Disturbance Decoupled Fault Reconstruction using Sliding Mode Observers

Kok Yew Ng * **Chee Pin Tan** ** **Rini Akmeliawati** *
Christopher Edwards **

* *School of Engineering, Monash University, Jalan Lagoon Selatan, Bandar Sunway, 46150 Selangor Darul Ehsan, Malaysia*
** *Engineering Department, Leicester University, University Road, LE1 7RH UK*

Abstract: The objective of a robust fault reconstruction scheme is to generate an accurate reconstruction of the fault that is unaffected by disturbances. A typical method for robust fault reconstruction is to reconstruct the faults and disturbances, which is conservative and requires stringent conditions. This paper investigates and presents conditions that guarantee a fault reconstruction that rejects the effects of disturbances, which are less stringent than those of previous work. A VTOL aircraft model is used to validate the work of this paper.

Keywords: Disturbance decoupling; fault reconstruction; actuator fault; 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 in the actuators or sensors. The fundamental purpose of an FDI scheme is to generate an alarm when a fault occurs and also to determine its location. An overview of work in this area appears in (Frank (1990)). The most commonly used FDI methods are observer-based, where the measured plant output is compared to the output of an observer, and the discrepancy is used to form a residual, which is then used to decide as to whether a fault condition is present.

A useful alternative to residual generation is *fault reconstruction*, which not only detects and isolates the fault, but provides an estimate of the fault so that its shape and magnitude can be better understood and more precise corrective action can be taken (Edwards et al. (2000); Saif & Guan (1993)). However, as the fault reconstruction scheme is observer-based, it 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 which will corrupt the reconstruction, producing a nonzero reconstruction when there are no faults, or mask the effect of a fault, producing a 'zero' reconstruction in the presence of faults (Patton & Chen (1993)). Hence, the scheme needs to be designed so that the reconstruction is robust to disturbances.

Edwards et al. (2000) used a sliding mode observer from (Edwards & Spurgeon (1994)) to reconstruct faults, but there was no explicit consideration of the disturbances. Tan & Edwards (2003) built on this work and designed the sliding mode observer using the Linear Matrix Inequalities (LMIs) method in (Boyd et al. (1994)) to minimize the $\mathcal{L}_2$ gain from the disturbances to the fault reconstruction. Saif & Guan (1993) combined the faults and disturbances to form a new 'fault' vector and used a linear observer to reconstruct the new 'fault'. Although this method manages to decouple the disturbances from the fault reconstruction, very stringent conditions need to be fulfilled, and is conservative because the disturbance vector does not need to be reconstructed. Edwards & Tan (2006) compared the fault reconstruction performances of (Edwards et al. (2000)) and (Saif & Guan (1993)), and found that it was redundant to reconstruct the disturbance in order to generate a fault reconstruction that is robust to disturbance. A counter example was presented to prove this point, but the conditions for disturbance decoupling were not formally investigated.

This paper builds on the work in (Edwards & Tan (2006)). Its main contribution is the investigation of conditions that guarantee the fault reconstruction is decoupled from disturbances. The conditions that guarantee disturbance decoupling are found to be less stringent than those in (Saif & Guan (1993)) which proves that disturbance reconstruction is redundant for disturbance decoupling. In addition, the conditions in this paper are easily testable on the original system matrices, making it possible to immediately determine whether disturbance decoupled fault reconstruction is feasible. A VTOL system taken from the FDI literature will be used to validate the results in this paper. The paper is organized as follows: Section 2 introduces the system and sets up the coordinate transformation and framework for the investigation of the existence conditions; Section 3 investigates the conditions

⋆ Correspondence to: School of Engineering, Monash University, Jalan Lagoon Selatan, Bandar Sunway, 46150 Selangor DE, Malaysia, tan.chee.pin@eng.monash.edu.my

such that disturbance decoupled fault reconstruction is achieved; an example to validate the conditions is given in Section 4 and finally Section 5 makes some conclusions.

The notation used throughout this paper is quite standard; in particular $\|.\|$ represents the Euclidean norm for vectors and the induced spectral norm for matrices, and $\lambda(.)$ denotes the spectrum of a square matrix. For an arbitrary matrix H with rank m, denote by $H^\dagger$ the left-pseudo-inverse with the property that

$$H^\dagger H = H_1 \begin{bmatrix} 0 & 0 \\ 0 & I_m \end{bmatrix} H_1^T$$

where H_1 is an orthogonal matrix. The matrix $H^\dagger$ can be easily found using the Singular Value Decomposition (SVD) and is not necessarily unique. For the case when H has full column rank, then $H^\dagger H = I_m$.

2. PRELIMINARIES AND PROBLEM STATEMENT

Consider the following system

$$\dot{x}(t) = Ax(t) + Mf(t) + Q\xi(t) \quad (1)$$

$$y(t) = Cx(t) \quad (2)$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$, $f \in \mathbb{R}^q$ and $\xi \in \mathbb{R}^h$ are the states, outputs, unknown faults and unknown disturbances respectively while ξ encapsulates all nonlinearities and unknowns in the system. Assume $rank(M) = q$, $rank(Q) = h$ and $rank(C) = p$. Also assume that $p > q$.

The main goal is to reconstruct f whilst being robust to ξ. The robust fault reconstruction scheme in (Tan & Edwards (2003)) minimized the effect of ξ on the fault reconstruction using Linear Matrix Inequalities if and only if the following conditions are satisfied

A1. $rank(CM) = rank(M)$
A2. (A, M, C) is minimum phase.

