The disturbance  $\tilde{\xi}$  is assumed to have a frequency content  $\omega < 10$  rad/s, and therefore the filter (3) will have the Bode plot as shown in Figure 2, resulting in  $A_\Omega = -10I_3$ ,  $B_\Omega = 10I_3$ .

Performing the co-ordinate transformation in Lemma 1 yields the following matrices:

$$T_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad T_2 = 1, \quad \tilde{T} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$M_{11} = -20, \quad M_{22} = \phi \text{ (the empty matrix)}$$

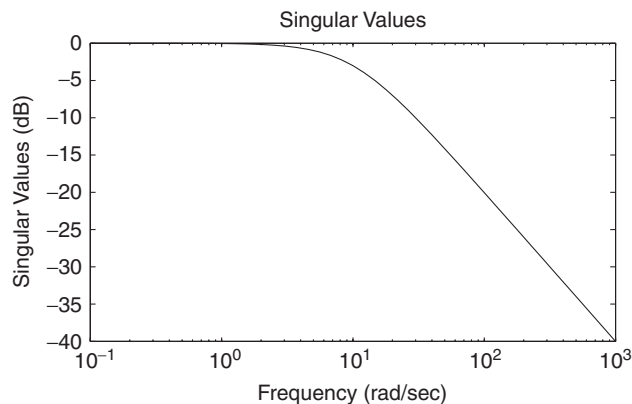


Figure 2. The gain plot of the filter (3).



with

$$\tilde{A} = \left[ \begin{array}{cccc|ccc} -25.0000 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1.1200 & -1.0000 & 0 & -0.3370 & -5.2000 & 0.2490 & 0 \\ 0 & 0 & -4.0000 & 0 & 0 & 0.5000 & 0 \\ \hline 0 & 0 & 0 & -20.0000 & 0 & 0 & 0 \\ 0 & -0.0003 & 0 & -0.0200 & -2.1170 & -0.9960 & 0.0386 \\ -0.0320 & -0.0042 & 0 & 0.7440 & 1.5400 & -0.1540 & 0 \\ 0 & 1.0000 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

$$\tilde{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which shows that  $\tilde{A}_{32}$  is full rank, and Condition C1 is fulfilled. Furthermore,  $\text{rank}[\tilde{Q}_{21} \tilde{A}_{31} \tilde{A}_{32}] = 3$ , which means that  $\bar{p} = 3$ .

Augmenting the system as in (7)–(8) yields a 10th order system. Performing the co-ordinate transformation in Lemma 4 gives

$$T_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.0335 & 0.0334 & 0.9989 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9994 & 0.0045 & 0.0334 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.0034 & 0.9994 & -0.0335 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$[A_{42}^0 \ A_{43}^0] = \begin{bmatrix} -0.0331 & 0.9995 & 0 \\ 0.9991 & 0.0334 & -0.0335 \\ 0.0267 & 0.0051 & 1.2465 \end{bmatrix}$$

The algorithm in Section 3.2 has been implemented to design the primary observer. In designing the primary observer,  $V_1 = 100I_{10}$ ,  $V_2 = I_3$  were chosen. The following gain matrices were obtained:

$$G_l = G_n = \begin{bmatrix} -0.0163 & -0.0391 & 0.0031 \\ 0.0005 & 0.3498 & -0.0230 \\ 0.0150 & -0.0302 & 0.0026 \\ -0.0019 & 0.0006 & 0.0001 \\ -0.1520 & -0.0133 & -0.1248 \\ 4.5838 & -0.3234 & -3.6522 \\ 0.0008 & -0.1192 & 0.0081 \\ 10.4474 & -0.0274 & -0.1554 \\ 0.1554 & -0.1322 & -8.1133 \\ 0.0274 & -9.8817 & -0.1322 \end{bmatrix}, \quad P_0 = \begin{bmatrix} 0.0957 & 0.0002 & 0.0018 \\ 0.0002 & 0.1012 & -0.0016 \\ 0.0018 & -0.0016 & 0.1233 \end{bmatrix}$$

