JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 25, No. 2, March–April 2002

Robust Fault Detection for Uncertain Discrete-Time Systems

Haibo Wang and James Lam
University of Hong Kong, Hong Kong, People's Republic of China

The fault detection problem under structured uncertainties in the system matrices is considered. The sensitivity of robust fault detection is one of the important issues considered in the fault detection and isolation development. To enhance this characteristic, an unconstrained optimization approach is taken to design a robust fault detection observer. The approach aims at enhancing the fault detection robustness to uncertainties without sacrificing the fault detection sensitivity, which was seldom addressed before. Furthermore, other objectives related to the observer gain and the eigenstructure conditioning of the observer system are also taken into account. The gradient-based optimization approach is facilitated by the explicit gradient expressions derived. Numerical simulation has also demonstrated the tradeoffs between different objectives as well as the effectiveness of the present methodology.

I. Introduction

THE research and application of fault detection and isolation in automated processes has received considerable attention during the last two decades,[1–6] both in a research context and also in the domain of application studies on real processes. One area of active research is the development of model-based fault detection systems. There are a great variety of methods in the literature to construct model-based fault detection systems. One of them, using observer techniques, has received much attention in the past.

Modern technology has increasingly led to the creation of highly complex dynamic systems that have demanding performance requirements in a variety of environments. These systems must be capable of meeting stringent specifications for reliability and operational safety over long periods of time, while operating under a great deal of uncertainty. However, in practice, such as in chemical processes or aerospace systems, fully accurate mathematical models of the systems cannot be obtained. There is always a mismatch between the actual process and its mathematical model even if there is no fault in the process. Such inaccuracies may give rise to false alarms and, thus, corrupt the performance of the fault detection system, which may even render it useless. To overcome this difficulty, the fault detection system has to be made robust, that is, insensitive or even invariant to such modeling errors. More specifically, a mere reduction of the sensitivity to modeling errors does not solve the problem because such a sensitivity reduction is generally associated with a reduction of the sensitivity to faults. A more meaningful formulation of the robust fault detection (RFD) problem must, therefore, require robustness to modeling errors without losing fault detection sensitivity. Such a system designed to provide satisfactory sensitivity to faults associated with the necessary robustness with respect to modeling errors is called an RFD scheme. In recent years the task of enhancing the robustness with respect to modeling uncertainties has been the subject of many published papers.[1,7–13] An overview of robustness issues and solutions can be found in the work by Frank.[7]

No matter which method is chosen to solve the robustness problem, there is often a point in common: to decrease the effects on the residues due to modeling uncertainties and simultaneously to increase that due to faults. In this context, Ding and Frank[14] presented a performance index expressed as a ratio of sensitivities of the residues due to the unknown inputs and the faults, respectively. The design goal is to then construct an observer for fault detection with the performance index being minimized. This or a similar idea is popular amongst a number of subsequent papers for model-based RFD systems.[7,8,15–19]

Following this idea, a method for designing an RFD observer in discrete time is studied in this paper. It is formulated as an optimization problem with observer poles as the constraints. Apart from

Haibo Wang was born in the People's Republic of China. He received his B.Sc. degree in Mathematics Education from Jinzhou Teacher College, Jinzhou, People's Republic of China, in 1992. He later obtained an M.Ph. degree in Applied Mathematics from the Northeastern University, Shenyang, People's Republic of China, in 1998. Since November 1998, he has been studying for his Ph.D. degree in Control Engineering at the University of Hong Kong. His research interests are in fault detection and isolation, reliable control, and H_∞ control.

James Lam received a first class B.Sc. degree in Mechanical Engineering from the University of Manchester in 1983. He was awarded the Ashbury Scholarship, the A. H. Gibson Prize, and the H. Wright Baker Prize for his academic performance. From the University of Cambridge, he obtained his M.Ph. and Ph.D. degrees in the area of Control Engineering in 1985 and 1988, respectively. His postdoctoral research was carried out in the Australian National University between 1990 and 1992. He is a Scholar (1984) and Fellow (1990) of the Croucher Foundation. He has held faculty positions at the now City University of Hong Kong and the University of Melbourne. He is now an Associate Professor in the Department of Mechanical Engineering at the University of Hong Kong, and holds a Concurrent Professorship at the Northeastern University, Guest Professorship at the Huazhong University of Science and Technology, and Consulting Professorship at South China University of China, People's Republic of China. His research interests include model reduction, delay systems, robust control, and generalized systems.

Received 12 January 2000; revision received 10 April 2001; accepted for publication 1 June 2001.

optimizing the sensitivity-related objective, the present formulation has also taken into account the observer gain size and the numerical conditioning of the observer. Moreover, we also considered the case where uncertainties may appear in the output. This has not been systematically treated in previous fault tolerant studies.[7,8,16,18] Numerical simulation is used to illustrate the effectiveness of the results. The paper is organized as follows. In Sec. II, the RFD problem is formulated and preliminary results are given. The problem is converted into a gradient-based optimization, and explicit gradient formulas are provided in Sec. III. The design process is summarized in an algorithmic form. Section IV provides numerical simulation of the results. Finally, a conclusion is provided in Sec. V.

II. Problem Formulation and Main Results

Throughout the paper, $\|\cdot\|$, $\|\cdot\|_F$, and tr $(\cdot)$ are used to denote the spectral norm, the Frobenius norm, and the trace of a matrix respectively. All matrices, if their dimensions are not explicitly stated, are assumed to have compatible dimensions.

Consider an uncertain discrete-time plant given by

$$\boldsymbol{x}(k+1) = [G + \Delta_G(k)]\boldsymbol{x}(k) + [B + \Delta_B(k)]\boldsymbol{u}(k) + K\boldsymbol{f}(k)$$
$$\boldsymbol{y}(k) = [C + \Delta_C(k)]\boldsymbol{x}(k) \tag{1}$$

where $\boldsymbol{x}(k) \in \mathbb{R}^n$ is the state vector, $\boldsymbol{u}(k) \in \mathbb{R}^m$ is the known control input vector, $\boldsymbol{y}(k) \in \mathbb{R}^r$ is the measurement vector, and $\boldsymbol{f}(k) \in \mathbb{R}^p$ is the actuator and component fault vector. Δ_G, Δ_B, and Δ_C are bounded time-varying uncertainties in G, B, and C, respectively. Notice that the fault vector $\boldsymbol{f}(k)$ may be input dependent or input independent. G, B, C, and K are known constant matrices with appropriate dimensions. The pair (C, G) is assumed to be observable, and the uncertain system is stable.

The observer for the generation of fault detection signal can be expressed as follows:

$$\hat{\boldsymbol{x}}(k+1) = (G - LC)\hat{\boldsymbol{x}}(k) + B\boldsymbol{u}(k) + L\boldsymbol{y}(k)$$
$$\hat{\boldsymbol{y}}(k) = C\hat{\boldsymbol{x}}(k), \qquad \boldsymbol{r}(k) = \boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k)$$

Denote the state estimation error and the closed-loop system matrix, respectively, as

$$\boldsymbol{e}(k) := \boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k), \qquad G_L := G - LC$$

where L is chosen such that G_L is stable. The error and residual dynamics of the fault diagnostic observer then become

$$\begin{aligned}\boldsymbol{e}(k+1) &= [G + \Delta_G(k)]\boldsymbol{x}(k) + [B + \Delta_B(k)]\boldsymbol{u}(k) + K\boldsymbol{f}(k) \\ &\quad - (G - LC)\hat{\boldsymbol{x}}(k) - B\boldsymbol{u}(k) - L[C + \Delta_C(k)]\boldsymbol{x}(k) \\ &= G\boldsymbol{e}(k) + \Delta_G(k)\boldsymbol{x}(k) + \Delta_B(k)\boldsymbol{u}(k) + K\boldsymbol{f}(k) - LC\boldsymbol{e}(k) \\ &\quad - L\Delta_C(k)\boldsymbol{x}(k) \\ &= G_L\boldsymbol{e}(k) + [\Delta_G(k) - L\Delta_C(k)]\boldsymbol{x}(k) \\ &\quad + \Delta_B(k)\boldsymbol{u}(k) + K\boldsymbol{f}(k)\end{aligned} \tag{2}$$

and the residual vector is given by

$$\boldsymbol{r}(k) = C\boldsymbol{e}(k) \tag{3}$$

Using Eq. (4), with $\tilde{\boldsymbol{x}}(k) := [\boldsymbol{x}^T(k) \;\; \boldsymbol{u}^T(k)]^T$, we can rewrite Eq. (2) as

$$\begin{aligned}&[\Delta_G(k) - L\Delta_C(k)]\boldsymbol{x}(k) + \Delta_B(k)\boldsymbol{u}(k) \\ &\quad = [\Delta_G(k) - L\Delta_C(k) \quad \Delta_B(k)]\tilde{\boldsymbol{x}}(k) \\ &\quad = \{[\Delta_G(k) \quad \Delta_B(k)] - L[\Delta_C(k) \quad 0]\}\tilde{\boldsymbol{x}}(k) \\ &\quad = [I_n \quad -L]\Delta(k)\tilde{\boldsymbol{x}}(k)\end{aligned}$$

where

$$\Delta(k) = \begin{bmatrix}\Delta_G(k) & \Delta_B(k) \\ \Delta_C(k) & 0_{n\times m}\end{bmatrix}$$

Assume the perturbation admitting the following structure:

$$\begin{aligned}\Delta(k) &= \Delta(E_1(k), E_2(k)) \\ &= \begin{bmatrix}F \\ 0\end{bmatrix}E_1(k)[H_G \quad H_B] + \begin{bmatrix}F_G \\ F_C\end{bmatrix}E_2(k)[H \quad 0]\end{aligned} \tag{4}$$

where F_G, F_C, F, H_G, H_B, and H are constant matrices defining the perturbation structure with

$$\|E_1(k)\|_F \le \varepsilon_1, \qquad \|E_2(k)\|_F \le \varepsilon_2 \tag{5}$$

The perturbation structure in the state equation is similar to that considered in Ref. 20. Although the robust stability of Eq. (1) with either E_1 or E_2 vanished has been established,[21] the general case is substantially more complicated.[22]

Remark 1: When the system is suffering from concurrent actuator and sensor faults as well as unmodeled noises, the system can be described as follows:

$$\boldsymbol{x}(k+1) = [G + \Delta_G(k)]\boldsymbol{x}(k) + [B + \Delta_B(k)]\boldsymbol{u}(k) + K_1\boldsymbol{f}_1(k)$$
$$\boldsymbol{y}(k) = [C + \Delta_C(k)]\boldsymbol{x}(k) + K_2\boldsymbol{f}_2(k) + \Delta_N(k)$$

where $\boldsymbol{f}_1(k)$ and $\boldsymbol{f}_2(k)$ are, respectively, the actuator and sensor fault vectors, K_1 and K_2 are known constant matrices, and $\Delta_N(k)$ is the unmodeled noise matrix. In this case, the error dynamics is governed by

$$\begin{aligned}\boldsymbol{e}(k+1) &= G_L\boldsymbol{e}(k) + [\Delta_G(k) - L\Delta_C(k)]\boldsymbol{x}(k) + \Delta_B(k)\boldsymbol{u}(k) \\ &\quad + \Delta_N(k) + [K_1 \quad -LK_2]\begin{bmatrix}\boldsymbol{f}_1(k) \\ \boldsymbol{f}_2(k)\end{bmatrix}\end{aligned}$$

and, hence, a similar development can be obtained. However, to maintain the lucidity of this work, such an extension is not treated here.

