Optimization of Observer Design for Sensor Fault Detection of Nonlinear Systems

Vasso Reppa, Stelios Timotheou, Marios M. Polycarpou and Christos G. Panayiotou

Abstract— **This paper presents an optimization methodology for the design of an observer-based sensor fault detection scheme for a class of nonlinear systems. Taking into account bounded system disturbances and measurement noise, we design an observer aiming at maximizing the set of faults that are guaranteed to be strongly detectable. Strong fault detectability conditions are derived based on the limit sets that bound the residual under healthy and faulty conditions. A novel optimization method is designed based on the separation of the healthy and faulty limit sets. The distance between these sets represents the trade-off between robustness and sensor fault sensitivity. Simulation results are used to show the effectiveness of the proposed methodology applied to a simple example of a flexible link robot.**

I. INTRODUCTION

Among the popular quantitative model-based sensor fault detection (FD) techniques are the observer-based techniques commonly developed in a bounded-error framework [1]. Their design is usually realized based on advanced control theory tools, a fact that has motivated several researchers to follow the observer-based approach for the FD of nonlinear systems [2]–[6]. The textbooks [7], [8] offer an in-depth description of observer-based FD methods for linear systems with emphasis on diagnostic and unknown input observer design.

The basic stages of the observer-based FD methods are the residual generation, the residual evaluation, the threshold computation and the fault detection decision logic. The existence of modeling uncertainties and the associated assumptions play a key role in these stages. One efficient way to deal with their effects is to use a priori known bounds of the modeling uncertainties in the residual evaluation and threshold computation [2], [3], [6]. Then, the residual can be evaluated using some norm, or component-wise bounds. The evaluated residual is compared to a threshold, which can be either fixed or adaptive. The fault detection decision logic is based on checking whether the evaluated residual exceeds the corresponding threshold. When the design specification focuses on eliminating false alarms, the thresholds can be designed to bound the evaluated residual taking into account the effects of modeling uncertainties. While this way of computing the thresholds ensures robustness by design, it does not ensure the fault detectability.

To tackle the trade-off between robustness and fault detectability, several optimization-based approaches have been proposed for the design of observer-based FD methods for linear systems. Following these approaches, the observer parameters are commonly selected such that the effects of modeling uncertainties on the residual are minimized assuming no faults, while the fault effects are maximized assuming however no modeling uncertainties [7]. Recently, optimization techniques were formulated taking into account the effects of disturbances in the fault sensitivity analysis, and the impact of a pre-defined threshold on the decision logic [8].

In general, (weak) fault detectability is not enough in order to guarantee the reliability of the FD method [7], [9]. Instead, strong fault detectability should be sought during the FD design. This property implies that the FD scheme generates a persistently excited or else asymptotically non-vanishing residual signal under a non-zero persistent excitation of a fault signal [10]. In previous work of the authors [6], [11], [12], guaranteed conditions for observer-based sensor fault diagnosis for nonlinear systems were derived for posterior analysis of the weak and strong fault detectability. However, these conditions were not exploited in the design phase of the residual generation and the adaptive threshold computation.

The objective and main contribution of this work is the optimization of an observer-based sensor fault detection method for nonlinear systems. The fault detection decision logic is based on checking at every time instant whether the residual lies within a convex set that is a function of the adaptive thresholds (see Section II). Under healthy conditions, it is proved that this adaptive set converges to a limit set associated with the robust positively invariant (RPI) set that asymptotically encloses the state estimation error dynamics (see Section III). Under faulty conditions, we define the residual as a function of the effects of faults and uncertainties. Based on this design, we compute a limit set, related to the RPI set that asymptotically encloses the fault effects on the state estimation error (see Section III). Strong detectability conditions are derived based on the separation between the limit sets that enclose the residual under healthy and faulty conditions. The optimization problem is to find the eigenvalues and eigenvectors of the observer that maximize the class of faults that are guaranteed to be strongly detectable, as shown in Section IV. The application of the proposed methodology to a simple example of a flexible link robot is illustrated in Section V.

This work was supported by the European Research Council (ERC) under the ERC Advanced Grant ERC-2011-AdG-291508 and by the European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE).