The secondary observer was designed using the second algorithm in Section 3.2. The user-defined matrices were specified to be  $\bar{D}_1 = 10I_3$ ,  $\bar{\gamma}_0 = 100$  and the following gains were obtained:

$$\bar{G}_l = \bar{G}_n = \begin{bmatrix} 0.1191 & 0.0019 & -0.1074 \\ 0.0003 & 0.0000 & -0.0002 \\ 0.1580 & 0.0021 & -0.1425 \\ 0.0013 & 0.0001 & -0.0012 \\ 0.0058 & -0.0745 & -0.0051 \\ -0.1310 & -0.0043 & 0.1182 \\ 0.0953 & 0.0695 & -2.5613 \end{bmatrix}, \quad \bar{P}_0 = \begin{bmatrix} 7.8917 & 0.1416 & 0.2962 \\ 0.1416 & 13.4004 & 0.3605 \\ 0.2962 & 0.3605 & 0.3338 \end{bmatrix}$$

$$\bar{W}\bar{T}^{-1}\bar{P}_0^{-1} = [0.0505 \quad -0.0050 \quad 0.1618]$$

The gains above provide an  $\mathcal{L}_2$  bound of  $\bar{\gamma} = 1.2241$ .

#### 4.2. Simulation results

In the simulations that follow, the parameters associated with  $v$  for the primary observer were chosen as  $\rho = 100$ ,  $\delta = 10^{-5}$  while for the secondary observer they were chosen as  $\bar{\rho} = 100$ ,  $\bar{\delta} = 10^{-5}$ . A fault was induced in the first actuator. Figure 3 shows the states that will affect  $\tilde{\xi}$ , i.e. the second, fourth, sixth and seventh states. Figure 4 shows the fault and its reconstruction, where the left subfigure is the fault and the right subfigure is the reconstruction. It can be clearly seen that the reconstruction signal  $\hat{f}$  provides a good estimate of the fault  $f$ , despite the parameter variation  $\Delta\hat{A}$  that causes a disturbance that could corrupt the reconstruction. The

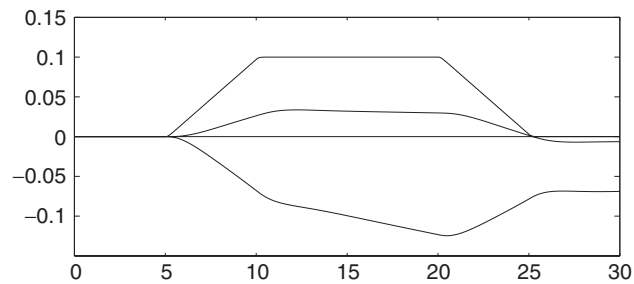


Figure 3. The components of  $\tilde{x}$  that generate  $\tilde{\xi}$ , namely the second, fourth, sixth and seventh components of  $\tilde{x}$ .

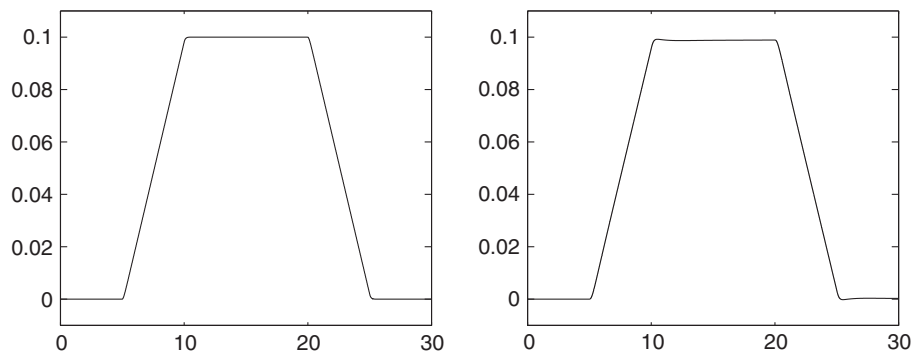


Figure 4. The left subfigure is the fault, the right subfigure is its reconstruction.

design algorithm has calculated the observer gains such that the reconstruction is least affected by the disturbances in an  $\mathcal{L}_2$  sense.