Their method however, does not fully reject the effects of ξ. Saif & Guan (1993) managed to reject the effects of the ξ by combining f and ξ vectors and then designed a fault reconstruction scheme to reconstruct this new 'fault'. The necessary and sufficient conditions for their scheme are

B1. $rank(C\,[\,M\ Q\,]) = rank\,[\,M\ Q\,]$
B2. $(A, [\,M\ Q\,], C)$ is minimum phase.

Since only the reconstruction of f is required, therefore reconstructing the new 'fault' is conservative. Therefore, this paper investigates the conditions that guarantee disturbance decoupled fault reconstruction with less stringent conditions than (Saif & Guan (1993)). Firstly, define $k = rank(CQ)$ where $k \le h$ and assume

N0. $rank(CM) = rank(M)$
N1. $rank\,(C\,[\,M\ Q\,]) = rank(CM) + rank(CQ)$

Notice that N1 implies $p > q + k$. The implication of this assumption will be discussed later in the paper.

Proposition 1. If N0 and N1 hold then there exist nonsingular linear transformations $x \mapsto T_2 x$, $\xi \mapsto T_1^{-1}\xi$ such that (A, M, C, Q) when partitioned have the structure

$$A = \overset{\;\;n\text{-}p\;\;\;\;p}{\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}} \begin{matrix} \updownarrow n\text{-}p \\ \updownarrow p \end{matrix}, M = \overset{q}{\begin{bmatrix} 0 \\ M_2 \end{bmatrix}}, C = \overset{n\text{-}p\;\;\;p}{[\,0\ \ C_2\,]}, Q = \overset{h}{\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}} \quad (3)$$

where $M_2 = \begin{bmatrix} 0 \\ M_o \end{bmatrix} \begin{matrix} \updownarrow p\text{-}q \\ \updownarrow q \end{matrix}$ with M_o, C_2 being invertible. Further partition Q_1, Q_2 to be

$$Q_1 = \begin{bmatrix} \bar{Q}_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} \updownarrow h\text{-}k \\ \updownarrow n\text{-}p\text{-}h\text{+}k \end{matrix}, Q_2 = \begin{bmatrix} 0 & 0 \\ 0 & \bar{Q}_2 \\ 0 & 0 \end{bmatrix} \begin{matrix} \updownarrow p\text{-}q\text{-}k \\ \updownarrow k \\ \updownarrow q \end{matrix} \quad (4)$$

where $\bar{Q}_1$,$\bar{Q}_2$ are square and invertible.

Proof. From (Edwards & Spurgeon (1994)), since CM is full rank, there exists a change of coordinates such that (A, M, C) can be written as

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 \\ \tilde{A}_3 & \tilde{A}_4 \end{bmatrix}, \tilde{M} = \begin{bmatrix} 0 \\ \tilde{M}_2 \end{bmatrix}, \tilde{C} = [\,0\ \ T\,], \tilde{Q} = \begin{bmatrix} \tilde{Q}_1 \\ \tilde{Q}_2 \end{bmatrix} \quad (5)$$

where $\tilde{M}_2 = \left[\,0\ \ \tilde{M}_o^T\,\right]^T$ with $\tilde{M}_o$ and T being full rank.

Let $T_1 \in \mathbb{R}^{h \times h}$ be an orthogonal matrix such that

$$\tilde{Q}_2 T_1 = \overset{\;h\text{-}k\;\;\;\;\;k}{[\,0\quad \tilde{Q}_{22}\,]} \updownarrow p \quad (6)$$

where $rank(\tilde{Q}_{22}) = k$. This follows since by assumption, $CQ = T\tilde{Q}_2$ is rank k. Therefore, $\tilde{Q}T_1$ can be partitioned to have the structure of

$$\tilde{Q}T_1 = \overset{\;h\text{-}k\;\;\;\;\;k}{\begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ 0 & \tilde{Q}_{22} \end{bmatrix}} \begin{matrix} \updownarrow n\text{-}p \\ \updownarrow p \end{matrix} \quad (7)$$

Using the partitions in (5), N1 results in

$$rank\left(T\left[\,\tilde{M}_2\ \ 0\ \ \tilde{Q}_{22}\,\right]\right) = q + k \quad (8)$$

Since T is orthogonal, using (6) and (7) result in

$$rank\left(\left[\,\tilde{M}_2\ \ 0\ \ \tilde{Q}_{22}\,\right]\right) = q + k \quad (9)$$

The expanded structure of (7) according to (9) is

$$\tilde{Q}T_1 = \overset{\;h\text{-}k\;\;\;\;\;k}{\begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ 0 & \tilde{Q}_{221} \\ 0 & \tilde{Q}_{222} \end{bmatrix}} \begin{matrix} \updownarrow n\text{-}p \\ \updownarrow p\text{-}q \\ \updownarrow q \end{matrix} \quad (10)$$

where $\tilde{Q}_{221}$ and $\tilde{Q}_{222}$ are appropriate partitions of $\tilde{Q}_{22}$. Hence, $\tilde{Q}_{221} \in \mathbb{R}^{(p-q)\times k}$ has full column rank k which means there exists a matrix $\tilde{Q}_{221}^\dagger$ such that $\tilde{Q}_{221}^\dagger \tilde{Q}_{221} = I_k$.