The authors are with the KIOS Research and Innovation Center of Excellence, Electrical & Computer Engineering Department, University of Cyprus, Nicosia, 1678, Cyprus reppavasso@gmail.com, {timotheou.stelios,mpolycar,christosp}@ucy.ac.cy

Notation: The symbols $\preceq$, $\succeq$ signify the element-wise inequalities, $|\cdot|$ denotes the element-wise absolute value (matrix modulus function), $\oplus$ symbolizes the Minkowski sum of sets, and $\|\cdot\|$ denotes a norm and $\mathbf{1}$ denotes a vector of all ones.

II. PROBLEM FORMULATION

We consider the class of nonlinear uncertain systems

$$\dot{x}(t) = Ax(t) + \gamma(x(t), u(t)) + \eta(t) \tag{1}$$

where $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^\ell$ is the input vector, $A \in \mathbb{R}^{n \times n}$ is a matrix representing the linearized part of the state equation, $\gamma : \mathbb{R}^n \times \mathbb{R}^\ell \mapsto \mathbb{R}^n$ characterizes the nonlinear part of system dynamics, and $\eta \in \mathbb{R}^n$ represents unknown additive disturbances. For the well-posedness of the problem, the state and input vectors satisfy the following assumption:

Assumption 1: The state vector $x(t)$ and input vector $u(t)$ generated by a feedback controller remain uniformly bounded before and after the occurrence of sensor faults; i.e., there exist a compact region of stability $\mathscr{X} \subset \mathbb{R}^n$ and admissible control inputs $u \in \mathscr{U}$ with $\mathscr{U}$ compact in $\mathbb{R}^\ell$ such that $(x(t), u(t)) \in \mathscr{X} \times \mathscr{U}$, for all $t \geq 0$.

The nonlinear system (1) is monitored by n sensors, characterized by the output vector $y \in \mathbb{R}^n$; i.e.,

$$y(t) = x(t) + d(t) + f(t), \tag{2}$$

where $d \in \mathbb{R}^n$ is the noise and $f \in \mathbb{R}^n$ represents the abrupt, permanent change in the output y due to a single fault in sensor i that occurs at $T \in \mathbb{R}$ [6].

The residual vector is defined as

$$\begin{aligned} \varepsilon_y(t) &= y(t) - \hat{x}(t), && (3)\\ \dot{\hat{x}}(t) &= (A-L)\hat{x}(t) + \gamma(y(t), u(t)) + Ly(t), && (4) \end{aligned}$$

where $\hat{x} \in \mathbb{R}^n$ is the estimation of x ($\hat{x}(0) \in \mathscr{X}$) and $L \in \mathbb{R}^{n \times m}$ is the observer gain matrix chosen such that $A-L$ is stable. Let us define the diagonizable form of $A-L$ as

$$A - L = V \Lambda V^{-1} \tag{5}$$

where Λ is a diagonal matrix containing the eigenvalues and V is a matrix comprised of the eigenvectors of $A-L$.

Assuming no faults (healthy conditions) , we define

$$\begin{aligned} y_H &= x + d && (6)\\ \dot{\hat{x}}_H &= V\Lambda V^{-1}\hat{x}_H + \gamma(y_H, u) + (A - V\Lambda V^{-1})y_H, && (7)\\ \varepsilon_{x_H} &= x - \hat{x}_H && (8)\\ \varepsilon_{y_H} &= y_H - \hat{x}_H = \varepsilon_{x_H} + d && (9) \end{aligned}$$

Based on (1), (5) and (7), ε_{x_H} defined in (8) satisfies

$$V^{-1}\dot{\varepsilon}_{x_H}(t) = \Lambda V^{-1}\varepsilon_{x_H}(t) + z(t) \tag{10}$$

where

$$z = V^{-1}\left(\gamma(x,u) - \gamma(y_H,u) + \eta\right) + \left(\Lambda V^{-1} - V^{-1}A\right)d \tag{11}$$

The adaptive threshold, denoted by $\bar{\varepsilon}_y(t)$ is computed to bound the residual ε_{y_H} defined in (9); i.e.,

$$|V^{-1}\varepsilon_{y_H}(t)| = |V^{-1}\left(\varepsilon_{x_H}(t) + d(t)\right)| \preceq \bar{\varepsilon}_y(t) \tag{12}$$

According to (12), the residual under healthy conditions lies in the convex set $\mathscr{P}_H(t)$ defined as

$$\mathscr{P}_H(t) = \left\{ \varepsilon \in \mathbb{R}^n : \begin{bmatrix} V^{-1} \\ -V^{-1} \end{bmatrix} \varepsilon \preceq \begin{bmatrix} \bar{\varepsilon}_y(t) \\ \bar{\varepsilon}_y(t) \end{bmatrix} \right\} \tag{13}$$

Thus, we monitor whether the residual defined in (3) lies in the parallelotope $\mathscr{P}_H(t)$ defined in (13).