## 5. CONCLUSION

This paper has proposed a new scheme for robust fault reconstruction in uncertain systems which is applicable to a wider class of systems than existing work. Early results were applicable only to systems which are relative degree one and minimum phase with respect to the effect of faults on the measured outputs. Recent work has relaxed these conditions by either considering a specific class of nonlinear systems (relative degree two plants arising from models of mechanical systems) or by considering linear (certain) systems subject to a class of unknown inputs (which could be an aggregation of faults and uncertainty). The method proposed in this paper uses two sliding mode observers in cascade; the equivalent output error injection term of the first observer is processed to form the measurable output of a fictitious system. Then a secondary observer is implemented for the fictitious system such that the fault can be reconstructed using existing methods. An aircraft model has demonstrated the validity of the proposed scheme.

## APPENDIX A: PROOFS

*A.1. Proof of Lemma 1*

Since  $\tilde{C}$  has full row rank, there exists an invertible change of co-ordinates  $T_a = \begin{bmatrix} N_c \\ \tilde{C} \end{bmatrix}$  where  $N_c^T$  spans the null space of  $\tilde{C}$ , such that

$$M_a := T_a \tilde{M} = \begin{bmatrix} M_{a,1} \\ M_{a,2} \end{bmatrix}, \quad C_a := \tilde{C} T_a^{-1} = [0 \quad I_p]$$

where  $M_{a,2} \in \mathbb{R}^{p \times q}$ . Since  $\text{rank}(\tilde{C}\tilde{M}) = r$ , then  $\text{rank}(M_{a,2}) = r$ . Hence, there exist orthogonal matrices  $\tilde{T} \in \mathbb{R}^{p \times p}$ ,  $T_2 \in \mathbb{R}^{q \times q}$  such that

$$\tilde{T}^T M_{a,2} T_2^T = \begin{bmatrix} 0 & 0 \\ 0 & M_{22} \end{bmatrix}$$

where  $M_{22} \in \mathbb{R}^{r \times r}$  is invertible. Partition  $M_{a,1} T_2^T = [M_{a,11} \quad M_{a,12}]$  where  $M_{a,12}$  has  $r$  columns. Then, let  $X_1$  be an orthogonal matrix such that

$$X_1 M_{a,11} = \begin{bmatrix} 0 \\ M_{11} \end{bmatrix}$$

where  $M_{11} \in \mathbb{R}^{(q-r) \times (q-r)}$  is invertible. Define  $X_2 = [0 \quad -X_1 M_{a,12} M_{22}^{-1}] \tilde{T}^T$ . Then, apply the change of co-ordinates  $T_1$  to  $M_a$  and  $C_a$  where

$$T_1 = \begin{bmatrix} X_1 & X_2 \\ 0 & \tilde{T}^T \end{bmatrix}$$

and post-multiply  $M_a$  by  $T_2^T$  to get

$$M = T_1 M_a T_2^T = \begin{bmatrix} 0 & 0 \\ M_{11} & 0 \\ 0 & 0 \\ 0 & M_{22} \end{bmatrix}, \quad C = C_a T_1^{-1} = [0 \quad \tilde{T}]$$

and the co-ordinate transformation is complete.

*A.2. Proof of Lemma 2*

Let the matrices  $(A, C)$  be in the partitions of (9)–(10). By performing the Popov–Hautus–Rosenbrock (PHR) rank test [18] on the pair  $(A, C)$  and using the fact that  $(\tilde{A}, \tilde{C})$  is observable if and only if

$$\text{rank} \begin{bmatrix} sI - \tilde{A}_1 \\ \tilde{A}_3 \end{bmatrix} = \tilde{n} - p$$

for all  $s \in \mathbb{C}$ , it is clear that the unobservable modes of  $(A, C)$  are given by  $\lambda(A_\Omega)$ . By assumption,  $A_\Omega$  is a stable matrix and therefore  $(A, C)$  is detectable.