Let X_1 and X_2 be orthogonal matrices such that

$$X_1\tilde{Q}_{11} = \overset{h\text{-}k}{\begin{bmatrix} \bar{Q}_1 \\ 0 \end{bmatrix}} \begin{matrix} \updownarrow h\text{-}k \\ \updownarrow n\text{-}p\text{-}h\text{+}k \end{matrix}\quad X_2\tilde{Q}_{221} = \overset{k}{\begin{bmatrix} 0 \\ \bar{Q}_2 \end{bmatrix}} \begin{matrix} \updownarrow p\text{-}q\text{-}k \\ \updownarrow k \end{matrix} \quad (11)$$

Now define a nonsingular change of coordinates represented by the matrix T_2 where

$$T_2 := \overset{\;n\text{-}p\;\;\;\;\;\;\;\;\;\;p\text{-}q\;\;\;\;\;\;\;\;\;\;q}{\begin{bmatrix} X_1 & -X_1\tilde{Q}_{12}\tilde{Q}_{221}^\dagger & 0 \\ 0 & X_2 & 0 \\ 0 & -\tilde{Q}_{222}\tilde{Q}_{221}^\dagger & I_q \end{bmatrix}} \quad (12)$$

such that the matrices $\tilde{Q}, \tilde{C}$ in their new coordinates have the following structure

$$T_2\tilde{Q}T_1 = \left[\begin{array}{cc} \bar{Q}_1 & 0 \\ 0 & 0 \\ \hline 0 & 0 \\ 0 & \bar{Q}_2 \\ 0 & 0 \end{array}\right] \begin{array}{l} \updownarrow h\text{-}k \\ \updownarrow n\text{-}p\text{-}h\text{+}k \\ \updownarrow p\text{-}q\text{-}k \\ \updownarrow k \\ \updownarrow q \end{array} , \quad \tilde{C}{T_2}^{-1} = \left[\, 0 \;\; TY_1^{-1} \,\right] \tag{13}$$

where $Y_1 = \begin{bmatrix} X_2 & 0 \\ -\tilde{Q}_{222}\tilde{Q}_{221}^{\dagger} & I_q \end{bmatrix}$ and TY_1^{-1} is invertible. The coordinate transformation T_2 does not alter $\tilde{M}$. Equating $M_o = \tilde{M}_o$ and $C_2 = TY_1^{-1}$, the proof is thus complete. ■

2.1 A sliding mode observer for fault reconstruction

A sliding mode observer from (Edwards & Spurgeon (1994)) for the system (1) - (2) is

$$\dot{\hat{x}}(t) = A\hat{x}(t) - G_l e_y(t) + G_n\nu \;, \quad \hat{y}(t) = C\hat{x}(t) \tag{14}$$

where $\hat{x} \in \mathbb{R}^n$ is the estimate of x and $e_y = \hat{y} - y$ is the output estimation error. The term ν is defined by

$$\nu = -\rho\frac{e_y}{||e_y||}, \quad e_y \neq 0, \quad \rho \in \mathbb{R}_+ \tag{15}$$

where ρ is an upper bound of f and ξ. The matrices $G_l, G_n \in \mathbb{R}^{n\times p}$ are the observer gains to be designed. In the coordinates of (3) in Proposition 1, G_n is assumed to have the structure

$$G_n = \begin{bmatrix} -L \\ I_p \end{bmatrix} C_2^{-1}P_o^{-1} \tag{16}$$

where $P_o \in \mathbb{R}^{p\times p}$ is symmetric positive definite (s.p.d.) and $L = [\, L_o \;\, 0 \,] \in \mathbb{R}^{(n-p)\times p}$ with $L_o \in \mathbb{R}^{(n-p)\times(p-q)}$.

Define the state estimation error as $e := \hat{x} - x$. Therefore, by combining (1) - (2) and (14), the state estimation error system can be expressed as

$$\dot{e}(t) = (A - G_lC)e(t) + G_n\nu - Mf(t) - Q\xi(t) \tag{17}$$

Introduce a change of coordinates such that $e_L := col(e_1, e_y) = T_e e$ where

$$T_e = \begin{bmatrix} I_{n-p} & L \\ 0 & C_2 \end{bmatrix} \tag{18}$$

Hence, (A, M, C, Q) in the new coordinates are

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 \\ \mathcal{A}_3 & \mathcal{A}_4 \end{bmatrix}, \; \mathcal{M} = \begin{bmatrix} 0 \\ \mathcal{M}_2 \end{bmatrix}, \; \mathcal{C} = [\, 0 \;\; I_p \,], \; \mathcal{Q} = \begin{bmatrix} \mathcal{Q}_1 \\ \mathcal{Q}_2 \end{bmatrix} \tag{19}$$

where $\mathcal{A}_1 = A_1 + LA_3,\; \mathcal{A}_3 = C_2A_3,\; \mathcal{M}_2 = C_2M_2,\; \mathcal{Q}_1 = Q_1 + LQ_2,\; \mathcal{Q}_2 = C_2Q_2$ and G_n will become

$$\mathcal{G}_n = \begin{bmatrix} 0 \\ P_o^{-1} \end{bmatrix} \tag{20}$$

Therefore, the state estimation error equation after the coordinates transformation is

$$\dot{e}_L(t) = (\mathcal{A} - \mathcal{G}_l\mathcal{C})e_L(t) + \mathcal{G}_n\nu - \mathcal{M}f(t) - \mathcal{Q}\xi(t) \tag{21}$$