Fault Detection Criterion: If there is a time instant $t \geq T$ such that

$$\varepsilon_y(t) \notin \mathscr{P}_H(t) \tag{14}$$

then a sensor fault $f(t)$ is guaranteed to be detected.

The time of fault detection is defined as

$$t_D = \min\left\{ t : \varepsilon_y(t) \notin \mathscr{P}_H(t) \right\} \tag{15}$$

Based on the aforementioned criterion, we characterize the detectability of permanent sensor faults as follows [10].

Weak Detectability: Permanent sensor faults are considered as weakly detectable, if (14) is valid for some time instant $t \geq t_D > T$.

Strong Detectability: Permanent sensor faults are considered as strongly detectable, if (14) is valid for all $t \geq t_D > T$.

The objective of this work is to design Λ, V defined in (5), which are used in the residual generation described by (3) and (4) and in the adaptive threshold computation given in (12), that maximize the set of sensor faults that are guaranteed to be strongly detectable, based on the following assumptions.

Assumption 2: The nonlinear function $\gamma_i \in \mathbb{R}$, $i \in \{1, \ldots, n\}$ is locally Lipschitz on x and for any $u \in \mathscr{U}$ and $t \geq 0$.

Assumption 3: The system disturbance η and the measurement noise d are uniformly bounded; i.e., $|\eta(t)| \preceq \overline{\eta}$, $|d(t)| \preceq \overline{d}$, where $\overline{\eta}, \overline{d} \in \mathbb{R}^n$ are known vector bounds.

Assumption 4: The sensor fault vector f is unknown but uniformly bounded; i.e., $|f(t)| \prec \infty$.

III. SENSOR FAULT DETECTABILITY

A. Adaptive and Limit Sets Under Healthy Conditions

In this section, we present the analytic computation of the adaptive set $\mathscr{P}_H(t)$ and its limit set $\mathscr{P}(\infty) = \lim_{t\to\infty} \mathscr{P}_H(t)$ that encloses the residual asymptotically; i.e.,

$$\varepsilon_{y_H}(\infty) \triangleq \limsup_{t\to\infty} \left|\varepsilon_{y_H}(t)\right| \in \mathscr{P}_H(\infty) \tag{16}$$

Lemma 3.1: If the eigenvalues Λ are selected to be real, negative, then for all initial conditions, the state estimation error under healthy conditions $\varepsilon_{x_H}(t)$ is: (a) *bounded at every time instant*, satisfying

$$\left|V^{-1}\varepsilon_{x_H}(t)\right| \preceq E(t) \tag{17}$$

where

$$\begin{aligned} E(t) &= e^{\Lambda t}\left|V^{-1}\right|\bar{x} + \left(I - e^{\Lambda t}\right)\chi_H && (18)\\ \chi_H &= -\Lambda^{-1}\left(\left|V^{-1}\right|(\Gamma\|\bar{d}\| + \bar{\eta}) + \left|\Lambda V^{-1} - V^{-1}A\right|\bar{d}\right) && (19) \end{aligned}$$

with $\Gamma = \begin{bmatrix} \lambda_{\gamma_1} & \cdots & \lambda_{\gamma_n} \end{bmatrix}^\top$, where λ_{γ_i} is the Lipschitz constant related to γ_i (see Assumption 2), and $\bar{x} \in \mathbb{R}^n$

denoting a bound such that $|\varepsilon_{x_H}(0)| \preceq \bar{x}$ (see Assumption 1), (b) *ultimately bounded*, satisfying

$$\limsup_{t\to\infty} \left|V^{-1}\varepsilon_{x_H}(t)\right| \preceq \chi_H \tag{20}$$

The ultimate bound in (20) defines the set $\Omega_{x_H}(\infty) = \{\varepsilon \in \mathbb{R}^n : \left|V^{-1}\varepsilon\right| \preceq \chi_H\}$, which is RPI with respect to (10); i.e. if $|V^{-1}\varepsilon_{x_H}(0)| \preceq \chi_H$, then $|V^{-1}\varepsilon_{x_H}(t)| \preceq \chi_H$ for all $t \geq 0$.