### A.3. Proof of Lemma 3

Define

$$H := \begin{bmatrix} \tilde{C}\tilde{A}\tilde{M} & \tilde{C}\tilde{M} \\ \tilde{C}\tilde{M} & 0 \end{bmatrix} \quad (\text{A1})$$

Therefore, from (4) to (5),

$$\begin{aligned} H &= \begin{bmatrix} \tilde{T} & 0 \\ 0 & \tilde{T} \end{bmatrix} \begin{bmatrix} \tilde{A}_3\tilde{M}_1 + \tilde{A}_4\tilde{M}_2 & \tilde{M}_2 \\ & \tilde{M}_2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \tilde{T} & 0 \\ 0 & \tilde{T} \end{bmatrix} \left[ \begin{array}{cc|cc} \tilde{A}_{32}M_{11} & \tilde{A}_{34}M_{22} & 0 & 0 \\ \tilde{A}_{42}M_{11} & \tilde{A}_{44}M_{22} & 0 & M_{22} \\ \hline 0 & 0 & 0 & 0 \\ 0 & M_{22} & 0 & 0 \end{array} \right] \end{aligned}$$

It is clear that  $\text{rank}(H) = \text{rank}(M_{22}) + \text{rank}(M_{22}) + \text{rank}(\tilde{A}_{32}M_{11})$ . Then, it follows that  $\text{rank}(H) = r + r + \text{rank}(\tilde{A}_{32}) = \text{rank}(\tilde{C}\tilde{M}) + r + \text{rank}(\tilde{A}_{32})$  since  $M_{11}, M_{22}$  are square and invertible. It is then straightforward to see that as  $\text{rank}(\tilde{M}) = q$ , condition B1 holds if and only if  $\text{rank}(\tilde{A}_{32}) = q - r$ .

### A.4. Proof of Lemma 4

In the co-ordinates of (9)–(10), define

$$R_1 = [\tilde{Q}_{21} \ \tilde{A}_{31} \ \tilde{A}_{32}], \quad R_2 = \begin{bmatrix} 0 \\ 0 \\ M_{11} \end{bmatrix} \begin{array}{l} \Downarrow h \\ \Downarrow n - h - p - q + r \\ \Downarrow q - r \end{array} \quad (\text{A2})$$

Therefore,  $M$  in (10) is

$$\begin{bmatrix} R_2 & 0 \\ 0 & 0 \\ 0 & M_{22} \end{bmatrix} \begin{array}{l} \Downarrow n - p \\ \Downarrow p - r \\ \Downarrow r \end{array}$$

Recall that by definition  $\text{rank}(R_1) = \bar{p} - r$  and that  $\tilde{A}_{32}$  has full column rank  $q - r$ . Let  $X_3 \in \mathbb{R}^{(n-p) \times (n-p)}$  and  $X_4 \in \mathbb{R}^{(p-r) \times (p-r)}$  be orthogonal matrices such that

$$X_4 R_1 X_3^T = [0 | A_{a,42} \ A_{a,43}] = \begin{bmatrix} 0 & 0 \\ 0 & U \end{bmatrix} \quad (\text{A3})$$

where  $U \in \mathbb{R}^{(\bar{p}-r) \times (\bar{p}-r)}$  is invertible. Then define a non-singular change of co-ordinates  $T_3 \in \mathbb{R}^{n \times n}$  where

$$T_3 = \left[ \begin{array}{c|cc} X_3 & 0 & 0 \\ \hline 0 & X_4 & 0 \\ 0 & 0 & I_r \end{array} \right]$$

Apply the change of co-ordinates  $T_3$  to  $A, M, Q, C$  in (9)–(10) to obtain

$$A_a = \begin{bmatrix} A_{a,1} & A_{a,2} \\ A_{a,3} & A_{a,4} \end{bmatrix}, \quad M_a = \begin{bmatrix} M_{a,1} \\ M_{a,2} \end{bmatrix}, \quad Q_a = \begin{bmatrix} Q_{a,1} \\ 0 \end{bmatrix}, \quad C_a = [0 \ T_a] \tag{A4}$$

where  $A_a := T_3 A T_3^{-1}, M_a := T_3 M, Q_a := T_3 Q, C_a := C T_3^{-1}$ . Further partition