Lemma 2. From (Tan & Edwards (2003)), if there exists a value of G_l to satisfy $P(A - G_lC) + (A - G_lC)^TP < 0$ where

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} > 0 \quad \text{with} \quad P_{12} = [\underbrace{P_{121}}_{p\text{-}q} \;\; 0] \updownarrow n\text{-}p$$

then if $P_o = P_{22} - P_{12}^TP_{11}^{-1}P_{12}$, and for a large enough choice of ρ, an ideal sliding motion takes place on $\mathcal{S} = \{e : Ce = 0\}$ in finite time. □

Assume that the observer (14) has been designed and a sliding motion has been achieved ($\dot{e}_y = e_y = 0$). Then, (21) can be partitioned according to (19) - (20) as

$$\dot{e}_1(t) = (A_1 + LA_3)e_1(t) - (Q_1 + LQ_2)\xi(t) \tag{22}$$

$$0 = C_2A_3e_1(t) + P_o^{-1}\nu_{eq} - C_2M_2f(t) - C_2Q_2\xi(t) \tag{23}$$

where ν_{eq} is the equivalent output error injection term required to maintain the sliding motion and can be approximated to any degree of accuracy by replacing ν with

$$\nu = -\rho\frac{e_y}{||e_y|| + \delta} \tag{24}$$

where δ is a small positive scalar. Since e_y is a measurable signal, therefore ν_{eq} can be computed online (Edwards et al. (2000); Edwards & Spurgeon (2000)) for full details.

To reconstruct the fault, define a fault reconstruction signal to be $\hat{f}(t) := WC_2^{-1}P_o^{-1}\nu_{eq}$ where $W := [\, W_1 \;\; M_o^{-1} \,]$ with $W_1 \in \mathbb{R}^{q\times(p-q)}$. Then define the fault reconstruction error signal $e_f = \hat{f} - f$. From (23) and $\hat{f}$ it becomes

$$e_f(t) = -WA_3e_1(t) + WQ_2\xi(t) \tag{25}$$

It is desired that $e_f = 0 \;\; (\hat{f} = f)$. From (22) and (25), it is clear that ξ is an excitation signal of e_f and that the design freedom is represented by L_o and W_1. Hence the goal is to decouple e_f from ξ by choice of L_o and W_1. Define Q_a to be the left $h - k$ columns of Q in (3). The following theorem states the main result of this paper.

Theorem 3. Suppose N0 and N1 hold. Then e_f will be decoupled from ξ by appropriate choice of L_o and W_1 if the following conditions are satisfied

C1. $rank\,[\, CAQ_a \;\, CM \;\, CQ \,] - rank(M) - rank(CQ) = rank\,[\, AQ_a \;\, Q \,] - rank(Q)$
C2. $(A, [\, M \;\, Q \,], C)$ is minimum phase.

Note that C1 and C2 are easily testable conditions onto the original system matrices. In addition C1 is not as stringent as B1 but the disturbance could still be decoupled from the fault reconstruction. The following section provides a constructive proof of Theorem 1.

3. DISTURBANCE DECOUPLED FAULT RECONSTRUCTION

In order to make e_f completely decoupled from ξ, a necessary condition is

$$WQ_2 = 0 \tag{26}$$

since it represents the direct feedthrough component in the system formed from (22) and (25).

Partition Q_2 from (3) and (13) such that

$$Q_2 = \begin{bmatrix} Q_{21} \\ Q_{22} \end{bmatrix} \begin{array}{l} \updownarrow p\text{-}q \\ \updownarrow q \end{array} = \left[\begin{array}{cc} 0 & 0 \\ 0 & \bar{Q}_2 \\ \hline 0 & 0 \end{array}\right] \tag{27}$$

which results in $WQ_2 = W_1Q_{21}$.

Thus the necessary and sufficient condition to ensure (26) holds is that $W_1Q_{21} = 0$.

Lemma 4. A general solution for W_1 that will satisfy $W_1Q_{21} = 0$ (and hence making $WQ_2 = 0$) is given by

$$W_1 = W_{12}(I_{p-q} - Q_{21}Q_{21}^{\dagger}) \tag{28}$$

where

$$Q_{21}^{\dagger} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{Q}_2^{-1} \end{bmatrix}$$

and $W_{12} \in \mathbb{R}^{q \times (p-q)}$ is design freedom such that

$$W_{12} = [\underbrace{W_{121}}_{p\text{-}q\text{-}k} \ \underbrace{W_{122}}_{k}]$$

Proof. This is straightforward; substituting (28) and (27) into WQ_2 results in (26) being satisfied. ■

Remark 1. If N1 is not satisfied, then $rank\,(C\,[\,M\ \ Q\,]) < rank(CM) + rank(CQ)$. Since $rank(CQ) = k$ if N1 is not satisfied, then from M in (3), it is clear that $rank(Q_{21}) < rank(Q_2) =: k$. Expanding (26) according to W gives $WQ_2 = W_1Q_{21} + M_o^{-1}Q_{22}$. In order to satisfy (26),

$$W_1Q_{21} = -M_o^{-1}Q_{22} \tag{29}$$

must be satisfied. To satisfy (29) by choice of W_1 requires

$$rank(Q_{21}) = rank\,\big[\,Q_{21}^T\ \ Q_{22}^T\,\big]^T = rank(Q_2) \tag{30}$$

which is a contradiction. Therefore, N1 is a necessary condition in order for (26) to be satisfied.