Proof: (a) Solving (10) yields

$$V^{-1}\varepsilon_{x_H} = e^{\Lambda t}V^{-1}\varepsilon_{x_H}(0) + \int_0^t e^{\Lambda(t-\tau)}z(\tau)d\tau \tag{21}$$

where z is defined in (11). Assumption 2 implies

$$|\gamma(x,u) - \gamma(y_H,u)| \preceq \Gamma\left\|\bar{d}\right\| \tag{22}$$

Based on (22) and Assumption 3, we have

$$|z| \preceq \left|V^{-1}\right|\left(\Gamma\left\|\bar{d}\right\| + \bar{\eta}\right) + \left|\Lambda V^{-1} - V^{-1}A\right|\bar{d} = -\Lambda\chi_H \tag{23}$$

where χ_H is defined in (19). A bound on $\left|V^{-1}\varepsilon_{x_H}\right|$ satisfies

$$\left|V^{-1}\varepsilon_{x_H}\right| \preceq e^{\Lambda t}\left|V^{-1}\varepsilon_{x_H}(0)\right| - \int_0^t e^{\Lambda(t-\tau)}\Lambda\chi_H d\tau \tag{24}$$

By using the conditions of Lemma 3.1, we obtain (17)-(18). (b)Equation (18) for $t \to \infty$ becomes $\lim_{t\to\infty} E(t) = \chi_H$. By using (23) and applying Theorem 1 of [13], it is proved that the set $\Omega_{x_H}(\infty)$ in Lemma 3.1 is RPI. ■

Equation (9) can be equivalently written as $V^{-1}\varepsilon_{y_H} = V^{-1}\varepsilon_{x_H} + V^{-1}d$. Based on (12) and (17), the adaptive threshold $\bar{\varepsilon}_y(t)$ is analytically defined as

$$\left|V^{-1}\varepsilon_{y_H}(t)\right| \preceq \left|V^{-1}\varepsilon_{x_H}(t)\right| + \left|V^{-1}d(t)\right| \preceq \bar{\varepsilon}_y(t) \tag{25}$$

$$\bar{\varepsilon}_y(t) = E(t) + \left|V^{-1}\right|\bar{d} \tag{26}$$

Taking into account (25) for $t \to \infty$ and (20), it yields that ε_{y_H} is ultimately bounded satisfying (16), where $\mathscr{P}_H(\infty)$ represents the limit set associated with the RPI set of the state estimation error dynamics defined in Lemma 3.1, defined as

$$\mathscr{P}_H(\infty) = \left\{\varepsilon \in \mathbb{R}^n : \begin{bmatrix} V^{-1} \\ -V^{-1} \end{bmatrix}\varepsilon \preceq \begin{bmatrix} \upsilon_H \\ \upsilon_H \end{bmatrix}\right\} \tag{27}$$

$$\upsilon_H = \chi_H + \left|V^{-1}\right|\bar{d} \tag{28}$$

B. Adaptive and Limit Sets under Faulty Conditions

The effects of sensor faults on the residual $\varepsilon_y(t)$ can be determined as

$$\varepsilon_{y_F}(t) = \varepsilon_y(t) - \varepsilon_{y_H}(t), \quad t \geq T \tag{29}$$

where ε_{y_H} is defined in (9). For $t \geq T$, $\hat{x}$ satisfies

$$\dot{\hat{x}} = V\Lambda V^{-1}\hat{x} + \gamma(y_H + f, u) + (A - V\Lambda V^{-1})(y_H + f) \tag{30}$$

with y_H defined in (6). Based on (2) and (6) the residual is defined as

$$\varepsilon_y = y - \hat{x} = y_H + f - \hat{x} \tag{31}$$

By combining (9), (29) and (31), the sensor fault effects can be expressed as

$$\varepsilon_{y_F} = -\varepsilon_{x_F} + f, \tag{32}$$

$$\varepsilon_{x_F} = \hat{x} - \hat{x}_H \tag{33}$$

where ε_{x_F} represents the sensor fault effects on the state estimation. For $t \geq T$, $f(t)$ can be described as

$$f(t) = f_0 + \tilde{f}(t), \tag{34}$$

where $f_0 \in \mathbb{R}^n$ is the (unknown) non-zero offset and $\tilde{f}(t) \in \mathbb{R}^n$ represents the deviation of $f(t)$ from f_0. Based on Assumption 4, $\tilde{f}(t)$ is assumed to be bounded; i.e.,