$$A_{a,1} = \begin{bmatrix} A_{a,11} & A_{a,12} & A_{a,13} \\ A_{a,21} & A_{a,22} & A_{a,23} \\ A_{a,31} & A_{a,32} & A_{a,33} \end{bmatrix} \begin{matrix} \Downarrow n - \bar{p} + r - p \\ \Downarrow \bar{p} - q \\ \Downarrow q - r \end{matrix}, \quad A_{a,3} = \begin{bmatrix} 0 & A_{a,42} & A_{a,43} \\ A_{a,51} & A_{a,52} & A_{a,53} \end{bmatrix} \begin{matrix} \Downarrow p - r \\ \Downarrow r \end{matrix} \tag{A5}$$

$$Q_{a,1} = \begin{bmatrix} Q_{a,11} \\ Q_{a,12} \\ Q_{a,13} \end{bmatrix}, \quad M_{a,1} = \begin{bmatrix} M_{a,11} & 0 \\ M_{a,12} & 0 \\ M_{a,13} & 0 \end{bmatrix}, \quad M_{a,2} = \begin{bmatrix} 0 & 0 \\ 0 & M_{22} \end{bmatrix} \tag{A6}$$

where  $T_a$  in (A4) is still orthogonal. It is straightforward to show that  $R_1 R_2 = \tilde{A}_{32} M_{11}$ . Since the matrix  $\tilde{A}_{32}$  has full column rank  $q - r$  and  $\det(M_{11}) \neq 0$ , then  $\text{rank}(R_1 R_2) = q - r$ . Clearly,  $R_1 R_2 = R_1 X_3^{-1} X_3 R_2$ , which from (A3) can be expanded to be

$$\underbrace{X_4^{-1} [0 \ A_{a,42} \ A_{a,43}]}_{R_1 X_3^{-1}} \underbrace{\begin{bmatrix} M_{a,11} \\ M_{a,12} \\ M_{a,13} \end{bmatrix}}_{X_3 R_2} = X_4^{-1} [A_{a,42} \ A_{a,43}] \begin{bmatrix} M_{a,12} \\ M_{a,13} \end{bmatrix} = X_4^{-1} \begin{bmatrix} 0 \\ U \end{bmatrix} \begin{bmatrix} M_{a,12} \\ M_{a,13} \end{bmatrix}$$

Since  $X_4$  is orthogonal and  $\bar{p} > q$ , it follows that

$$\text{rank}(\tilde{A}_{32}) = q - r \Rightarrow \text{rank}(R_1 R_2) = q - r \Rightarrow \text{rank} \begin{bmatrix} M_{a,12} \\ M_{a,13} \end{bmatrix} = q - r$$

Define two matrices  $X_5 \in \mathbb{R}^{(q-r) \times (\bar{p}-r)}$  and  $X_6 \in \mathbb{R}^{(\bar{p}-r) \times (\bar{p}-r)}$  so that

$$X_5 \begin{bmatrix} M_{a,12} \\ M_{a,13} \end{bmatrix} = I_{q-r}, \quad X_6 \begin{bmatrix} M_{a,12} \\ M_{a,13} \end{bmatrix} = \begin{bmatrix} 0 \\ M_{11} \end{bmatrix}$$

Then introduce the final change of co-ordinates

$$T_4 = \left[ \begin{array}{cc|c} I_{n-\bar{p}+r-p} & -M_{a,11}X_5 & 0 \\ 0 & X_6 & 0 \\ \hline 0 & 0 & I_p \end{array} \right]$$

so that  $A_a, M_a, Q_a, C_a$  are transformed to be

$$A_b = \begin{bmatrix} A_{b,1} & A_{b,2} \\ A_{b,3} & A_{b,4} \end{bmatrix}, \quad M_b = \begin{bmatrix} M_{b,1} \\ M_{b,2} \end{bmatrix}, \quad C_b = [0 \ T_a], \quad Q_b = \begin{bmatrix} Q_{b,1} \\ 0 \end{bmatrix} \tag{A7}$$