Lemma 4 has shown that N1 is a sufficient condition for (26) to be satisfied as it makes the coordinate transformation in Proposition 1 feasible. Therefore, N1 is a necessary and sufficient condition for (26) to be satisfied. ♯

From the solution in (28), W_1 is constrained to be

$$W_1 = W_{12}(I_{p-q} - Q_{21}Q_{21}^{\dagger}) = [\,W_{121}\ \ 0\,] \tag{31}$$

Partition A_1, A_3, L_o from (3) and (2.1) as

$$A_1 = \begin{bmatrix} A_{11} & A_{12} \\ A_{13} & A_{14} \end{bmatrix} \begin{matrix} \updownarrow h\text{-}k \\ \updownarrow n\text{-}p\text{-}h\text{+}k \end{matrix}, \quad A_3 = \begin{bmatrix} A_{31} & A_{32} \\ A_{33} & A_{34} \\ A_{35} & A_{36} \end{bmatrix} \begin{matrix} \updownarrow p\text{-}k\text{-}q \\ \updownarrow k \\ \updownarrow q \end{matrix}$$

$$L_o = \begin{bmatrix} \underbrace{L_{11}}_{p\text{-}k\text{-}q} & \underbrace{L_{12}}_{k} \\ L_{21} & L_{22} \end{bmatrix} \tag{32}$$

Then, from W_1 in (31), WA_3 can be written as

$$WA_3 = \big[\,W_{121}A_{31} + M_o^{-1}A_{35}\ \ W_{121}A_{32} + M_o^{-1}A_{36}\,\big] \tag{33}$$

The terms $(A_1 + LA_3)$ and $(Q_1 + LQ_2)$ in (22), when expressed using the system partitions in (32) will produce

$$A_1{+}LA_3{=}\begin{bmatrix} A_{11}{+}L_{11}A_{31}{+}L_{12}A_{33} & A_{12}{+}L_{11}A_{32}{+}L_{12}A_{34} \\ A_{13}{+}L_{21}A_{31}{+}L_{22}A_{33} & A_{14}{+}L_{21}A_{32}{+}L_{22}A_{34} \end{bmatrix}$$

$$Q_1{+}LQ_2{=}\begin{bmatrix} \bar{Q}_1 & L_{12}\bar{Q}_2 \\ 0 & L_{22}\bar{Q}_2 \end{bmatrix}$$

Therefore, from W_1 in (31) and L_o in (32), the system (22) and (25) form the following state space system

$$\begin{bmatrix} \dot{e}_{11}(t) \\ \dot{e}_{12}(t) \end{bmatrix} = \underbrace{(A_1 + LA_3)}_{\bar{A}} \begin{bmatrix} e_{11}(t) \\ e_{12}(t) \end{bmatrix} - \underbrace{(Q_1 + LQ_2)}_{\bar{B}} \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} \tag{34}$$

$$e_f(t) = \underbrace{(-WA_3)}_{\bar{C}} \begin{bmatrix} e_{11}(t) \\ e_{12}(t) \end{bmatrix} \tag{35}$$

Expressing (34) - (35) in terms of the triple $(\bar{A}, \bar{B}, \bar{C})$ the state space matrices will have the form

$$\bar{A}{=}\begin{bmatrix} \underbrace{\bar{A}_1}_{h\text{-}k} & \underbrace{\bar{A}_2}_{n\text{-}p\text{-}h\text{+}k} \\ \bar{A}_3 & \bar{A}_4 \end{bmatrix} \begin{matrix} \updownarrow h\text{-}k \\ \updownarrow n\text{-}p\text{-}h\text{+}k \end{matrix}, \bar{B}{=}\begin{bmatrix} \underbrace{\bar{B}_1}_{h\text{-}k} & \underbrace{\bar{B}_2}_{k} \\ 0 & \bar{B}_4 \end{bmatrix}, \bar{C}{=}[\underbrace{\bar{C}_1}_{h\text{-}k}\ \ \underbrace{\bar{C}_2}_{n\text{-}p\text{-}h\text{+}k}] \updownarrow q \tag{36}$$

It is obvious that e_{11} will be affected by ξ_1 because $\bar{B}_1$ is full rank. However, e_{12} can be decoupled from ξ by setting $\bar{B}_4 = 0$ $(L_{22} = 0)$ and $\bar{A}_3 = 0$. In order for e_f not to be affected by e_{11} (and subsequently ξ), it is essential to make $\bar{C}_1 = 0$. Making $\bar{A}_3 = 0$ and $\bar{C}_1 = 0$ respectively requires

$$rank(A_{31}) = rank \begin{bmatrix} A_{13} \\ A_{31} \end{bmatrix} \ \text{ and } \ rank(A_{31}) = rank \begin{bmatrix} A_{31} \\ A_{35} \end{bmatrix}$$

since M_o is nonsingular.

Therefore, combining the requirements that satisfy both $\bar{A}_3 = 0$ and $\bar{C}_1 = 0$ requires

E1. $rank(A_{31}) = rank\,\big[\,A_{13}^T\ \ A_{31}^T\ \ A_{35}^T\,\big]^T$

From (35), to make $\bar{C}_1 = 0 \Rightarrow -W_{121}A_{31} - M_o^{-1}A_{35} = 0$. Hence, a general solution for W_{121} that satisfies $\bar{C}_1 = 0$ is

$$W_{121} = -M_o^{-1}A_{35}A_{31}^{\dagger} + W_{1211}(I - A_{31}A_{31}^{\dagger}) \tag{37}$$

where W_{1211} is design freedom.