$$\left|\tilde{f}(t)\right| = \left|f(t) - f_0\right| \preceq \bar{f} \tag{35}$$

where $\bar{f} \in \mathbb{R}^n$. In the special case of bias (e.g. constant) faults, it is considered that $\tilde{f} = \bar{f} = 0$. Using (5), (7), (30) and (34), ε_{x_F} defined in (33) satisfies

$$V^{-1}\dot{\tilde{\varepsilon}}_{x_F} = \Lambda V^{-1}\tilde{\varepsilon}_{x_F} + z_f \tag{36}$$

$$z_f = V^{-1}(\gamma(y_H + f_0 + \tilde{f}, u) - \gamma(y_H, u)) + (V^{-1}A - \Lambda V^{-1})\tilde{f} \tag{37}$$

$$V^{-1}\tilde{\varepsilon}_{x_F}(t) = V^{-1}\varepsilon_{x_F}(t) + (\Lambda^{-1}V^{-1}A - V^{-1})f_0 \tag{38}$$

where $\varepsilon_{x_F}(T) = \hat{x}(T) - \hat{x}_H(T) = 0$, $V^{-1}\tilde{\varepsilon}_{x_F}(T) = (\Lambda^{-1}V^{-1}A - V^{-1})f_0$ and z_f satisfies

$$\left|z_f\right| \preceq |V^{-1}|\Gamma\left(\|f_0\| + \left\|\bar{f}\right\|\right) + |\Lambda V^{-1} - V^{-1}A|\bar{f} \tag{39}$$

The derivation of bounds for $V^{-1}\tilde{\varepsilon}_{x_F}(t)$ are obtained next.

Lemma 3.2: If the eigenvalues Λ are real, negative, $\tilde{\varepsilon}_{x_F}$ defined in (36)-(38) is: (a) *bounded at every time instant*; i.e.,

$$\left|V^{-1}\tilde{\varepsilon}_{x_F}(t)\right| \preceq E_F(t), \tag{40}$$

$$E_F(t) = e^{\Lambda(t-T)}|(\Lambda^{-1}V^{-1}A - V^{-1})f_0| + \left(I - e^{\Lambda(t-T)}\right)\chi_F \tag{41}$$

$$\chi_F = -\Lambda^{-1}\left(|V^{-1}|\Gamma\left(\|f_0\| + \left\|\bar{f}\right\|\right) + |\Lambda V^{-1} - V^{-1}A|\bar{f}\right) \tag{42}$$

with $\Gamma = \begin{bmatrix} \lambda_{\gamma_1} & \cdots & \lambda_{\gamma_n} \end{bmatrix}^\top$, where λ_{γ_i} is the Lipschitz constant related to γ_i (see Assumption 2), and (b) *ultimately bounded*; i.e.

$$\left|V^{-1}\limsup_{t\to\infty} \tilde{\varepsilon}_{x_F}(t)\right| \preceq \chi_F \tag{43}$$

The ultimate bound in (43) defines the set $\Omega_{x_F} = \left\{\varepsilon \in \mathbb{R}^n : \left|V^{-1}\varepsilon\right| \preceq \chi_F\right\}$, which is RPI with respect to (36); i.e., if $|V^{-1}\tilde{\varepsilon}_{x_F}(T)| \preceq \chi_F$, then $|V^{-1}\tilde{\varepsilon}_{x_F}(t)| \preceq \chi_F$ for all $t \geq T$.

Proof: The proof is similar to Lemma 3.1. ■

Taking into account (32), (34) and (38), we have

$$V^{-1}\tilde{\varepsilon}_{y_F} = -V^{-1}\tilde{\varepsilon}_{x_F} + V^{-1}\tilde{f} \tag{44}$$

$$V^{-1}\tilde{\varepsilon}_{y_F} = V^{-1}\varepsilon_{y_F} - \Lambda^{-1}V^{-1}Af_0 \tag{45}$$