Consider the error dynamics governed by

$$\begin{aligned}\boldsymbol{e}(k+1) &= G_L\boldsymbol{e}(k) + [I_n \quad -L]\left\{\begin{bmatrix}F \\ 0\end{bmatrix}E_1(k)[H_G \quad H_B]\right. \\ &\quad \left. + \begin{bmatrix}F_G \\ F_C\end{bmatrix}E_2(k)[H \quad 0]\right\}\tilde{\boldsymbol{x}}(k) + K\boldsymbol{f}(k)\end{aligned} \tag{6}$$

Thus, the residues $\boldsymbol{r}(k)$ are influenced by two inputs, namely, $\tilde{\boldsymbol{x}}(k)$ and $\boldsymbol{f}(k)$. Because the system in Eq. (1) is assumed to be stable, we have $\|\tilde{\boldsymbol{x}}(k)\| \le \|\boldsymbol{x}(k)\| + \|\boldsymbol{u}(k)\|$ and, hence, $\tilde{\boldsymbol{x}}(k)$ is a bounded sequence if the input $\boldsymbol{u}(k)$ is bounded. In other words, when the input $\boldsymbol{u}(k)$ and fault $\boldsymbol{f}(k)$ are bounded, then the error dynamics described by Eq. (6) can be considered as one subjected to bounded disturbances. To measure the magnitude of their effects, we use the H_2 norm of the system (6) relating to $\tilde{\boldsymbol{x}}(k)$ and $\boldsymbol{f}(k)$ taken individually. That is, the H_2 norm of the systems with realizations

$$\mathcal{G}_E\colon (G_L, [I_n \quad -L]\Delta(k), C), \qquad \mathcal{G}_K\colon (G_L, K, C)$$

are given, respectively, by

$$\|\mathcal{G}_E\|_2 = \lim_{N\to\infty}\sqrt{\frac{1}{N+1}\sum_{h=0}^{N}\sum_{k=0}^{N}\mathrm{tr}[R(k,h)^T R(k,h)]}$$

$$\|\mathcal{G}_K\|_2 = \left\|Q^{\frac{1}{2}}K\right\|_F$$

where

$$R(k,h) = \begin{cases}CG_L^{k-h}[I_n \quad -L]\Delta(h), & k \ge h \\ 0, & k < h\end{cases}$$

and Q satisfies

$$G_L^T Q G_L - Q + C^T C = 0 \tag{7}$$

Notice that $\mathcal{G}_E$ is a time-varying system with $\|\mathcal{G}_E\|_2$ dependent on the parameter sequences $E_1(k)$ and $E_2(k)$. The following proposition provides an upper bound on $\|\mathcal{G}_E\|_2$ and the following lemma is required in the process (see Appendix A for its proof).

Lemma 1: Let A, B, and $\mathcal{E}$ be real matrices of compatible dimensions. Then

$$\max_{\|\mathcal{E}\|_F \le \epsilon} \|A\mathcal{E}B\|_F = \epsilon\|A\|\|B\|$$

Proposition 1: With Q defined in Eq. (7), and $E_1(k)$ and $E_2(k)$ satisfying Eq. (5),

$$\|\mathcal{G}_E\|_2 \le \sqrt{2}\Big[\varepsilon_1 \Big\|Q^{\frac{1}{2}}F\Big\|\|[H_G \quad H_B]\| + \varepsilon_2 \Big\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\Big\|\|H\|\Big] \tag{8}$$

In particular, if $\Delta_1(k) = 0$ [respectively, $\Delta_2(k) = 0$], then

$$\|\mathcal{G}_E\|_2 \le \varepsilon_1 \Big\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\Big\|\|H\|$$

$$\Big(\text{respectively } \|\mathcal{G}_E\|_2 \le \varepsilon_2 \Big\|Q^{\frac{1}{2}}F\Big\|\|[H_G \quad H_B]\|\Big)$$

Proof: See Appendix B. □

It is clear that the part of the "energy" contained in the residues is contributed by $\|\mathcal{G}_E\|_2$, which is dependent on the actually perturbation sequences $E_1(k)$ and $E_2(k)$. In practice, these sequences are not known a priori. For this reason, one should consider the worst-case sensitivity of the residues due to $E_1(k)$ and $E_2(k)$. The following function related to the worst-case sensitivity of residues due to these parameters is defined:

$$S_E := \varepsilon_1 \Big\|Q^{\frac{1}{2}}F\Big\|\|[H_G \quad H_B]\| + \varepsilon_2 \Big\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\Big\|\|H\|$$

Similarly, the sensitivity of residues due to faults is defined as

$$S_K := \Big\|Q^{\frac{1}{2}}K\Big\|_F$$

The ratio of sensitivities S_E/S_K, referred to as the noise-fault sensitivity, is a noise–signal measure in the robust fault detection context. It is also easily seen that

$$\frac{\|\mathcal{G}_E\|_2}{\|\mathcal{G}_K\|_2} \le \sqrt{2}\frac{S_E}{S_K}$$

Clearly, S_E should be kept small to desensitize the influence of uncertainties in the residual vector whereas S_K should be made large to enhance the sensitivity due to faults. Unfortunately, there is in general a tradeoff between the two sensitivities.

For the observer system matrix G_L, the choice of L is such that the eigenvalues are distinct and $\operatorname{spec}(G_L) \cap \operatorname{spec}(G) = \emptyset$ where $\operatorname{spec}(\cdot)$ denotes the spectrum of a matrix. The reason that the eigenvalues are chosen to be distinct is due to an eigenvalue sensitivity consideration (less susceptible to perturbation). This is always possible because (C, G) is observable. Then there exists an invertible V such that

$$VG_LV^{-1} = \Lambda \tag{9}$$

where Λ is a real pseudodiagonal matrix with $\operatorname{spec}(G_L) = \operatorname{spec}(\Lambda)$. Specifically,

$$\Lambda = \operatorname{diag}\left[\begin{pmatrix}\alpha_1 & \beta_1\\ -\beta_1 & \alpha_1\end{pmatrix}, \ldots, \begin{pmatrix}\alpha_{n'} & \beta_{n'}\\ -\beta_{n'} & \alpha_{n'}\end{pmatrix}, \gamma_1, \ldots, \gamma_{n-2n'}\right]$$

with the eigenvalues of Λ as the desired observer eigenvalues, $\alpha_i \pm \beta_i j$, $i = 1, \ldots, n'$, γ_k, $k = 1, \ldots, (n - 2n')$. Although V is not an eigenvector matrix, there exists a unitary U such that VU is an eigenvector matrix of G_L. Nevertheless, V defined via Eq. (9) is nonunique. By writing Eq. (9) as

$$VG - \Lambda V = MC, \qquad L = V^{-1}M \tag{10}$$

then for each M, a unique V is obtained because $\operatorname{spec}(\Lambda) \cap \operatorname{spec}(G) = \emptyset$. Moreover, the set

$$\mathcal{M} := \{M \in \mathbb{R}^{n\times r} \mid V \text{satisfies } VG - \Lambda V = MC \text{ is invertible}\}$$

is open and dense in $\mathbb{R}^{n\times r}$. To improve the numerical conditioning of the observer, the spectral condition number of the eigenvector matrix of the closed-loop system matrix given by $\|V\|\|V^{-1}\|$ should be small due to the Bauer–Fike theorem.[23] On the other hand, the size of the observer gain measured by $\|L\|$ should not be excessively large. Moreover, a small $\|L\|$ will also reduce the effects of the any perturbations in C. For a given fixed set of desired observer poles characterized by Λ, the residual vector mixed sensitivity minimization problem can be formulated as follows:

$$\min_{M\in\mathcal{M}} \big[\alpha(S_E/S_K) + \beta\|L\| + (1-\alpha-\beta)\|V\|\|V^{-1}\|\big] \tag{11}$$

where $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ are scalar weightings subject to $\alpha + \beta \le 1$. These parameters are used to control the relative emphasis of the individual objectives. Because of the tradeoffs between these objectives, the parameters α and β are required to be tuned so as to obtain an acceptable compromised solution.

Remark 2: Similar to almost all relevant works,[1,7,8,12,24–27] we considered the system in the open loop. Our task is to design a fault detection observer, and it does not alter the structure of system subject to faults. It should be understood that the open-loop system in Eq. (1) may also represent some system under closed-loop feedback that is subjected to uncertainties. Thus, it makes no difference for both open-loop or closed-loop systems using our approach as long as they can be equivalently described by Eq. (1). On the other hand, if $\boldsymbol{u}(k) = P\boldsymbol{x}(k) + \boldsymbol{v}(k)$ is given a priori with $\boldsymbol{v}(k)$ as the new reference input, then the closed-loop system obtained with Eq. (1) is given by

$$\boldsymbol{x}(k+1) = [G + BP + \Delta_G(k) + \Delta_B(k)P]\boldsymbol{x}(k) + [B + \Delta_B(k)]\boldsymbol{v}(k) + K\boldsymbol{f}(k) \tag{12}$$

and, hence, a similar formulation can be obtained by identifying $G + BP \to G$, $\Delta_G(k) + \Delta_B(k)P \to \Delta_G(k)$ and $H_G + H_BP \to H_G$. In this case, the fault vector $\boldsymbol{f}(k)$ must be input independent and treated as input disturbance to the system (12). If $\boldsymbol{f}(k)$ is state dependent, such as that due to actuator faults, then it would be difficult to ensure the closed-loop stability, and our designed observer does not have the appropriate structure to estimate the state of Eq. (12).

Remark 3: The choice of $\Delta_G(k)$, $\Delta_B(k)$, and $\Delta_C(k)$, usually of small magnitude, is primarily used to structurally represent modeling uncertainties in G, B, and C, respectively. In the closed-loop case, although $\Delta_B(k)$ can also be used to reflect actuator faults, such $\Delta_B(k)$ may be large enough to affect the performance of the observer constructed based on a noise–signal measure. For closed-loop systems with significant degradation of the control channels such as actuator failures, fault detection will not be considered as the major issue because stability may not be preserved anyway. These are some reasons why fault detection is normally formulated in the open loop, and along this direction the present paper is developed.

III. Gradient-Based Optimization

Consequently, an equivalent minimization problem to that in Eq. (11) can be formulated as

$$\text{Problem RFD:} \quad \min_{M\in\mathcal{M}} \mathcal{J} \tag{13}$$

with

$$\mathcal{J} := \alpha J_1 + \beta J_2 + (1-\alpha-\beta)J_3 \tag{14}$$

where

$$J_1 := \frac{\varepsilon_1 \Big\|Q^{\frac{1}{2}}F\Big\|\|[H_G \quad H_B]\| + \varepsilon_2 \Big\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\Big\|\|H\|}{\Big\|Q^{\frac{1}{2}}K\Big\|_F}$$

$$J_2 := \|L\|, \qquad J_3 := \|V\|\|V^{-1}\| \tag{15}$$

Notice that Eq. (13) corresponds to an unconstrained minimization problem with Eq. (14) differentiable almost everywhere (apart from possibly at those maximum singular values in $Q^{1/2}F$, $Q^{1/2}[F_G \; -LF_C]$, L, V, and V^{-1}, with multiplicity greater than unity).