where

$$A_{b,3} = \begin{bmatrix} 0 & A_{b,42} & A_{b,43} \\ A_{b,51} & A_{b,52} & A_{b,53} \end{bmatrix} \begin{matrix} \Downarrow p-r \\ \Downarrow r \end{matrix}, \quad M_{b,1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ M_{11} & 0 \end{bmatrix} \begin{matrix} \Downarrow n-p-\bar{p}+r \\ \Downarrow \bar{p}-q \\ \Downarrow q-r \end{matrix}, \quad M_{b,2} = \begin{bmatrix} 0 & 0 \\ 0 & M_{22} \end{bmatrix} \begin{matrix} \Downarrow p-r \\ \Downarrow r \end{matrix}$$

and from (A3)

$$[A_{b,42} \ A_{b,43}] = \begin{bmatrix} 0 \\ UX_6^{-1} \end{bmatrix} \begin{matrix} \Downarrow p-\bar{p} \\ \Downarrow \bar{p}-r \end{matrix}$$

By defining the non-singular transformation matrix  $T_5 := T_4T_3$  and partitioning

$$UX_6^{-1} = [A_{42}^0 \ A_{43}^0] \tag{A8}$$

where  $A_{43}^0 \in \mathbb{R}^{(\bar{p}-r) \times (q-r)}$ , the matrices  $A_b, M_b, Q_b, C_b$  and their partitions are now in the same form as the matrices  $A, M, Q, C$  in (11)–(12) in the statement of Lemma 4. To prove that  $\text{rank}(A_{43}^0) = \text{rank}(\tilde{A}_{32})$ , define

$$X_7 = \begin{bmatrix} I_{n-\bar{p}-p+r} & -M_{a,11}X_5 \\ 0 & X_6 \end{bmatrix}$$

From the co-ordinate transformations  $T_3, T_4$  and by observing the structure of  $A_b$  and  $M_b$ , it is clear from (A3) and (A8) that

$$X_4R_1X_3^{-1}X_7^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & A_{42}^0 & A_{43}^0 \end{bmatrix}, \quad X_7X_3R_2 = \begin{bmatrix} 0 \\ 0 \\ M_{11} \end{bmatrix} \Rightarrow X_4R_1R_2 = \begin{bmatrix} 0 \\ A_{43}^0M_{11} \end{bmatrix}$$

Recalling that  $\text{rank}(R_1R_2) = \text{rank}(\tilde{A}_{32}) = q-r$ , and using the fact that  $X_4$  and  $M_{11}$  are invertible, it follows that  $\text{rank}(A_{43}^0) = \text{rank}(\tilde{A}_{32})$ .

### A.5. Lemma 5 and its Proof

#### Lemma 5

The invariant zeros of  $(\bar{A}, \bar{M}, \bar{C})$  are identical to the invariant zeros of  $(\tilde{A}, \tilde{M}, \tilde{C})$  together with the eigenvalues of  $A_\Omega$ .

#### Proof

The Rosenbrock system matrix [18] of  $(\bar{A}, \bar{M}, \bar{C})$  is given by

$$E_{a,1}(s) = \begin{bmatrix} sI - \bar{A} & \bar{M} \\ \bar{C} & 0 \end{bmatrix}$$

and the invariant zeros of a system are the values of  $s$  that cause its Rosenbrock matrix to lose normal rank. From (31) to (32),  $E_{a,1}(s)$  can be expanded to be

$$E_{a,1}(s) = \begin{bmatrix} sI - \bar{A}_1 & -\bar{A}_2 & 0 \\ -\bar{A}_3 & sI - \bar{A}_4 & \bar{M}_2 \\ 0 & \bar{T} & 0 \end{bmatrix}$$

Since  $\bar{T}$  has full rank, it is clear that  $E_{a,1}(s)$  loses rank if and only if the following matrix loses rank

$$E_{a,2}(s) := \begin{bmatrix} sI - \bar{A}_1 & 0 \\ -\bar{A}_3 & \bar{M}_2 \end{bmatrix}$$