Bounding (44) and using (40) results in

$$\left|V^{-1}\tilde{\varepsilon}_{y_F}\right| \preceq \left|V^{-1}\tilde{\varepsilon}_{x_F}\right| + \left|V^{-1}\right|\bar{f} \preceq E_F + \left|V^{-1}\right|\bar{f} \tag{46}$$

where E_F is defined in (41). Taking into account (46) for $t \to \infty$ and that $\lim_{t\to\infty} E_F(t) = \chi_F$, it yields that ε_{y_F} lies in the limit set (parallelotope) $\mathscr{P}_F(\infty)$, which is associated with the RPI set Ω_{x_F} defined in Lemma 3.2; i.e. $\varepsilon_{y_F}(\infty) \triangleq \lim_{t\to\infty} \varepsilon_{y_F}(t)$ lies in $\mathscr{P}_F(\infty)$, which is centered at $V\Lambda^{-1}V^{-1}Af_0$; i.e.,

$$\varepsilon_{y_F}(\infty) \in \mathscr{P}_F(\infty) \tag{47}$$

$$\mathscr{P}_F(\infty) = \left\{\varepsilon \in \mathbb{R}^n : \begin{bmatrix} V^{-1} \\ -V^{-1} \end{bmatrix}(\varepsilon - Qf_0) \preceq \begin{bmatrix} \upsilon_F \\ \upsilon_F \end{bmatrix}\right\} \tag{48}$$

where $Q = V\Lambda^{-1}V^{-1}A$ and $\upsilon_F = \chi_F + \left|V^{-1}\right|\bar{f}$. Taking into account (29), $\varepsilon_y(\infty) = \varepsilon_{y_F}(\infty) + \varepsilon_{y_H}(\infty)$. Based on (16), (27), (28), (47) and (48), we have

$$\varepsilon_y(\infty) \in \{\mathscr{P}_H(\infty) \oplus \mathscr{P}_F(\infty)\} \triangleq \mathscr{P}(\infty) \tag{49}$$

$$\mathscr{P}(\infty) = \left\{\varepsilon \in \mathbb{R}^n : \begin{bmatrix} V^{-1} \\ -V^{-1} \end{bmatrix} (\varepsilon - Qf_0) \preceq \begin{bmatrix} \upsilon_F + \upsilon_H \\ \upsilon_F + \upsilon_H \end{bmatrix}\right\} \tag{50}$$

C. Strong Detectability Conditions

Strong fault detectability can be guaranteed based on the separation of the limit sets $\mathscr{P}(\infty)$ defined in (50) and $\mathscr{P}_H(\infty)$ defined in (27) that characterize the asymptotic behavior of ε_y under faulty and healthy conditions, respectively [14].

Lemma 3.3: If $\mathscr{P}(\infty) \cap \mathscr{P}_H(\infty) = \emptyset$, then $\varepsilon_y(\infty) \notin \mathscr{P}_H(\infty)$ and (14) is valid for $t \to \infty$.

Proof: Suppose that $\varepsilon_y(\infty) \in \mathscr{P}_H(\infty)$. Given (49), it yields $\varepsilon_y(\infty) \in \mathscr{P}(\infty) \cap \mathscr{P}_H(\infty)$. This contradicts our assumption that $\mathscr{P}(\infty) \cap \mathscr{P}_H = \emptyset$. So $\varepsilon_y(\infty) \notin \mathscr{P}_H(\infty)$. ■

The limit sets $\mathscr{P}(\infty)$ and $\mathscr{P}_H(\infty)$ are separated, if their distance along the j-th eigenvector, denoted by D_j is positive, where D_j is defined as

$$D_j = (|\lambda_j^{-1}V_j^{-1}Af_0| - \chi_{Fj} - |V_j^{-1}|\bar{f} - 2\upsilon_{Hj})/||V_j^{-1}|| \tag{51}$$

with V_j^{-1} and λ_j being the j-th row of V^{-1} and the j-th diagonal element of Λ, respectively; i.e. given (19), (28) and (42), if for at least one $j \in \{1,\ldots,n\}$

$$\begin{aligned} |\lambda_j^{-1}V_j^{-1}Af_0| >& |\lambda_j^{-1}V_j^{-1}|(2\bar{\eta} + \Gamma(2||d|| + ||f_0|| + ||\bar{f}||)) \\ &+ (|V_j^{-1} - \lambda_j^{-1}V_j^{-1}A| + |V_j^{-1}|)(2\bar{d} + \bar{f}) \end{aligned} \tag{52}$$