Thus, a gradient-based optimization procedure can be applied. The gradient of $\mathcal{J}$ with respect to M is then summarized in the following proposition (see Appendix C for its proof).

Proposition 2: Suppose

$$Q^{\frac{1}{2}}F\boldsymbol{v}_{11} = \left\|Q^{\frac{1}{2}}F\right\|\boldsymbol{u}_{11}$$

$$Q^{\frac{1}{2}}[F_G \quad -LF_C]\boldsymbol{v}_{12} = \left\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\|\boldsymbol{u}_{12}$$

$$L\boldsymbol{v}_2 = \|L\|\boldsymbol{u}_2, \qquad V\boldsymbol{x}_1 = \|V\|\boldsymbol{y}_1, \qquad V^{-1}\boldsymbol{x}_n = \|V^{-1}\|\boldsymbol{y}_n$$

where $(\boldsymbol{v}_{11}, \boldsymbol{u}_{11})$, $(\boldsymbol{v}_{12}, \boldsymbol{u}_{12})$, $(\boldsymbol{v}_2, \boldsymbol{u}_2)$, $(\boldsymbol{x}_1, \boldsymbol{y}_1)$, and $(\boldsymbol{x}_n, \boldsymbol{y}_n)$ are the corresponding singular vector pairs (unit norm). If the maximum singular values of $Q^{1/2}F$, $Q^{1/2}[F_G \; -LF_C]$, L, V, and V^{-1} are distinct, then

$$\begin{aligned}\frac{\partial J}{\partial M} &= \alpha\frac{\partial J_1}{\partial M} + \beta\frac{\partial J_2}{\partial M} + (1-\alpha-\beta)\frac{\partial J_3}{\partial M} \\ &= \alpha\left[Y^T - V^{-T}QG_L(X+X^T)\right]C^T \\ &\quad + \alpha\frac{\varepsilon_2\|H\|V^{-T}Q^{\frac{1}{2}}\boldsymbol{u}_1\boldsymbol{v}_1^T[0 \quad -F_C]^T}{\left\|Q^{\frac{1}{2}}K\right\|_F} \\ &\quad + \beta\left(-Z^TC^T + V^{-T}\boldsymbol{u}_2\boldsymbol{v}_2^T\right) + (1-\alpha-\beta)U^TC^T \end{aligned} \tag{16}$$

where

$$G_L^TQG_L - Q + C^TC = 0 \tag{17}$$

$$W_1Q^{\frac{1}{2}} + Q^{\frac{1}{2}}W_1 - F\boldsymbol{v}_{11}\boldsymbol{u}_{11}^T = 0 \tag{18}$$

$$W_2Q^{\frac{1}{2}} + Q^{\frac{1}{2}}W_2 - [F_G \quad -LF_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T = 0 \tag{19}$$

$$\begin{aligned} G_LXG_L^T - X &+ \frac{\varepsilon_1\|[H_G \quad H_B]\|W_1 + \varepsilon_2\|H\|W_2}{\left\|Q^{\frac{1}{2}}K\right\|_F} \\ &- \frac{\varepsilon_1\|[H_G \quad H_B]\|\left\|Q^{\frac{1}{2}}F\right\| + \varepsilon_2\|H\|\left\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\|}{2\left\|Q^{\frac{1}{2}}K\right\|_F^3} \\ &\times KK^T = 0 \end{aligned} \tag{20}$$

$$\begin{aligned} GY - Y\Lambda - LC(X+X^T)G_L^TQV^{-1} \\ + \frac{\varepsilon_2\|H\|L[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^TQ^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F} = 0 \end{aligned} \tag{21}$$

$$GZ - Z\Lambda - \|L\|\boldsymbol{u}_2\boldsymbol{u}_2^TV^{-1} = 0 \tag{22}$$

$$GU - U\Lambda - \left[\|V^{-1}\|\boldsymbol{x}_1\boldsymbol{y}_1^T - \|V\|\|V^{-1}\|^2\boldsymbol{x}_n\boldsymbol{y}_n^T\right] = 0 \tag{23}$$

Remark 4: For a particular $M \in \mathcal{M}$, the computation of the gradient starts with solving V and, hence, L from the Sylvester equation in Eq. (10). Then $G_L = G - LC$ can be formed and the matrices required for $\partial J/\partial M$ can be obtained from the matrix equations in Proposition 2 when successively evaluated. There are eight Sylvester equations to be solved and singular value decompositions (SVD) are required to compute the singular vectors.

Now, we summarize the process of obtaining the fault detection observer gain in the following schematic RFD algorithm. G, B, C, K, F, F_G, F_C, H, H_G, H_B, and Λ are given.

1) Choose nonnegative numbers α and β subject to $\alpha + \beta \leq 1$ and nonnegative numbers ε_1 and ε_2.

2) Select an initial guess $M_0 \in \mathcal{M}$ and solve minimization problem (13) based on the objective function (14), and function (15) and its gradient function (16). (Off-the-shelf numerical routines such as those from MATLAB® may be used, or gradient-based numerical algorithms in Ref. 28 may be implemented.)

3) Let M_{opt} be the optimal solution obtained in step 2. Compute J_1, J_2, and J_3.

4) IF J_1, J_2, and J_3 meet the requirements, GOTO step 5, ELSE GOTO step 1 and modify the values of α and β (relax the perturbation constraint values ε_1 and ε_2 if possible).

5) Solve Eq. (10), the required observer gain is given by $L_{\text{opt}} = V^{-1}M_{\text{opt}}$.

IV. Numerical Simulation

Example 1: Consider a linearized discrete-time model of a manipulator with two rotational joints studied in Ref. 29,

$$\boldsymbol{x}(k+1) = G\boldsymbol{x}(k) + B\boldsymbol{u}(k) + K\boldsymbol{f}(k), \qquad \boldsymbol{y}(k) = C\boldsymbol{x}(k)$$

where

$$G = \begin{bmatrix} 0.9627 & 0.0019 & -0.1963 & -0.0002 & 0.1778 & 0.0002 \\ 0.0019 & 0.9588 & -0.0002 & -0.1959 & 0.0002 & 0.1774 \\ 0.0196 & 0 & 0.9980 & 0 & 0.0018 & 0 \\ 0 & 0.0196 & 0 & 0.9980 & 0 & 0.0018 \\ 0 & 0 & 0 & 0 & 0.8187 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8187 \end{bmatrix}$$

$$B = K = \begin{bmatrix} 0.0185 & 0 \\ 0 & 0.0185 \\ 0.0001 & 0 \\ 0 & 0.0001 \\ 1.1813 & 0 \\ 0 & 0.1813 \end{bmatrix}, \qquad C = [0 \quad I_4]$$

Notice that the dynamics of the actuators, the field-controlled dc motors, are included in the model. The state variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ represent the joint angle velocities, whereas $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$ represent the joint angles. Finally, $\boldsymbol{x}_5$ and $\boldsymbol{x}_6$ are the actuators command outputs. The system was simulated with an actuator fault starting at $k = 100$ given by $[0.1, -0.8]^T$. In this example, the system uncertainty in Eq. (4) has

$$F = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \\ 0.2 & 0 \\ 0 & 0.2 \\ 0 & 0.2 \\ 0 & 0 \end{bmatrix}, \qquad F_G = \begin{bmatrix} 0.25 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.25 \end{bmatrix}$$

$$F_C = \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \\ 0 & 0.25 \\ 0 & 0 \end{bmatrix}$$

$$H = \begin{bmatrix} 0.05 & 0 & 0 & 0.05 & 0.05 & 0 \\ 0 & 0.05 & 0 & 0 & 0 & 0.05 \end{bmatrix}$$

$$H_G = \begin{bmatrix} 0.25 & 0 & 0 & 0 & 0 & 0.25 \\ 0 & 0.25 & 0 & 0 & 0.25 & 0 \end{bmatrix}, \qquad H_B = \begin{bmatrix} 0.25 & 0 \\ 0 & 0 \end{bmatrix}$$

and $E_1(k)$ and $E_2(k)$ are random processes with $\|E_1(k)\|_F = \varepsilon_1$ and $\|E_2(k)\|_F = \varepsilon_2$ (each element is a random variable drawn from a normal distribution with zero mean and unit variance and then normalized to give the required norm). Notice that this represents the worst sequence, in terms of perturbation norm, in the simulation process. In the present example, ε_1 and ε_2 are taken as unity. Two cases of $\boldsymbol{u}(k)$ will be used, $\boldsymbol{u}_1(k) = [1, 1]^T$ and $\boldsymbol{u}_2(k) = [\sin(0.1k), \cos(0.1k)]^T$.

The choice of the target poles of observer (different from those of the original system) should reflect the response speed requirement, and a faster speed generally lead to a larger feedback gain. Thus, a compromise between them should be reached. In our example, the poles of the diagnostic observer are designed at 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6, which give a nonoscillatory response behavior. That is,

$$\Lambda = \text{diag}(0.1, 0.2, 0.3, 0.4, 0.5, 0.6)$$

When the approach developed in this paper is used, the optimization is initiated with a random initial value of M. Numerical experience indicates that the initial choice of M has little influence on the optimal results. However, it is suggested that a number of initial guesses should be tested. The numerical simulation was carried out using MATLAB 4.2 (Control Toolbox 3.0b, Optimization Toolbox 1.0d), and an optimal observer gain L_{opt1} is obtained. For comparison, an observer gain L_{place}, which gives the same spectrum is obtained from the command place.m (the command has also taken the sensitivities of the eigenvalues into account). We have

$$L_{\mathrm{opt1}} = \begin{bmatrix} 18.3138 & -2.9365 & -4.0858 & -3.7116 \\ -3.0069 & 14.7346 & -8.0278 & -6.6398 \\ 1.3656 & -0.0583 & -0.1146 & -2.0516 \\ 0.0204 & 1.2612 & -0.2245 & -3.7898 \\ 0.2065 & 0.1421 & 0.5564 & -5.6251 \\ 0.0000 & -0.0091 & 0.0136 & 0.2718 \end{bmatrix}$$

$$L_{\mathrm{place}} = \begin{bmatrix} 25.3413 & 3.6371 & 0.1778 & 0.0002 \\ 3.8877 & 25.0679 & 0.0002 & 0.1774 \\ 1.4613 & 0.0977 & 0.0018 & 0 \\ 0.1061 & 1.4562 & 0 & 0.0018 \\ 0 & 0 & 0.3187 & 0 \\ 0 & 0 & 0 & 0.2187 \end{bmatrix}$$

In Table 1, different choices of α and β are compared with L_{place}. Ideally, all J_i, $i = 1, 2, 3$, should be uniformly small. However, because of the tradeoffs between these objectives, a compromise is necessary. Notice that cases 1–3 correspond to special situations where only one particular objective J_i, $i = 1, 2, 3$, is minimized, respectively. Thus, the values in cases 1–3 may be considered as the best values obtained when J_i is considered individually. The choice of α and β in case OPTIMAL1 is considered to be better as compared with case PLACE. The tradeoffs between these objectives can be observed.