Lemma 5. The condition E1 is satisfied if and only if

$$\begin{aligned} rank\,[\,CAQ_a\ \ CM\ \ CQ\,] - rank\,[\,CM\ \ CQ\,] \\ = rank\,[\,AQ_a\ \ Q\,] - rank(Q) \end{aligned} \tag{38}$$

Proof. See §A in the appendix. ■

Since it has been assumed in §2 that $rank(CM) = rank(M)$, then Condition N1 becomes

$$rank\,(C\,[\,M\ \ Q\,]) = rank(M) + rank(CQ) \tag{39}$$

Substituting this result into (38) in Lemma 5 yields

$$\begin{aligned} rank\,[\,CAQ_a\ \ CM\ \ CQ\,] - rank(M) - rank(CQ) \\ = rank\,[\,AQ_a\ \ Q\,] - rank(Q) \end{aligned} \tag{40}$$

which corresponds to Condition C1.

Substitute $L_{22} = 0$ into (34) to get $\bar{A}_3 = A_{13} + L_{21}A_{31}$ and choosing

$$L_{21} = -A_{13}A_{31}^{\dagger} + L_{211}(I - A_{31}A_{31}^{\dagger}) \tag{41}$$

where L_{211} is design freedom, makes $\bar{A}_3 = 0$. As a result,

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & \bar{A}_2 \\ 0 & \bar{A}_4 \end{bmatrix} \tag{42}$$

In order for $\bar{A}$ to be stable, $\bar{A}_1$ and $\bar{A}_4$ have to be stable. Using (41), $\bar{A}_4$ from (34) and (36) can be written as

$$\bar{A}_4 = A_{14} - A_{13}A_{31}^{\dagger}A_{32} + L_{211}(I - A_{31}A_{31}^{\dagger})A_{32} \tag{43}$$

So for $\bar{A}_4$ to be stable, $\big(A_{14} - A_{13}A_{31}^{\dagger}A_{32}, (I - A_{31}A_{31}^{\dagger})A_{32}\big)$ must be detectable. Likewise $\bar{A}_1$ can be written as

$$\bar{A}_1{=}A_{11} + L_{11}A_{31} + L_{12}A_{33}{=}A_{11} + [L_{11}\ \ L_{12}] \begin{bmatrix} A_{31} \\ A_{33} \end{bmatrix} \tag{44}$$

This implies that $\left(A_{11}, \begin{bmatrix} A_{31} \\ A_{33} \end{bmatrix}\right)$ has to be detectable if $\bar{A}_1$ is to be stable.

Proposition 6. $\big(A_{14} - A_{13}A_{31}^{\dagger}A_{32}, (I - A_{31}A_{31}^{\dagger})A_{32}\big)$ and $\left(A_{11}, \begin{bmatrix} A_{31} \\ A_{33} \end{bmatrix}\right)$ are detectable if $(A, [\,M\ \ Q\,], C)$ is minimum phase.

Proof. See §B in the appendix. ■

Proposition 6 matches Condition C2, guaranteeing a stable sliding motion and the proof of Theorem 1 is complete. □

Remark 3.6 of (Edwards & Tan (2006)) provided a counter-example where B1 is not satisfied but it is still possible to reconstruct the fault robustly. In the notation of (1) - (2), the matrices that describe the example are

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}, M = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, Q = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (45)$$

C1 and C2 are satisfied for the matrices in (45) above.

Remark 2. Note that C1 is less conservative than B1, and hence can be applied to a wider class of systems. Condition B1 implies that $Q_a = \emptyset$ (empty matrix), which will satisfy C1. However, the converse is not necessarily true. ♯

Remark 3. There have been efforts to generate disturbances decoupled fault detection residuals using linear observers in (Xiong & Saif (2000)) and (Patton & Chen (2000)) which respectively utilize the Special Coordinate Basis and Eigenstructure Assignment. However, their methods required certain elements in the matrix A to be zero. From the analysis in this paper, no such condition is required; the only requirement on the matrix A is that E1 is satisfied. Hence, this paper has also shown how the conditions for robust fault reconstruction using sliding mode observer is less stringent than if linear observers are used. ♯

4. AN EXAMPLE

The method proposed in this paper will be verified using a VTOL aircraft model taken from (Saif & Guan (1993)) where the states are the horizontal velocity, vertical velocity, pitch rate and pitch angle. The inputs are the collective pitch control and the longitudal cyclic pitch control. Assume that the horizontal and vertical velocities and the pitch angle are measurable and that both inputs are faulty. Thus,

$$B = M = \begin{bmatrix} 0.4422 & 0.1761 \\ 3.5446 & -7.5922 \\ -5.5200 & 4.4900 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Suppose that the matrix A is imprecisely known and that there exists parametric uncertainty such that

$$\dot{x} = (A + \Delta A)x + Bu + Mf \quad (46)$$

where ΔA is the discrepancy between the known matrix A and its actual value. Due to the nature of the state equations, any parametric uncertainty will appear in the third row of A. Writing (46) in the framework of (1) using

$$\Delta A x = Q\xi = \underbrace{[\,0\ \ 0\ \ 1\ \ 0\,]^T}_{Q} \underbrace{[\,0\ \ 0.5\ \ 0\ \ 2\,]\,x}_{\xi}$$

The coordinate transformation in Proposition 1 shows that $\bar{Q}_1 = -1, A_{11} = 1.6459$ and $\bar{Q}_2, A_{12}, A_{13}$ and A_{14} are all empty matrices. Also, $Q_2 = 0_{3\times 1}$ and from (27), it is easy to see that $Q_{21} = 0$. Choosing $W_{1211} = [\,-0.75\ \ 1\,]^T$ and substituting into (37) and then into (31), yields $W_1 = [\,-0.7475\ \ 0.3824\,]^T$ which satisfies (26).