then $\mathscr{P}(\infty) \cap \mathscr{P}_H(\infty) = \emptyset$. Let us consider an abrupt permanent fault affecting a single sensor, i.e. $f_i(t) \neq 0$ for all $t \geq T$, while $f_q(t) = 0$ for all $q \neq i$, $q \in \{1,\ldots,n\}$. By applying conditions (52) for a single sensor fault $f_i = f_{0i} + \tilde{f}_i$, we define the minimum offset magnitude with a known bound $\bar{f}_i^*$ that ensures that f_i is strongly detectable as:

$$f_{0i}^{\min} = \min\{f_i : (|\lambda_j^{-1}V_j^{-1}A_i| - |\lambda_j^{-1}V_j^{-1}|\Gamma)|f_i| > c\} \tag{53}$$

$$\begin{aligned} c =& \left|\lambda_j^{-1}V_j^{-1}\right|\left(2\bar{\eta} + \Gamma\left(2\left\|\bar{d}\right\| + \bar{f}_i^*\right)\right) + \left(\left|V_j^{-1} - \lambda_j^{-1}V_j^{-1}A\right|\right. \\ &\left. + |V_j^{-1}|\right)\left(2\bar{d} + [0,\ldots,\bar{f}_i^*,\ldots,0]^\top\right) \end{aligned} \tag{54}$$

where A_i denotes the i-th column of A.

Corollary 3.4: Every single sensor fault $f_i = f_{0i} + \tilde{f}_i$ with $|f_{0i}| \geq f_{0i}^{\min}$ and $|\tilde{f}_i| \leq \bar{f}_i^*$ is strongly detectable, if $|V_j^{-1}A_i| > |V_j^{-1}|\Gamma$.

IV. OPTIMIZATION APPROACH

The goal of the developed optimization approach is to select Λ and V for maximizing the set of faults f_i, i.e. $\mathscr{F}_i = \left\{f_i = f_{0i} + \tilde{f}_i : |f_{0i}| \geq f_{0i}^{\min}, |\tilde{f}_i| \leq \bar{f}_i^*\right\}$ that satisfy Corollary 3.4, for all $i \in \{1,\ldots,n\}$. Nonetheless, there is a trade-off in selecting a common Λ and V with good performance for all possible faults, as distinct sensors measure different quantities (e.g., speed, proximity) so that fault offsets may considerably vary. Hence, we develop a two-phase approach:

- **Phase 1:** For each sensor i, find a different Λ and V that yields the minimum fault offset, ϕ_i^{min}, with strong detectability guarantees.
- **Phase 2:** Find a common Λ and V that minimizes a multiplicative factor β, such that single faults with offset $f_{0i}^{\min} = \beta\phi_i^{min}$ are strongly detectable for all $i = 1,\ldots,n$.

A. Phase 1

During this phase, one optimization problem is solved for each sensor to minimize the fault-offsets that guarantee strong detectability. Specifically, the optimization problem for sensor i aims to select Λ and V in order to minimize ϕ when considering a fault offset $F_0^i(\phi) = [\underbrace{0,\cdots,0}_{i-1}, \phi, \underbrace{0,\cdots,0}_{n-i}]^\top$, with known $\bar{f}^* = [\underbrace{0,\cdots,0}_{i-1}, \bar{f}_i^*, \underbrace{0,\cdots,0}_{n-i}]^\top$, such that (52) is satisfied along the direction of at least one eigenvector with $\bar{f} = \bar{f}^*$. For a specific sensor i and eigenvector direction j and constant value ϕ, the arising optimization problem is the following:

$$\begin{aligned} \min_{\lambda_j, W_j}\ & |\lambda_j^{-1}||W_j|\delta_1(\phi) + (|W_j - \lambda_j^{-1}W_jA| + |W_j|)\left(2\bar{d} + \bar{f}^*\right) \\ & - |\lambda_j^{-1}W_jAF_0^i(\phi)| \end{aligned} \tag{55a}$$

$$\text{s.t. } |W_jA_i| > |W_j|\Gamma, \tag{55b}$$

$$\lambda_j^l \leq \lambda_j \leq \lambda_j^u < 0, \quad W_j^l \preceq W_j \preceq W_j^u, \tag{55c}$$

where $\delta_1(\phi) = \left(2\bar{\eta} + \Gamma\left(2\left\|\bar{d}\right\| + \left\|F_0^i(\phi)\right\| + \left\|\bar{f}^*\right\|\right)\right)$ and $W_j = V_j^{-1}$, while (55b) and (55c) denote the condition from Corollary 3.4 and the bounds on variables λ_j and W_j, respectively. Optimization problem (55) is non-convex and nonlinear mainly due to the presence of terms $\lambda_j^{-1}W_j$. Because λ_j is one-dimensional, we consider a simple one-dimensional nonlinear optimization method that solves the problem for a predetermined number of equally spaced points in the interval $[\lambda_j^l, \lambda_j^u]$ and selects the best value [15].