Substituting for  $\bar{A}_1, \bar{A}_3, \bar{M}_2$  from (31) to (32),  $E_{a,2}(s)$  can be expanded to be

$$E_{a,2}(s) = \left[ \begin{array}{c|cc} sI - A_{11} & 0 & 0 \\ \hline -A_{21} & 0 & 0 \\ -A_{31} & -M_{11} & 0 \\ A_f A_{51} & 0 & A_f M_{22} \end{array} \right]$$

It is then obvious to see that  $E_{a,2}(s)$  loses rank if and only if  $E_{a,3}(s)$  loses rank where

$$E_{a,3}(s) := \begin{bmatrix} sI - A_{11} \\ -A_{21} \end{bmatrix}$$

From the PHR rank test [18], the values of  $s$  that make  $E_{a,3}(s)$  lose rank are the unobservable modes of  $(A_{11}, A_{21})$ .

The invariant zeros of  $(A, M, C)$  are given by the values of  $s$  that cause the following matrix to lose rank:

$$E_{b,1}(s) = \begin{bmatrix} sI - A & M \\ C & 0 \end{bmatrix}$$



From (11) to (12),  $E_{b,1}(s)$  is

$$E_{b,1}(s) = \left[ \begin{array}{cc|c} sI - A_1 & -A_2 & M_1 \\ -A_3 & sI - A_4 & M_2 \\ \hline 0 & T & 0 \end{array} \right]$$

Since  $T$  is orthogonal, then  $E_{b,1}(s)$  loses rank if and only if  $E_{b,2}(s)$  loses rank where

$$E_{b,2}(s) = \left[ \begin{array}{cc} sI - A_1 & M_1 \\ -A_3 & M_2 \end{array} \right]$$

Substituting for  $A_1, A_3, M_1, M_2$  from (11) to (12),  $E_{b,2}(s)$  becomes

$$E_{b,2}(s) = \left[ \begin{array}{ccc|cc} sI - A_{11} & -A_{12} & -A_{13} & 0 & 0 \\ -A_{21} & sI - A_{22} & -A_{23} & 0 & 0 \\ -A_{31} & -A_{32} & sI - A_{33} & M_{11} & 0 \\ \hline 0 & -A_{42} & -A_{43} & 0 & 0 \\ -A_{51} & -A_{52} & -A_{53} & 0 & M_{22} \end{array} \right]$$

From Lemma 4,  $[A_{42} \ A_{43}]$  has a special structure which together with the fact that  $M_{11}, M_{22}$  are square and invertible, means  $E_{b,2}(s)$  loses rank if and only if  $E_{b,3}(s)$  loses rank where

$$E_{b,3}(s) = \left[ \begin{array}{c} sI - A_{11} \\ -A_{21} \end{array} \right]$$

which loses rank if and only if  $s$  is an unobservable mode of  $(A_{11}, A_{21})$ . This shows that  $(\tilde{A}, \tilde{M}, \tilde{C})$  and  $(A, M, C)$  have the same invariant zeros.

By using the partitions of  $(A, M, C)$  in (9)–(10), it can be easily shown the Rosenbrock matrix of  $(A, M, C)$  loses rank if and only if the following matrix  $E_{c,1}(s)$  loses rank

$$E_{c,1}(s) = \left[ \begin{array}{ccc} sI - A_\Omega & 0 & 0 \\ -\tilde{Q}_{11} & sI - \tilde{A}_{11} & -\tilde{A}_{12} \\ -\tilde{Q}_{21} & -\tilde{A}_{31} & -\tilde{A}_{32} \end{array} \right]$$

It is clear that  $E_{c,1}(s)$  loses rank when  $s = \lambda(A_\Omega)$  or when  $s$  is an invariant zero of  $(\tilde{A}_{11}, \tilde{A}_{12}, \tilde{A}_{31}, \tilde{A}_{32})$ . Then, by finding the Rosenbrock matrix of  $(\tilde{A}, \tilde{M}, \tilde{C})$  using the partitions in (4)–(6), it can be proven that the invariant zeros of  $(\tilde{A}, \tilde{M}, \tilde{C})$  are the invariant zeros of  $(\tilde{A}_{11}, \tilde{A}_{12}, \tilde{A}_{31}, \tilde{A}_{32})$ .