It can be shown that C1 and C2 are satisfied, hence it is possible to obtain a fault reconstruction that is decoupled from the disturbances. Since $rank(Q) = 1$ and $rank(M) = 2$ and $rank\,(C\,[\,M\ \ Q\,]) = 2$, then B1 is not satisfied and hence the method in (Saif & Guan (1993)) cannot be used.

In the following simulation, the parameters associated with ν were chosen as $\rho = 50$, $\delta = 0.001$. The faults were induced in both actuators and Figure 1 shows the faults and their reconstructions. It can be seen that $\hat{f}$ provides accurate estimates of f that are independent of ξ.

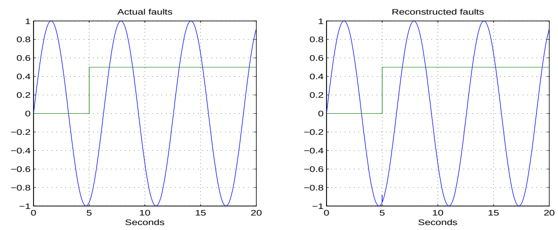

Fig. 1. The left subfigure is f, the right subfigure is $\hat{f}$.

5. CONCLUSION

This paper has investigated and presented conditions that guarantee disturbance decoupled fault reconstruction, which are easily testable onto the original system matrices. In previous work, the effects of the disturbances on the fault reconstruction were only minimized, and conditions that guarantee disturbance decoupling were not known. Other work has achieved disturbance decoupling by reconstructing the disturbances together with the fault, but this requires stringent conditions to be fulfilled and the disturbance reconstruction is not needed for analysis. This paper has successfully investigated the conditions such that the fault reconstruction is decoupled from the disturbance, and it was found that the conditions are less stringent compared to earlier work. A VTOL aircraft model was used to validate the proposed method.

REFERENCES

S.P. Boyd, L. El-Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Systems and Control Theory.* SIAM: Philadelphia, 1994.

C. Edwards and C.P. Tan. A comparison of sliding mode and unknown input observers for fault reconstruction. *European Journal of Control*, 12:245–260, 2006.

C. Edwards and S.K. Spurgeon. A sliding mode observer based FDI scheme for the ship benchmark. *European Journal of Control*, 6:341–356, 2000.

C. Edwards, S.K. Spurgeon, and R.J. Patton. Sliding mode observers for fault detection and isolation. *Automatica*, 36:541–553, 2000.

C. Edwards and S.K. Spurgeon. On the development of discontinuous observers. *Int. Journal of Control*, 59:1211–1229, 1994.

P.M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge based redundancy - a survey and some new results. *Automatica*, 26:459–474, 1990.

R.J. Patton and J. Chen. Optimal unknown input distribution matrix selection in robust fault diagnosis. *Automatica*, 29:837–841, 1993.

R.J. Patton and J. Chen. On eigenstructure assignment for robust fault diagnosis. *International Journal of Robust and Nonlinear Control*, 10:1193–1208, 2000.

M. Saif and Y. Guan. A new approach to robust fault detection and identification. *IEEE Trans. Aerospace and Electronic Systems*, 29:685–695, 1993.

C.P. Tan and C. Edwards. Sliding mode observers for robust detection and reconstruction of actuator and sensor faults. *Int. Journal of Robust and Nonlinear Control*, 13:443–463, 2003.

Y. Xiong and M. Saif. Robust fault detection and isolation via a diagnostic observer. *International Journal of Robust and Nonlinear Control*, 10:1175–1192, 2000.

Appendix A. PROOF OF LEMMA 5

From the partitions in (3) - (4) and (32) as well as the structure of Q_a in Theorem 3, it can be shown that

$$CAQ_a = C_2 \begin{bmatrix} A_{31}\bar{Q}_1 \\ A_{33}\bar{Q}_1 \\ A_{35}\bar{Q}_1 \end{bmatrix} \tag{A.1}$$

Using the partitions of M, Q from (3) - (4) results in

$$rank[CAQ_a\ CM\ CQ] - rank[CM\ CQ] = rank(A_{31}) \tag{A.2}$$

since C_2, M_o and $\bar{Q}_2$ are full rank and invertible.