In example 1, Figs. 1–4 show the residual responses due to different choices of α and β, respectively, and Fig. 5 shows the residual responses due to L_{place}. In case OPTIMAL1, despite the influence of E_1 and E_2, a threshold at ± 1 can easily be imposed on the residual signals to indicate the occurrence of fault at the sampling instant $k = 100$, which disappeared at the sampling instant $k = 200$. In other words, the robust fault detection sensitivity as indicated by J_1 in case PLACE is comparatively smaller than in case OPTIMAL1. Moreover, Figs. 1–3 also demonstrate the effectiveness of our method by showing the residual responses when objective functions $J = J_1$, J_2, and J_3 are used, respectively. It can be seen that no reasonable threshold can be deduced to distinguish the influence between faults and uncertainties in these three cases. The comparisons, thus, illustrate that J_1, J_2, and J_3 all played important roles when designing an RFD observer.

Table 1 Summary of simulation results of example 1

Case	Parameters	J_1	J_2	J_3
1	$\alpha = 1, \beta = 0$	0.1197	9.4×10^4	7.3×10^5
2	$\alpha = 0, \beta = 1$	0.8853	3.3	8.6×10^3
3	$\alpha = 0, \beta = 0$	0.5797	19.9	100.4
OPTIMAL1	$\alpha = 0.98, \beta = 0.01$	0.1960	20.0	101.2
PLACE	L_{place}	0.5592	29.0	293.3

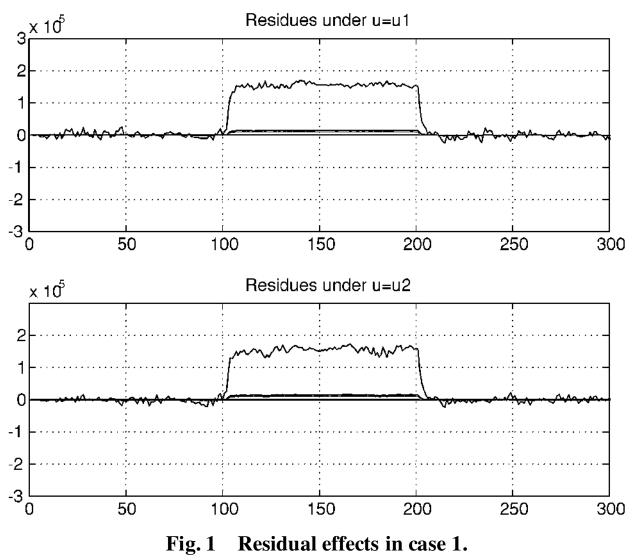

Fig. 1 Residual effects in case 1.

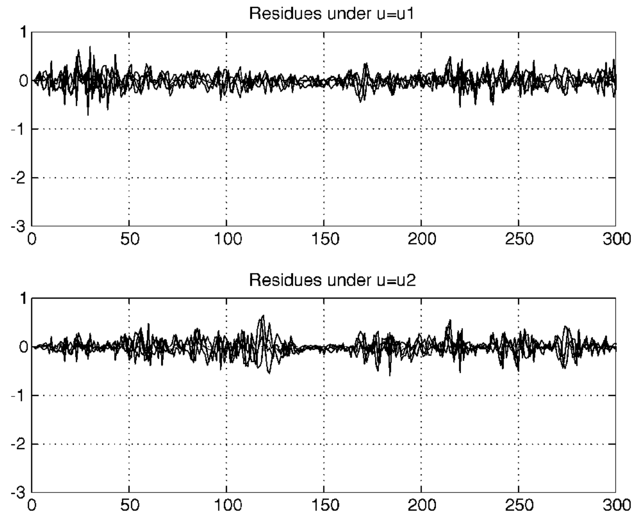

Fig. 2 Residual effects in case 2.

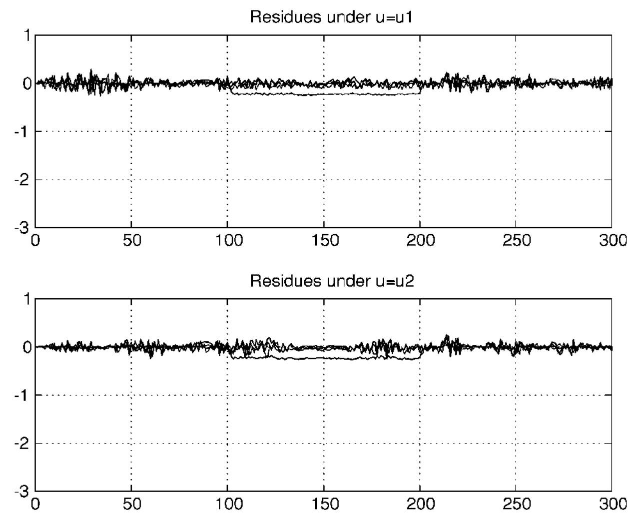

Fig. 3 Residual effects in case 3.

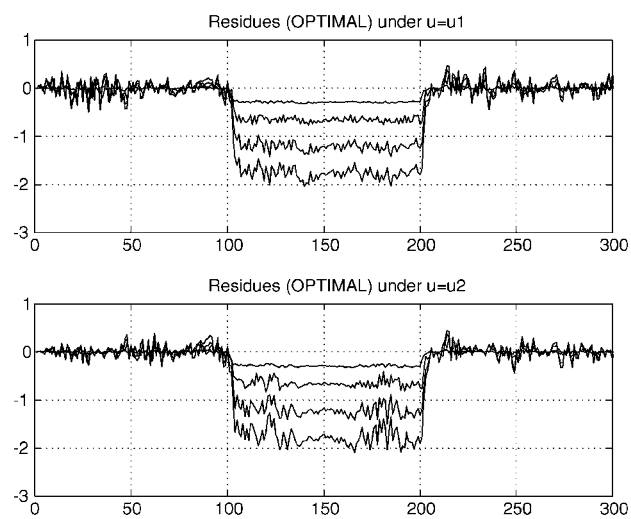

Fig. 4 Residual effects in case OPTIMAL1.

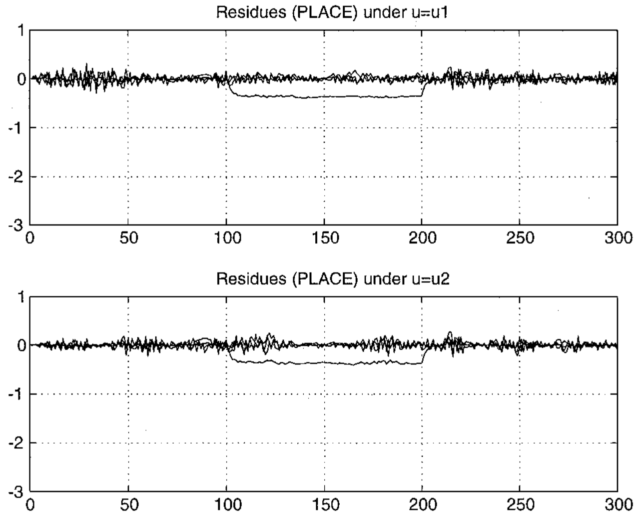

Fig. 5 Residual effects in case PLACE.

To assess the achieved performance in the optimization process of case OPTIMAL1, the improvement in J_1, J_2, and J_3 are compared with the values at the random initial guess. The noise-fault sensitivity (measured by J_1) is decreased by 45% (from 0.3550 to 0.1960), the magnitude of the observer gain L (measured by J_2) is decreased by 75% (from 80.3 to 20.0) and the condition number of the observer system matrix (measured by J_3) is decreased by 92% (from 1252.4 to 102.5).

Example 2: Consider the linearized dynamics of a vertical takeoff and landing aircraft in the vertical plane as proposed by Saif and Guan[30] and Narendra and Tripathi.[31] The continuous-time state-space description is

$$\dot{x}(t) = \bar{G}x(t) + \bar{B}u(t), \qquad y(t) = \bar{C}x(t)$$

where the states are the horizontal velocity (knot), vertical velocity (knot), and pitch rate (degree), respectively. The actuator inputs are the collective pitch control and the longitudinal pitch control, respectively. The system is sampled at 0.5 s. to result in a discrete-time state-space realization (G, B, C) with fault structure given by K, where

$$G = \begin{bmatrix} 0.9813 & 0.0083 & -0.0454 & -0.2459 \\ 0.0117 & 0.5813 & -0.3898 & -1.6662 \\ 0.0457 & 0.1274 & 0.8230 & 0.4803 \\ 0.0117 & 0.0358 & 0.4433 & 1.1361 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.2664 & 0.0365 \\ 1.7629 & -3.2664 \\ -2.3152 & 1.7209 \\ -0.6083 & 0.4660 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

$$K = 0.2B$$

The open-loop eigenvalues of the discrete-time nominal plant are located at $1.1383 \pm 0.1474i$, 0.3548, and 0.8902. A state feedback gain

$$P = \begin{bmatrix} -1.2707 & -0.0005 & 0.6865 & 1.4542 \\ -1.3140 & -0.0452 & 0.3938 & 0.9317 \end{bmatrix}$$

is employed to stabilize the aircraft such that the closed-loop eigenvalues are located at 0.25, 0.42, 0.65, and 0.60.

The faults, considered to be input independent, start at $k = 100$ and end at $k = 200$ given by $[1.5, -1.5]^T$. The system uncertainty structure is as in Eq. (4), where

$$F = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \\ 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}, \qquad F_G = \begin{bmatrix} 0.25 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.25 \end{bmatrix}$$

$$F_C = \begin{bmatrix} 0 & 0.25 \\ 0.25 & 0 \\ 0 & 0.25 \\ 0.25 & 0 \end{bmatrix}, \qquad H = \begin{bmatrix} 0.05 & 0.05 & 0 & 0 \\ 0.05 & 0 & 0 & 0.05 \end{bmatrix}$$

$$H_G = \begin{bmatrix} 0.2 & 0 & 0 & 0.2 \\ 0 & 0.2 & 0.2 & 0 \end{bmatrix}, \qquad H_B = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}$$

All other settings of $E_1(k)$, $E_2(k)$, ε_1, ε_2, and $\boldsymbol{u}_1(k)$ are the same with example 1 and $\boldsymbol{u}_2(k) = [0.5\sin(0.5k), 0.5\cos(0.5k)]^T$. The poles of the diagnostic observer are designed at 0.1, 0.2, 0.3, and 0.4. The design parameters α and β are taken as 0.990, and 0.001, respectively.

As discussed in Remark 2, if we identify $G + BP \to G$ and $H_G + H_B P \to H_G$, the closed-loop system

$$\boldsymbol{x}(k+1) = [G + \Delta_G(k) + BP + \Delta_B(k)P]\boldsymbol{x}(k) + [B + \Delta_B(k)]\boldsymbol{v}(k) + K\boldsymbol{f}(k)$$

$$\boldsymbol{y}(k) = (C + \Delta_C)\boldsymbol{x}(k)$$

can be handled by the proposed method. The numerical simulation was carried out using MATLAB 4.2 (Control Toolbox 3.0b, Optimization Toolbox 1.0d), and an optimal observer gain L_{opt2} is given by

$$L_{\text{opt2}} = \begin{bmatrix} 0.2598 & -0.2367 & -0.0216 & 0.2114 \\ 1.7974 & 1.4972 & 2.2948 & -1.6902 \\ 0.6519 & 0.8558 & 1.2574 & -1.0945 \\ -0.0864 & -0.9554 & -0.0616 & 0.6903 \end{bmatrix}$$

Notice that C is an invertible matrix, and an observer gain L_{exact} can be exactly solved with

$$L_{\text{exact}} = (G + BP - \Lambda)C^{-1} = \begin{bmatrix} 0.4949 & -0.1690 & -0.0236 & 0.1755 \\ 2.0638 & 2.6741 & 1.6803 & -2.1460 \\ 0.7262 & 1.3336 & 0.8941 & -1.2829 \\ 0.1722 & -0.2708 & -0.0766 & 0.2858 \end{bmatrix}$$

Because Λ is diagonal, the best (and lowest possible) condition number equals unity is obtained, and such L_{exact} will be used for comparison.