Hence, it is proven that the invariant zeros of  $(\tilde{A}, \tilde{M}, \tilde{C})$  are the invariant zeros of  $(\tilde{A}, \tilde{M}, \tilde{C})$  and  $\lambda(A_\Omega)$ .

#### REFERENCES

1. Chen J, Patton RJ. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers: Dordrecht, 1999.
2. Frank PM. Fault diagnosis in dynamic systems using analytical and knowledge based redundancy—a survey and some new results. *Automatica* 1990; **26**:459–474.
3. Frank PM. Analytical and qualitative model-based fault diagnosis—a survey and some new results. *European Journal of Control* 1996; **2**:6–28.

4. Patton RJ, Chen J. A survey of robustness problems in quantitative model-based fault diagnosis. *Applied Maths and Computer Science* 1993; 3:339–416.
5. Magni JF, Mouyon P. On residual generation by observer and parity space approaches. *IEEE Transactions on Automatic Control* 1994; 39:441–447.
6. Shen LC, Hsu PL. Robust design of fault isolation observers. *Automatica* 1998; 34:1421–1429.
7. Yuan Z, Vansteenkiste GC, Wen C. Improving the observer-based FDI design for efficient fault isolation. *International Journal of Control* 1997; 68:197–218.
8. Edwards C, Spurgeon SK, Patton RJ. Sliding mode observers for fault detection and isolation. *Automatica* 2000; 36:541–553.
9. Edwards C, Spurgeon SK. A sliding mode observer based FDI scheme for the ship benchmark. *European Journal of Control* 2000; 6:341–356.
10. Saif M, Guan Y. A new approach to robust fault detection and identification. *IEEE Transactions on Aerospace and Electronic Systems* 1993; 29:685–695.
11. Patton RJ, Chen J. Optimal unknown input distribution matrix selection in robust fault diagnosis. *Automatica* 1993; 29:837–841.
12. Edwards C, Spurgeon SK. On the development of discontinuous observers. *International Journal of Control* 1994; 59:1211–1229.
13. Tan CP, Edwards C. Sliding mode observers for robust detection and reconstruction of actuator and sensor faults. *International Journal of Robust and Nonlinear Control* 2003; 13:443–463.
14. Boyd SP, El-Ghaoui L, Feron E, Balakrishnan V. *Linear Matrix Inequalities in Systems and Control Theory*. SIAM: Philadelphia, PA, 1994.
15. Floquet T, Barbot JP. An observability form for linear systems with unknown inputs. *International Journal of Control* 2006; 79:132–139.
16. Davila J, Fridman L, Levant A. Second-order sliding mode observer for mechanical systems. *IEEE Transactions on Automatic Control* 2005; 50:1785–1789.
17. Edwards C, Tan CP. A comparison of sliding mode and unknown input observers for fault reconstruction. *European Journal of Control* 2006; 12:245–260.
18. Rosenbrock HH. *State Space and Multivariable Theory*. Wiley: New York, 1970.
19. Gahinet P, Nemirovski A, Laub AJ, Chilali M. *LMI Control Toolbox, Users Guide*. The MathWorks, Inc.: MA, U.S.A, 1995.
20. Tan CP, Edwards C. An LMI approach for designing sliding mode observers. *International Journal of Control* 2001; 74:1559–1568.
21. Utkin VI. *Sliding Modes in Control Optimization*. Springer: Berlin, 1992.
22. Zhou K, Doyle J, Glover K. *Robust and Optimal Control*. Prentice-Hall: Englewood Cliffs, NJ, 1995.
23. Heck BS, Yallapragada SV, Fan MKH. Numerical methods to design the reaching phase of output feedback variable structure control. *Automatica* 1995; 31:275–279.
24. Chen J, Patton RJ, Zhang HZ. Design of unknown input observers and robust fault detection filters. *International Journal of Control* 1996; 63:85–105.
25. Chen J, Zhang H. Robust detection of faulty actuators via unknown input observers. *International Journal of Systems Science* 1991; 22:1829–1839.
26. Darouach M, Zasadzinski M, Xu SJ. Full order observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control* 1994; 39:606–609.