Also, it can be shown that

$$AQ_a = \begin{bmatrix} A_{11}\bar{Q}_1 \\ A_{13}\bar{Q}_1 \\ A_{31}\bar{Q}_1 \\ A_{33}\bar{Q}_1 \\ A_{35}\bar{Q}_1 \end{bmatrix} \tag{A.3}$$

As a result, the right hand side of (38) becomes

$$rank\,[\,AQ_a\ \ Q\,] - rank(Q) = rank \begin{bmatrix} A_{13}\bar{Q}_1 \\ A_{31}\bar{Q}_1 \\ A_{35}\bar{Q}_1 \end{bmatrix} \tag{A.4}$$

Since $\bar{Q}_1$ is invertible, (38) is equivalent to

$$rank(A_{31}) = rank\,[\,A_{13}^T\ A_{31}^T\ A_{35}^T\,]^T \tag{A.5}$$

which corresponds to E1 and the proof is complete. ■

Appendix B. PROOF OF PROPOSITION 6

From (Edwards & Spurgeon (1994)), the Rosenbrock system matrix associated with $(A, [\,M\ Q\,], C)$ is

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

and the zeros of the system are the values of s that cause its Rosenbrock matrix to lose normal rank. From (3) and (32), it can be shown that $E_{a,1}(s)$ loses rank if and only if $E_{a,2}(s)$ loses rank where

$$E_{a,2}(s) := \begin{bmatrix} sI - A_{14} & -A_{13} \\ -A_{32} & -A_{31} \end{bmatrix} \tag{B.2}$$

Hence, the invariant zeros of $(A, [\,M\ Q\,], C)$ are the invariant zeros of $(A_{14}, A_{13}, A_{32}, A_{31})$.

Define $r = rank(A_{31})$, by a singular value decomposition

$$A_{31} = R_1 \begin{bmatrix} 0 & 0 \\ 0 & A_{312} \end{bmatrix} R_2 \tag{B.3}$$

where R_1, R_2 are orthogonal matrices and $A_{312} \in \mathbb{R}^{r\times r}$ is invertible. Define

$$A_{31}^{\dagger} = R_2^T \begin{bmatrix} 0 & 0 \\ 0 & A_{312}^{-1} \end{bmatrix} R_1^T$$

Since by assumption $rank(A_{31}) = rank\,[\,A_{13}^T\ A_{31}^T\,]^T$ then the structure of A_{31} in (B.3) results in

$$A_{13} = [\,0\ A_{132}\,]\,R_2 \tag{B.4}$$

where $A_{132} \in \mathbb{R}^{(n-p-h+k)\times r}$. Partition

$$A_{32} = R_1 \begin{bmatrix} A_{321} \\ A_{322} \end{bmatrix} \begin{matrix} \updownarrow p\text{-}q\text{-}k\text{-}r \\ \updownarrow r \end{matrix} \tag{B.5}$$

Substituting (B.3) - (B.5) into (B.2) and pre-multiply it with T_z where

$$T_z = \begin{bmatrix} I_{n-p-h+k} & 0 & -A_{132}A_{312}^{-1} \\ 0 & I_{p-q-k-r} & 0 \\ 0 & 0 & I_r \end{bmatrix} \tag{B.6}$$

then it is clear that $E_{a,2}(s)$ loses normal rank when the following matrix pencil loses rank

$$E_{a,3}(s) = \begin{bmatrix} sI - (A_{14} - A_{132}A_{312}^{-1}A_{322}) \\ A_{321} \end{bmatrix} \tag{B.7}$$

From the Popov-Belevitch-Hautus (PBH) rank test (Edwards & Spurgeon (1994)), the values of s that cause $E_{a,3}(s)$ to lose rank are the unobservable modes of $(A_{14} - A_{132}A_{312}^{-1}A_{322}, A_{321})$. Therefore, the invariant zeros of $(A_{14}, A_{13}, A_{32}, A_{31})$ are the unobservable modes of $(A_{14} - A_{132}A_{312}^{-1}A_{322}, A_{321})$.

Now evaluate the pair $\left(A_{14} - A_{13}A_{31}^{\dagger}A_{32}, (I - A_{31}A_{31}^{\dagger})A_{32}\right)$ using the new structures of A_{31}, A_{13} and A_{32} introduced in (B.3) - (B.5). It is easy to verify that

$$A_{14} - A_{13}A_{31}^{\dagger}A_{32} = A_{14} - A_{132}A_{312}^{-1}A_{322} \tag{B.8}$$

$$(I - A_{31}A_{31}^{\dagger})A_{32} = R_1 \begin{bmatrix} A_{321} \\ 0 \end{bmatrix} \tag{B.9}$$

Therefore, if $(A, [\,M\ Q\,], C)$ is minimum phase, then $\left(A_{14} - A_{13}A_{31}^{\dagger}A_{32}, (I - A_{31}A_{31}^{\dagger})A_{32}\right)$ is detectable ♯.

If $(A, [\,M\ Q\,], C)$ is minimum phase, then (A, C) is detectable. Performing the PBH rank test on (A, C) and expanding in the coordinates of (3), then the detectability of (A, C) implies the detectability of (A_1, A_3), which (by expanding further in the coordinates of (32)) further implies the detectability of $\left(A_{11}, [\,A_{13}^T\ A_{31}^T\ A_{33}^T\ A_{35}^T\,]^T\right)$.

However, since by assumption the condition in E1 holds, the detectability of $\left(A_{11}, [\,A_{13}^T\ A_{31}^T\ A_{33}^T\ A_{35}^T\,]^T\right)$ implies that $\left(A_{11}, \begin{bmatrix} A_{31} \\ A_{33} \end{bmatrix}\right)$ is detectable.

Hence, if $(A, [\,M\ Q\,], C)$ is minimum phase, then $\left(A_{11}, \begin{bmatrix} A_{31} \\ A_{33} \end{bmatrix}\right)$ is detectable and the proof is complete. ■