In Table 2, different choices of α and β are compared with L_{exact}. Figures 6–8 show the residual responses due to different choices of α and β, respectively, and Fig. 9 shows the residual responses due to L_{exact}. In case OPTIMAL2, a threshold at ± 5 can be imposed

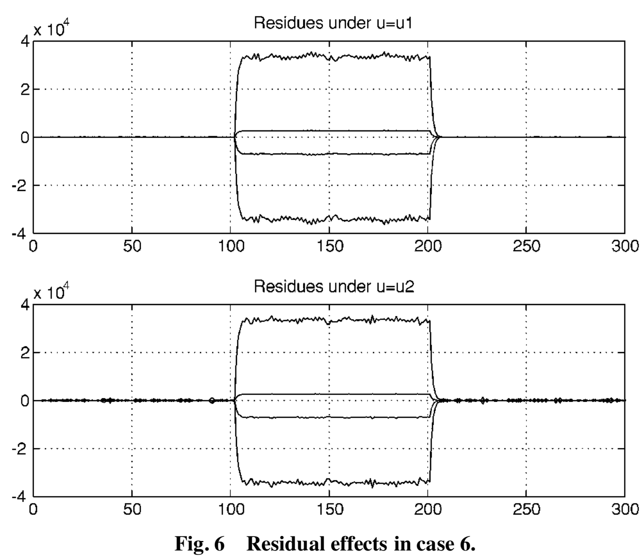

Fig. 6 Residual effects in case 6.

Table 2 Summary of simulation results of example 2

Case	Parameters	J_1	J_2	J_3
6	$\alpha = 1,\ \beta = 0$	0.0253	2.1×10^4	1.9×10^6
7	$\alpha = 0,\ \beta = 1$	0.2554	0.3	7.7×10^2
OPTIMAL2	$\alpha = 0.990,\ \beta = 0.001$	0.1248	4.3	10.8
EXACT	L_{exact}	0.3574	4.9	1

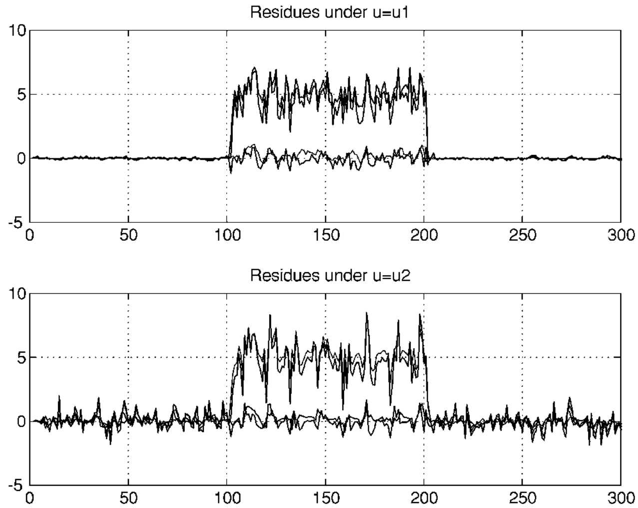

Fig. 7 Residual effects in case 7.

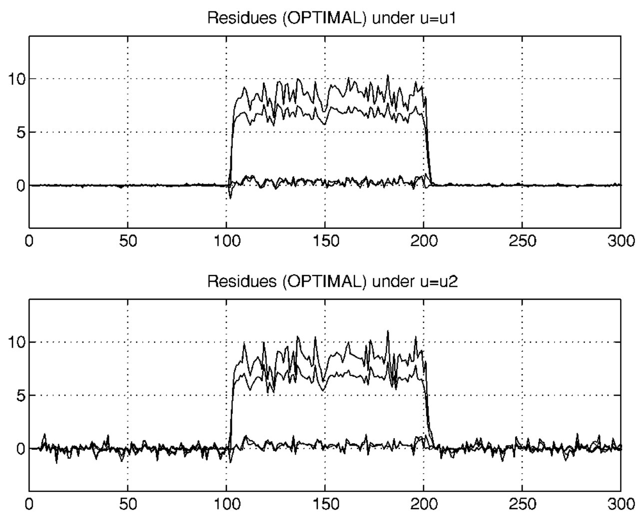

Fig. 8 Residual effects in case OPTIMAL2.

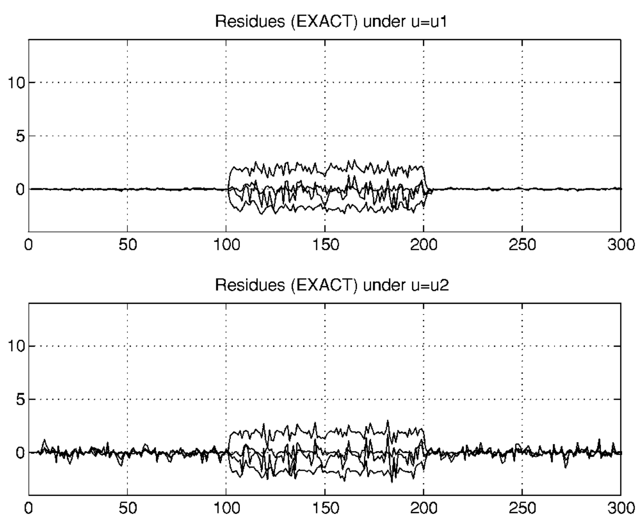

Fig. 9 Residual effects in case EXACT.

on the residual signals to indicate the occurrence of fault. In other words, even though L_{exact} has the lowest condition number, the RFD sensitivity as indicated by J_1 in case EXACT is smaller than in case OPTIMAL2. Moreover, Figs. 6 and 7 demonstrate the effectiveness of our method by showing the residual responses when objective functions $J = J_1$ and J_2 are used, respectively. Similar to example 1, the improvement in J_1, J_2, and J_3 with respect to the random initial guess are respectively 81, 75, and 99.0%.

V. Conclusions

In this paper, a method is proposed for designing an RFD observer when the system under consideration is subjected to structured perturbations. The approach aims at enhancing the fault detection robustness to uncertainties without sacrificing the fault detection sensitivity, which was seldom addressed. Furthermore, other objectives related to the observer gain and the eigenstructure conditioning of the observer system are also considered. The gradient-based optimization approach is facilitated by the explicit gradient expressions derived. Numerical simulation has also demonstrated the tradeoffs between different objectives as well as the effectiveness of the present methodology.

Appendix A: Proof of Lemma 1

Let the SVD of A and B be, respectively, $A = U_A \Sigma_A V_A^T$ and $B = U_B \Sigma_B V_B^T$ where the singular values are arranged in descending order. Thus,

$$A\mathcal{E}B = U_A \Sigma_A V_A^T \mathcal{E} U_B \Sigma_B V_B^T \Rightarrow \|A\mathcal{E}B\|_F = \left\| \Sigma_A V_A^T \mathcal{E} U_B \Sigma_B \right\|$$

and with $\mathcal{E}' := V_A^T \mathcal{E} U_B$, we have $\|\mathcal{E}'\|_F = \|V_A^T \mathcal{E} U_B\|_F = \|\mathcal{E}\|_F \le 1$. Hence,

$$\max_{\|\mathcal{E}\|_F \le \epsilon} \|A\mathcal{E}B\|_F = \max_{\|\mathcal{E}'\|_F \le \epsilon} \|\Sigma_A \mathcal{E}' \Sigma_B\|_F$$

Since

$$\|\Sigma_A \mathcal{E}' \Sigma_B\|_F \le \|\Sigma_A\| \|\mathcal{E}' \Sigma_B\|_F \le \|\Sigma_A\| \|\Sigma_B\| \|\mathcal{E}'\|_F \le \epsilon \|\Sigma_A\| \|\Sigma_B\|$$

and when $\mathcal{E}' = \epsilon \boldsymbol{u}_1 \boldsymbol{v}_1^T$, where $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are, respectively, the first standard basis of $\mathbb{R}^q$ and $\mathbb{R}^r$, respectively, the upper bound is achieved. That is,

$$\left\| \Sigma_A \left(\epsilon \boldsymbol{u}_1 \boldsymbol{v}_1^T \right) \Sigma_B \right\|_F = \epsilon \|\Sigma_A\| \|\Sigma_B\| = \epsilon \|A\| \|B\|$$

Therefore, the result follows. □

Appendix B: Proof of Proposition 1

Define

$$\|\mathcal{G}_E\|_{2,[0,N]} = \sqrt{\frac{1}{N+1} \sum_{h=0}^{N} \sum_{k=0}^{N} \operatorname{tr}[R(k,h)^T R(k,h)]}$$

We have

$$\sum_{k=0}^{N} \operatorname{tr}[R(k,h)^T R(k,h)]$$

$$= \sum_{k=0}^{N} \operatorname{tr}\left(\Delta(h)^T \begin{bmatrix} I_n \\ L^T \end{bmatrix} \left(G_L^{k-h}\right)^T C^T C G_L^{k-h} [I_n \quad -L] \Delta(h) \right)$$

$$= \operatorname{tr}\left(\Delta(h)^T \begin{bmatrix} I_n \\ L^T \end{bmatrix} Q(h) [I_n \quad -L] \Delta(h) \right)$$

where

$$Q(h) = \sum_{k=h}^{N} \left(G_L^T\right)^{k-h} C^T C G_L^{k-h}$$

satisfies

$$G_L^T Q(h) G_L - Q(h-1) + C^T C = 0, \qquad \Delta(h) = \Delta_1(h) + \Delta_2(h)$$

$$\Delta_1(h) = [I_n \quad -L] \begin{bmatrix} F \\ 0 \end{bmatrix} E_1(h)[H_G \quad H_B]$$

$$\Delta_2(h) = [I_n \quad -L] \begin{bmatrix} F_G \\ F_C \end{bmatrix} E_2(h)[H \quad 0]$$

Therefore,

$$\mathrm{tr}[\Delta(h)^T Q(h)\Delta(h)] \le 2\left\| Q^{\frac{1}{2}}(h)\Delta_1(h)\right\|_F^2 + 2\left\| Q^{\frac{1}{2}}(h)\Delta_2(h)\right\|_F^2$$

From the form of $Q(h)$, it is easy to see that $Q(h) \ge 0$ and for $t_1 \le t_2$, we have $Q(t_1) \ge Q(t_2)$. Moreover, G_L is stable; thus, $Q(h)$ is a uniformly bounded sequence and, hence, $\lim_{h\to-\infty} Q(h) = Q$, where Q satisfies Eq. (7). In fact, it is easy to see that

$$\lim_{h\to-\infty} Q(h) = \lim_{h\to\infty} \sum_{k=-h}^{N} \left(G_L^T\right)^{k+h} C^T C G_L^{k+h}$$

$$= \lim_{h\to\infty} \sum_{m=0}^{N+h} \left(G_L^T\right)^m C^T C G_L^m = \sum_{m=0}^{\infty} \left(G_L^T\right)^m C^T C G_L^m$$

which is the observability gramian. Now,

$$\|\mathcal{G}_E\|_{2,[0,N]}^2 \le \frac{2}{N+1} \sum_{h=0}^{N} \left[\left\| Q^{\frac{1}{2}}(h)\Delta_1(h)\right\|_F^2 + \left\| Q^{\frac{1}{2}}(h)\Delta_2(h)\right\|_F^2 \right]$$

$$\le \frac{2}{N+1} \sum_{h=0}^{N} \left[\left\| Q^{\frac{1}{2}}\Delta_1(h)\right\|_F^2 + \left\| Q^{\frac{1}{2}}\Delta_2(h)\right\|_F^2 \right]$$

$$\le 2\left[\max_{h=0,1,\ldots,N} \left\| Q^{\frac{1}{2}}\Delta_1(h)\right\|_F^2 + \max_{h=0,1,\ldots,N} \left\| Q^{\frac{1}{2}}\Delta_2(h)\right\|_F^2 \right]$$

Clearly, because $E_1(h)$ and $E_2(h)$ are uniformly bounded, the right-hand side term is well defined. Moreover, we have the 2-norm of $\mathcal{G}$ over an infinite interval given by

$$\|\mathcal{G}_E\|_2 \le \sqrt{2}\left[\max_{h=0,1,\ldots} \left\| Q^{\frac{1}{2}}\Delta_1(h)\right\|_F^2 + \max_{h=0,1,\ldots} \left\| Q^{\frac{1}{2}}\Delta_2(h)\right\|_F^2 \right]^{\frac{1}{2}}$$

$$\le \sqrt{2}\left[\max_{h=0,1,\ldots} \left\| Q^{\frac{1}{2}}\Delta_1(h)\right\|_F + \max_{h=0,1,\ldots} \left\| Q^{\frac{1}{2}}\Delta_2(h)\right\|_F \right]$$

Therefore, we obtain Eq. (8) according to Lemma 1. Suppose, for any $k = 0, 1, \ldots,$ the worst Δ_1 and Δ_2 are such that

$$\max_{\|E_1\|_F \le \varepsilon_1} \left\| Q^{\frac{1}{2}}\Delta_1\right\|_F = \varepsilon_1 \left\| Q^{\frac{1}{2}}[I_n \quad -L] \begin{bmatrix} F \\ 0 \end{bmatrix}\right\| \|[H_G \quad H_B]\|$$

$$= \varepsilon_1 \left\| Q^{\frac{1}{2}} F\right\| \|[H_G \quad H_B]\|$$

$$\max_{\|E_2\|_F \le \varepsilon_2} \left\| Q^{\frac{1}{2}}\Delta_2\right\|_F = \varepsilon_2 \left\| Q^{\frac{1}{2}}[I_n \quad -L] \begin{bmatrix} F_G \\ F_C \end{bmatrix}\right\| \|[H \quad 0]\|$$

$$= \varepsilon_2 \left\| Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\| \|H\|$$

In particular, if $\Delta_1(k) = 0$ [respectively $\Delta_2(k) = 0$], then

$$\|\mathcal{G}_E\|_2 \le \max_{k=0,1,\ldots} \left\| Q^{\frac{1}{2}}\Delta_2(k)\right\|_F$$

$$\times \left[\text{respectively}, \|\mathcal{G}_E\|_2 \le \max_{k=0,1,\ldots} \left\| Q^{\frac{1}{2}}\Delta_1(k)\right\|_F\right]$$

$$\|\mathcal{G}_E\|_2 \le \varepsilon_2 \left\| Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\| \|H\|$$

$$\times \left[\text{respectively}, \|\mathcal{G}_E\|_2 \le \varepsilon_1 \left\| Q^{\frac{1}{2}}F\right\| \|[H_G \quad H_B]\|\right] \qquad \square$$

Appendix C: Proof of Proposition 2

To prove Proposition 2, we require the following lemma.

Lemma 2: For real matrices M, N, Q, R, X, and Y satisfying either of the following Sylvester equations

$$MX + XN + Q = 0, \qquad YM + NY + R = 0$$

$$MXN - X + Q = 0, \qquad NYM - Y + R = 0$$

then $\mathrm{tr}(XR) = \mathrm{tr}(YQ)$.

For brevity, we only prove the second Sylvester equation case of Lemma 2.

Proof of Lemma 2 (second case): Postmultiplying the two Sylvester equations by R and Q gives

$$MXNR - XR + QR = 0, \qquad NYMQ - YQ + RQ = 0$$

and, hence, $\mathrm{tr}(MXNR - XR + QR) = \mathrm{tr}(NYMQ - YQ + RQ)$. Since $\mathrm{tr}(QR) = \mathrm{tr}(RQ)$, we have

$$\mathrm{tr}(MXNR) = \mathrm{tr}[(MXN)(Y - NYM)] = \mathrm{tr}[(MXNY - MXNNYM)]$$

$$= \mathrm{tr}(NYMX - NYMMXN) = \mathrm{tr}(NYMQ) \qquad \square$$

The result follows.

We divide the whole proof into three parts for J_1, J_2, and J_3, respectively.

First, consider $J_1 = J_{11} + J_{12}$, where

$$J_{11} := \frac{\varepsilon_1 \left\| Q^{\frac{1}{2}} F\right\| \|[H_G \quad H_B]\|}{\left\| Q^{\frac{1}{2}} K\right\|_F}$$

$$J_{12} := \frac{\varepsilon_2 \left\| Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\| \|H\|}{\left\| Q^{\frac{1}{2}} K\right\|_F}$$

We first consider J_{11}. With $M = [m_{ij}]$,

$$\frac{\partial J_{11}}{\partial m_{ij}} = \frac{\varepsilon_1 \|[H_G \quad H_B]\|}{\left\| Q^{\frac{1}{2}} K\right\|_F} \frac{\partial \left\| Q^{\frac{1}{2}} F\right\|}{\partial m_{ij}}$$

$$- \frac{\varepsilon_1 \|[H_G \quad H_B]\| \left\| Q^{\frac{1}{2}} F\right\|}{\left\| Q^{\frac{1}{2}} K\right\|_F^2} \frac{\partial \left\| Q^{\frac{1}{2}} K\right\|_F}{\partial m_{ij}}$$

Because the maximum singular value of $Q^{1/2}F$ is distinct and

$$Q^{\frac{1}{2}} F \boldsymbol{v}_{11} = \left\| Q^{\frac{1}{2}} F\right\| \boldsymbol{u}_{11}, \qquad \boldsymbol{u}_{11}^T Q^{\frac{1}{2}} F = \left\| Q^{\frac{1}{2}} F\right\| \boldsymbol{v}_{11}^T$$

$$\boldsymbol{u}_{11}^T \boldsymbol{u}_{11} = 1, \qquad \boldsymbol{v}_{11}^T \boldsymbol{v}_{11} = 1$$

Hence,

$$\frac{\partial Q^{\frac{1}{2}} F}{\partial m_{ij}} \boldsymbol{v}_{11} + Q^{\frac{1}{2}} F \frac{\partial \boldsymbol{v}_{11}}{\partial m_{ij}} = \frac{\partial \left\| Q^{\frac{1}{2}} F\right\|}{\partial m_{ij}} \boldsymbol{u}_{11} + \left\| Q^{\frac{1}{2}} F\right\| \frac{\partial \boldsymbol{u}_{11}}{\partial m_{ij}}$$

which, with $\boldsymbol{u}_{11}^T \boldsymbol{u}_{11} = 1$, implies

$$\frac{\partial \left\| Q^{\frac{1}{2}} F\right\|}{\partial m_{ij}} = \boldsymbol{u}_{11}^T \frac{\partial Q^{\frac{1}{2}} F}{\partial m_{ij}} \boldsymbol{v}_{11} = \boldsymbol{u}_{11}^T \frac{\partial Q^{\frac{1}{2}}}{\partial m_{ij}} F \boldsymbol{v}_{11} = \mathrm{tr}\left(\frac{\partial Q^{\frac{1}{2}}}{\partial m_{ij}} F \boldsymbol{v}_{11} \boldsymbol{u}_{11}^T \right)$$

Observe that

$$Q^{\frac{1}{2}} \frac{\partial Q^{\frac{1}{2}}}{\partial m_{ij}} + \frac{\partial Q^{\frac{1}{2}}}{\partial m_{ij}} Q^{\frac{1}{2}} = \frac{\partial Q}{\partial m_{ij}}$$

Thus, with Lemma 2,

$$\frac{\partial \left\| Q^{\frac{1}{2}} F\right\|}{\partial m_{ij}} = \mathrm{tr}\left(\frac{\partial Q}{\partial m_{ij}} W_1 \right)$$

where W_1 is obtained from Eq. (18). On the other hand,

$$\frac{\partial \left\| Q^{\frac{1}{2}} K \right\|_F}{\partial m_{ij}} = \frac{\partial}{\partial m_{ij}} \sqrt{\operatorname{tr}(K^T Q K)} = \frac{1}{2\left\| Q^{\frac{1}{2}} K \right\|_F} \operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}} K K^T \right)$$

Here, $\partial Q / \partial m_{ij}$ is obtained via differentiating Eq. (7), which gives

$$G_L^T \frac{\partial Q}{\partial m_{ij}} G_L - \frac{\partial Q}{\partial m_{ij}} + \left(V^{-1} \left[\frac{\partial V}{\partial m_{ij}} LC - e_i e_j^T C \right] \right)^T Q G_L + G_L^T Q V^{-1} \left[\frac{\partial V}{\partial m_{ij}} LC - e_i e_j^T C \right] = 0$$

Now, using Lemma 2 again,

$$\begin{aligned} \frac{\partial J_{11}}{\partial m_{ij}} &= \frac{\varepsilon_1 \|[H_G \quad H_B]\|}{\left\| Q^{\frac{1}{2}} K \right\|_F} \operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}} W_1 \right) - \frac{\varepsilon_1 \|[H_G \quad H_B]\| \left\| Q^{\frac{1}{2}} F \right\|}{2\left\| Q^{\frac{1}{2}} K \right\|_F^3} \\ &\quad \times \operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}} K K^T \right) = \operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}} \left[\frac{\varepsilon_1 \|[H_G \quad H_B]\|}{\left\| Q^{\frac{1}{2}} K \right\|_F} W_1 - \frac{\varepsilon_1 \|[H_G \quad H_B]\| \left\| Q^{\frac{1}{2}} F \right\|}{2\left\| Q^{\frac{1}{2}} K \right\|_F^3} K K^T \right] \right) \\ &= \operatorname{tr}\left(\left(X_1 + X_1^T \right) G_L^T Q V^{-1} \left[\frac{\partial V}{\partial m_{ij}} LC - e_i e_j^T C \right] \right) \\ &= \operatorname{tr}\left(\frac{\partial V}{\partial m_{ij}} LC \left(X_1 + X_1^T \right) G_L^T Q V^{-1} \right) - e_j^T C \left(X_1 + X_1^T \right) G_L^T Q V^{-1} e_i \end{aligned}$$

where

$$G_L X_1 G_L^T - X_1 + \left[\frac{\varepsilon_1 \|[H_G \quad H_B]\|}{\left\| Q^{\frac{1}{2}} K \right\|_F} W_1 - \frac{\varepsilon_1 \|[H_G \quad H_B]\| \left\| Q^{\frac{1}{2}} F \right\|}{2\left\| Q^{\frac{1}{2}} K \right\|_F^3} K K^T \right] = 0$$

Since

$$\frac{\partial V}{\partial m_{ij}} G - \Lambda \frac{\partial V}{\partial m_{ij}} = e_i e_j^T C$$

and, hence,

$$\begin{aligned} \frac{\partial J_{11}}{\partial m_{ij}} &= \operatorname{tr}\left[\frac{\partial V}{\partial m_{ij}} LC \left(X_1 + X_1^T \right) G_L^T Q V^{-1} \right] - e_j^T C \left(X_1 + X_1^T \right) G_L^T Q V^{-1} e_i \\ &= \operatorname{tr}\left(Y_1 e_i e_j^T C \right) - e_j^T C \left(X_1 + X_1^T \right) G_L^T Q V^{-1} e_i \\ &= e_j^T C \left[Y_1 - \left(X_1 + X_1^T \right) G_L^T Q V^{-1} \right] e_i \end{aligned}$$

where

$$G Y_1 - Y_1 \Lambda = LC \left(X_1 + X_1^T \right) G_L^T Q V^{-1}$$

Consequently,

$$\frac{\partial J_{11}}{\partial M} = \left[Y_1^T - V^{-T} Q G_L \left(X_1 + X_1^T \right) \right] C^T$$

Now we consider J_{12},

$$\frac{\partial J_{12}}{\partial m_{ij}} = \frac{\varepsilon_2 \|H\|}{\left\| Q^{\frac{1}{2}} K \right\|_F} \frac{\partial \left\| Q^{\frac{1}{2}} [F_G \quad -L F_C] \right\|}{\partial m_{ij}} - \frac{\varepsilon_2 \|H\| \left\| Q^{\frac{1}{2}} [F_G \quad -L F_C] \right\|}{\left\| Q^{\frac{1}{2}} K \right\|_F^2} \frac{\partial \left\| Q^{\frac{1}{2}} K \right\|_F}{\partial m_{ij}}$$

Under the assumption, the maximum singular value of $Q^{1/2} F$ is distinct and

$$Q^{\frac{1}{2}} [F_G \quad -L F_C] \boldsymbol{v}_{12} = \left\| Q^{\frac{1}{2}} [F_G \quad -L F_C] \right\| \boldsymbol{u}_{12}$$

$$\boldsymbol{u}_{12}^T Q^{\frac{1}{2}} [F_G \quad -L F_C] = \left\| Q^{\frac{1}{2}} [F_G \quad -L F_C] \right\| \boldsymbol{v}_{12}^T$$

$$\boldsymbol{u}_{12}^T \boldsymbol{u}_{12} = 1, \qquad \boldsymbol{v}_{12}^T \boldsymbol{v}_{12} = 1$$

Hence,

$$\begin{aligned} &\frac{\partial Q^{\frac{1}{2}} [F_G \quad -L F_C]}{\partial m_{ij}} \boldsymbol{v}_{12} + Q^{\frac{1}{2}} [F_G \quad -L F_C] \frac{\partial \boldsymbol{v}_{12}}{\partial m_{ij}} \\ &\quad = \frac{\partial \left\| Q^{\frac{1}{2}} [F_G \quad -L F_C] \right\|}{\partial m_{ij}} \boldsymbol{u}_{12} + \left\| Q^{\frac{1}{2}} [F_G \quad -L F_C] \right\| \frac{\partial \boldsymbol{u}_{12}}{\partial m_{ij}} \end{aligned}$$

which, with $\boldsymbol{u}_{12}^T \boldsymbol{u}_{12} = 1$, implies

$$\begin{aligned} \frac{\partial \left\| Q^{\frac{1}{2}} [F_G \quad -L F_C] \right\|}{\partial m_{ij}} &= \boldsymbol{u}_{12}^T \frac{\partial Q^{\frac{1}{2}} [F_G \quad -L F_C]}{\partial m_{ij}} \boldsymbol{v}_{12} \\ &= \operatorname{tr}\left(\frac{\partial Q^{\frac{1}{2}} [F_G \quad -L F_C]}{\partial m_{ij}} \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T \right) \\ &= \operatorname{tr}\left(\frac{\partial Q^{\frac{1}{2}}}{\partial m_{ij}} [F_G \quad -L F_C] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T + Q^{\frac{1}{2}} \left[0 \quad -\frac{\partial L}{\partial m_{ij}} F_C \right] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T \right) \end{aligned}$$

Thus, with Lemma 2,

$$\operatorname{tr}\left(\frac{\partial Q^{\frac{1}{2}}}{\partial m_{ij}} [F_G \quad -L F_C] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T \right) = \operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}} W_2 \right)$$

where W_2 is obtained from Eq. (19). Also,

$$\operatorname{tr}\left(Q^{\frac{1}{2}} \left[0 \quad -\frac{\partial L}{\partial m_{ij}} F_C \right] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T \right) = \operatorname{tr}\left([0 \quad -F_C] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T Q^{\frac{1}{2}} \right)$$

where

$$\frac{\partial L}{\partial m_{ij}} = -V^{-1} \frac{\partial V}{\partial m_{ij}} L + V^{-1} e_i e_j^T$$

Hence,

$$\begin{aligned} &\operatorname{tr}\left(\frac{\partial L}{\partial m_{ij}} [0 \quad -F_C] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T Q^{\frac{1}{2}} \right) \\ &\quad = -\operatorname{tr}\left(\frac{\partial V}{\partial m_{ij}} L [0 \quad -F_C] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T Q^{\frac{1}{2}} V^{-1} \right) + e_j^T [0 \quad -F_C] \boldsymbol{v}_{12} \boldsymbol{u}_{12}^T Q^{\frac{1}{2}} V^{-1} e_i \end{aligned}$$

Now, when Lemma 2 is used again,

$$\frac{\partial J_{12}}{\partial m_{ij}} = \frac{\varepsilon_2\|H\|}{\left\|Q^{\frac{1}{2}}K\right\|_F}\left[\operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}}W - \frac{\partial V}{\partial m_{ij}}L[0 \quad -F_C] \times \boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}\right) + e_j^T[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}e_i\right]$$
$$- \frac{\varepsilon_2\|H\|\left\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\|}{2\left\|Q^{\frac{1}{2}}K\right\|_F^3}\operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}}KK^T\right)$$
$$= \operatorname{tr}\left(\frac{\partial Q}{\partial m_{ij}}\left[\frac{\varepsilon_2\|H\|}{\left\|Q^{\frac{1}{2}}K\right\|_F}W_2 - \frac{\varepsilon_2\|H\|\left\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\|}{2\left\|Q^{\frac{1}{2}}K\right\|_F^3}KK^T\right]\right)$$
$$- \operatorname{tr}\left(\frac{\partial V}{\partial m_{ij}}\frac{\varepsilon_2\|H\|L[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}\right) + e_j^T\frac{\varepsilon_2\|H\|[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}e_i$$
$$= \operatorname{tr}\left(\left(X_2 + X_2^T\right)G_L^T QV^{-1}\left[\frac{\partial V}{\partial m_{ij}}LC - e_i e_j^T C\right]\right)$$
$$+ \operatorname{tr}\left(-\frac{\partial V}{\partial m_{ij}}\frac{\varepsilon_2\|H\|L[0 \quad F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}\right) + e_j^T\frac{\varepsilon_2\|H\|[0 \quad F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}e_i$$
$$= \operatorname{tr}\left(\frac{\partial V}{\partial m_{ij}}LC\left(X_2 + X_2^T\right)G_L^T QV^{-1}\right) - e_j^T C\left(X_2 + X_2^T\right)G_L^T QV^{-1}e_i$$
$$+ \operatorname{tr}\left(-\frac{\partial V}{\partial m_{ij}}\frac{\varepsilon_2\|H\|L[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}\right) + e_j^T\frac{\varepsilon_2\|H\|[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}e_i$$
$$= \operatorname{tr}\left(\frac{\partial V}{\partial m_{ij}}\left(LC\left(X_2 + X_2^T\right)G_L^T QV^{-1} - \frac{\varepsilon_2\|H\|L[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}\right)\right)$$
$$+ e_j^T\left(\frac{\varepsilon_2\|H\|[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F} - C\left(X_2 + X_2^T\right)G_L^T QV^{-1}\right)e_i$$
$$= \operatorname{tr}\left(Y_2 e_i e_j^T C\right) + e_j^T\left(\frac{\varepsilon_2\|H\|[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F} - C\left(X_2 + X_2^T\right)G_L^T QV^{-1}\right)e_i$$
$$= e_j^T\left(CY_2 + \frac{\|H\|[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F} - C\left(X_2 + X_2^T\right)G_L^T QV^{-1}\right)e_i$$

where

$$G_L X_2 G_L^T - X_2 + \left[\frac{\varepsilon_2\|H\|}{\left\|Q^{\frac{1}{2}}K\right\|_F}W_2 - \frac{\varepsilon_2\|H\|\left\|Q^{\frac{1}{2}}[F_G \quad -LF_C]\right\|}{2\left\|Q^{\frac{1}{2}}K\right\|_F^3}KK^T\right] = 0$$
$$GY_2 - Y_2\Lambda = LC\left(X_2 + X_2^T\right)G_L^T QV^{-1} - \frac{\varepsilon_2\|H\|L[0 \quad -F_C]\boldsymbol{v}_{12}\boldsymbol{u}_{12}^T Q^{\frac{1}{2}}V^{-1}}{\left\|Q^{\frac{1}{2}}K\right\|_F}$$

Consequently,

$$\frac{\partial J_{12}}{\partial M} = \left[Y_2^T - V^{-T}QG_L\left(X_2 + X_2^T\right)\right]C^T + \frac{\varepsilon_2\|H\|V^{-T}Q^{\frac{1}{2}}\boldsymbol{u}_{12}\boldsymbol{v}_{12}^T[0 \quad -F_C]^T}{\left\|Q^{\frac{1}{2}}K\right\|_F}$$

They sum up to give

$$\frac{\partial J_1}{\partial M} = \frac{\partial J_{11}}{\partial M} + \frac{\partial J_{12}}{\partial M} = \left[Y_1^T - V^{-T}QG_L\left(X_1 + X_1^T\right)\right]C^T + Y_2^T C^T + \frac{\varepsilon_2\|H\|V^{-T}Q^{\frac{1}{2}}\boldsymbol{u}_{12}\boldsymbol{v}_{12}^T[0 \quad -F_C]^T}{\left\|Q^{\frac{1}{2}}K\right\|_F} - V^{-T}QG_L\left(X_2 + X_2^T\right)C^T$$
$$= \left[Y_1^T + Y_2^T - V^{-T}QG_L\left(X_1 + X_2 + X_1^T + X_2^T\right)\right]C^T + \frac{\varepsilon_2\|H\|V^{-T}Q^{\frac{1}{2}}\boldsymbol{u}_{12}\boldsymbol{v}_{12}^T[0 \quad -F_C]^T}{\left\|Q^{\frac{1}{2}}K\right\|_F}$$
$$= \left[Y^T - V^{-T}QG_L(X + X^T)\right]C^T + \frac{\varepsilon_2\|H\|V^{-T}Q^{\frac{1}{2}}\boldsymbol{u}_{12}\boldsymbol{v}_{12}^T[0 \quad -F_C]^T}{\left\|Q^{\frac{1}{2}}K\right\|_F}$$

where X and Y are solutions from Eqs. (20) and (21).

Second, we consider J_2, with the assumption that the maximum singular value of L is distinct and

$$L\boldsymbol{v}_2 = \|L\|\boldsymbol{u}_2, \qquad \boldsymbol{u}_2^T L = \|L\|\boldsymbol{v}_2^T$$
$$\boldsymbol{u}_2^T\boldsymbol{u}_2 = 1, \qquad \boldsymbol{v}_2\boldsymbol{v}_2^T = 1$$

Then, similar to J_1, we have

$$\frac{\partial\|L\|}{\partial m_{ij}} = \boldsymbol{u}_2^T\frac{\partial L}{\partial m_{ij}}\boldsymbol{v}_2 = \operatorname{tr}\left(\frac{\partial L}{\partial m_{ij}}\boldsymbol{v}_2\boldsymbol{u}_2^T\right)$$

Hence,

$$\frac{\partial J_2}{\partial m_{ij}} = -\mathrm{tr}\left(\frac{\partial V}{\partial m_{ij}} L \boldsymbol{v}_2 \boldsymbol{u}_2^T V^{-1}\right) + e_j^T \boldsymbol{v}_2 \boldsymbol{u}_2^T V^{-1} e_i$$

$$= -\|L\| \mathrm{tr}\left(\frac{\partial V}{\partial m_{ij}} \boldsymbol{u}_2 \boldsymbol{u}_2^T V^{-1}\right) + e_j^T \boldsymbol{v}_2 \boldsymbol{u}_2^T V^{-1} e_i$$

$$= -\mathrm{tr}\left(Z e_i e_j^T C\right) + e_j^T \boldsymbol{v}_2 \boldsymbol{u}_2^T V^{-1} e_i$$

where Z is obtained from Eq. (22). Consequently,

$$\frac{\partial J_2}{\partial M} = -Z^T C^T + V^{-T} \boldsymbol{u}_2 \boldsymbol{v}_2^T$$

Third, consider $J_3 = \|V\| \|V^{-1}\|$ and the assumption that the maximum singular values of V and V^{-1} are distinct. We have

$$\frac{\partial J_3}{\partial m_{ij}} = \|V^{-1}\| \frac{\partial \|V\|}{\partial m_{ij}} + \|V\| \frac{\partial \|V^{-1}\|}{\partial m_{ij}}$$

with a SVD on V,

$$V x_1 = \|V\| \boldsymbol{y}_1, \qquad \boldsymbol{y}_1^T V = \|V\| \boldsymbol{x}_1^T, \qquad \boldsymbol{x}_1^T \boldsymbol{x}_1 = 1, \qquad \boldsymbol{y}_1^T \boldsymbol{y}_1 = 1$$

$$V \boldsymbol{x}_n = (1/\|V^{-1}\|) \boldsymbol{y}_n, \qquad \boldsymbol{y}_n^T V = (1/\|V^{-1}\|) \boldsymbol{x}_n^T$$

$$\boldsymbol{x}_n^T \boldsymbol{x}_n = 1, \qquad \boldsymbol{y}_n^T \boldsymbol{y}_n = 1$$

then

$$\frac{\partial \|V\|}{\partial m_{ij}} = \mathrm{tr}\left(\frac{\partial V}{\partial m_{ij}} \boldsymbol{x}_1 \boldsymbol{y}_1^T\right), \qquad \frac{\partial \|V^{-1}\|}{\partial m_{ij}} = -\|V^{-1}\|^2 \mathrm{tr}\left(\frac{\partial V}{\partial m_{ij}} \boldsymbol{x}_n \boldsymbol{y}_n^T\right)$$

Thus,

$$\frac{\partial J_3}{\partial m_{ij}} = \mathrm{tr}\left(\frac{\partial V}{\partial m_{ij}} \left[\|V^{-1}\| \boldsymbol{x}_1 \boldsymbol{y}_1^T - \|V\| \|V^{-1}\|^2 \boldsymbol{x}_n \boldsymbol{y}_n^T\right]\right) = e_j^T C U e_i$$

where U is obtained from Eq. (23). Hence, $\partial J_3 / \partial M = U^T C^T$. □

Acknowledgment

This research was partially supported by Germany/Hong Kong Joint Research Scheme G_HK008/00.

References

[1]Frank, P. M., "Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-Based Redundancy—A Survey and Some New Results," *Automatica*, Vol. 26, No. 3, 1990, pp. 459–474.

[2]Gertler, J., "Survey of Model-Based Failure Detection and Isolation in Complex Plants," *IEEE Control Systems Magazine*, Vol. 8, No. 6, 1988, pp. 3–11.

[3]Isermann, R., "Process Fault Detection Based on Modelling and Estimation Methods: A Survey," *Automatica*, Vol. 20, No. 4, 1984, pp. 387–404.

[4]Patton, R. J., "Robust Model-Based Fault Diagnosis: The State of the Art," *IFAC Symposium on Fault Detection, Supervision and Safety for Technical Process*, Academic Press, New York, 1994, pp. 1–27.

[5]Patton, R. J., Frank, P. M., and Clark, R. N., *Fault Diagnosis in Dynamic Systems, Theory and Applications*, Prentice–Hall, London, 1989.

[6]Willsky, A. S., "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, Vol. 12, No. 4, 1976, pp. 601–611.

[7]Frank, P. M., "Enhancement of Robustness in Observer-Based Fault Detection," *International Journal of Control*, Vol. 59, No. 4, 1994, pp. 955–981.

[8]Frank, P. M., and Ding, X., "Survey of Robust Residual Generation and Evaluation Methods in Observer-Based Fault Detection Systems," *Journal of Process Control*, Vol. 7, No. 6, 1997, pp. 403–424.

[9]Frank, P. M., Koeppen, B., and Wunnenberg, J., "General Solution of Robustness Problem in Linear-Fault Detection Filters," *1st European Control Conference*, Hérmes, Paris, 1991, pp. 1407–1412.

[10]Lou, X. C., Willsky, A. S., and Verghese, G. L., "Optimally Robust Redundancy Relations for Failure Detection in Uncertain Systems," *Automatica*, Vol. 22, No. 3, 1986, pp. 333–344.

[11]Patton, R. J., and Chen, J., "Robust Fault Detection of Jet Engine Sensor Systems Using Eigenstructure Assignment," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 6, 1992, pp. 1491–1497.

[12]Shen, L. C., Chang, S. K., and Hsu, P. L., "Robust Fault Detection and Isolation with Unstructured Uncertainty Using Eigenstructure Assignment," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 1, 1997, pp. 50–57.

[13]Viswanadham, N., and Srichander, R., "Fault Detection Using Unknown Input Observers," *Control Theory and Advanced Technology*, Vol. 3, No. 2, 1987, pp. 91–101.

[14]Ding, X., and Frank, P. M., "Fault Detection via Optimally Robust Detection Filters," *Proceedings of the 28th IEEE Conference on Decision and Control*, Vol. 2, Inst. of Electrical and Electronics Engineers, New York, 1989, pp. 1767–1772.

[15]Ding, X., and Frank, P. M., "Fault Detection via Factorization Approach," *Systems and Control Letters*, Vol. 14, No. 5, 1990, pp. 431–436.

[16]Ding, X., Guo, L., and Frank, P. M., "A Frequency Domain Approach to Fault Detection of Uncertain Dynamic Systems," *Proceedings of the 32nd IEEE Conference on Decision and Control*, Vol. 2, Inst. of Electrical and Electronics Engineers, New York, 1993, pp. 1722–1727.

[17]Ding, X., Guo, L., and Frank, P. M., "Parameterization of Linear Observers and Its Application to Observer Design," *IEEE Transactions on Automatic Control*, Vol. 39, No. 8, 1994, pp. 1648–1652.

[18]Patton, R. J., and Chen, J., "Review of Parity Space Approaches to Fault Diagnosis for Aerospace Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 17, No. 2, 1994, pp. 278–284.

[19]Wuennenberg, J., "Observer-Based Fault Detection in Dynamic Systems," Ph.D. Dissertation, Univ. of Duisburg, Duisburg, Germany, 1990.

[20]Xie, L., and de Souza, C. E., "Robust H_∞ Control for Linear Time-Invariant Systems with Norm-Bounded Uncertainty in the Input Matrix," *Systems and Control Letters*, Vol. 14, No. 5, 1990, pp. 389–396.

[21]Qiu, L., Bernhardsson, B., Rantzer, A., Davison, E. J., Young, P. M., and Doyle, J. C., "A Formula for Computation of the Real Stability Radius," *Automatica*, Vol. 31, No. 6, 1995, pp. 879–890.

[22]Rotea, M. A., Corless, M., Da, D., and Petersen, I. R., "Systems with Structured Uncertainty: Relations Between Quadratic and Robust Stability," *IEEE Transactions on Automatic Control*, Vol. 93, No. 5, 1993, pp. 799–803.

[23]Bauer, F. L., and Fike, C. T., "Norms and Exclusion Theorems," *Numerische Mathematik*, Vol. 2, 1960, pp. 137–141.

[24]Chen, J., Patton, R. J., and Zhang, H., "Design of Unknown Input Observers and Robust Fault Detection Filters," *International Journal of Control*, Vol. 63, No. 1, 1996, pp. 85–105.

[25]Edelmayer, A., Bokor, J., Szigeti, F., and Keviczky, L., "Robust Detection Filter Design in the Presence of Time-Varying System Perturbations," *Automatica*, Vol. 33, No. 3, 1997, pp. 471–475.

[26]Park, J., and Rizzoni, G., "An Eigenstructure Assignment Algorithm for the Design of Detection Filters," *IEEE Transactions on Automatic Control*, Vol. 39, No. 7, 1994, pp. 1521–1524.

[27]Isermann, R., "Supervision, Fault-Detection and Fault-Diagnosis Methods—An Introduction," *Control Engineering Practice*, Vol. 5, No. 5, 1997, pp. 639–652.

[28]Dennis, J., and Schnabel, R., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, Upper Saddle River, NJ, 1983.

[29]Saif, M., "Robust Discrete Time Observer with Application to Fault Diagnosis," *IEE Proceedings: Control Theory and Applications*, Vol. 145, No. 3, 1998, pp. 353–357.

[30]Saif, M., and Guan, Y., "A New Approach to Robust Fault Detection and Identification," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 29, No. 3, 1993, pp. 685–695.

[31]Narendra, K., and Tripathi, S. S., "Identification and Optimization of Aircraft Dynamics," *Journal of Aircraft*, Vol. 10, No. 4, 1973, pp. 193–199.