For a specific value of λ_j, say $\bar{\lambda}$, the problem remains non-convex due to the presence of the term $-\left|(1/\bar{\lambda})W_jAF_0^i(\phi)\right| = -\left|(\phi/\bar{\lambda})W_jA_i\right|$ in the objective function and the presence of $|W_jA_i|$ in (55b). Nonetheless, the particular issue can be addressed by solving two optimization problems one for $W_jA_i \geq 0$ and one for $W_jA_i \leq 0$ and then selecting the best value. For the former case, problem (55) becomes:

$$\min_{W_j, Z_j, \Xi_j} Z_j\delta_2(\phi) + \Xi_j\left(2\bar{d} + \bar{f}^*\right) - \left|\phi/\bar{\lambda}\right|W_jA_i \tag{56a}$$

$$\text{s.t. } W_jA_i \geq Z_j\Gamma, \qquad W_j^l \preceq W_j \preceq W_j^u, \tag{56b}$$

$$Z_j \succeq W_j,\ Z_j \succeq -W_j, \tag{56c}$$

$$\Xi_j \succeq W_j - (1/\bar{\lambda})W_jA,\ \Xi_j \succeq -W_j + (1/\bar{\lambda})W_jA, \tag{56d}$$

where $\delta_2(\phi) = (1/\left|\bar{\lambda}\right|)\delta_1(\phi) + 2\bar{d} + \bar{f}^*$. Auxiliary variables Ξ_j and Z_j along with constraints (56c)-(56d) are introduced to ensure that $\Xi_j = |W_j - \frac{1}{\bar{\lambda}}W_jA|$ and $Z_j = |W_j|$. Problem (56) is a Linear Program (LP) that can be solved using standard optimization solvers. For the case $W_jA_i \leq 0$, problem (55) becomes:

$$\min_{W_j, Z_j, \Xi_j} Z_j\delta_2(\phi) + \Xi_j(2\bar{d} + \bar{f}^*) + |\phi/\bar{\lambda}|W_jA_i \tag{57a}$$

$$\text{s.t. } -W_jA_i \geq Z_j\Gamma, \qquad W_j^l \preceq W_j \preceq W_j^u, \tag{57b}$$

$$\text{Constraints (56c) - (56d).} \tag{57c}$$

Algorithm 1 : Solution approach for problem (55)

1: **Input**: A, $\bar{d}$, $\bar{f}^*$, δ_1, λ_j^l, λ_j^u, W_j^l, W_j^u, λ_s, $F_0^i(\phi)$;
2: **Output**:λ_j^*, W_j^*, D_j^*, $j = 1,...,n$;
3: **for** ($\bar{\lambda} = \lambda_j^l : \lambda_s : \lambda_j^u$) **do**
4: Solve problem (56) to obtain $\hat{\lambda}_j$ and $\hat{W}_j$;
5: Using $\hat{\lambda}_j$ and $\hat{W}_j$ compute D_j from (51);
6: **if** ($D_j \geq D_j^*$) **then**
7: Set $\lambda_j^* = \hat{\lambda}_j$, $W_j^* = \hat{W}_j$ and $D_j^* = D_j$;
8: Execute lines 4-7 for problem (57);

Algorithm 2 : Minimum fault offset determination

1: **Input**: A, $\bar{d}$, $\bar{f}^*$, δ_1, λ_j^l, λ_j^u, W_j^l, W_j^u, ϕ^l, ϕ^u, λ_s;
2: **Output**: Λ, V, ϕ_i^{min};
3: **repeat**
4: $\phi = (\phi^l + \phi^u)/2$ and compute $F_0^i(\phi)$;
5: **for** $(j = 1,...,n)$ **do**
6: Run Alg. 1 to obtain λ_j^*, W_j^*, and D_j^* from (51);
7: **if** any$(D_j^* \geq 0)$ **then**
8: $\phi_i^{min} = \phi^u = \phi$, $\Lambda = \mathrm{diag}([\lambda_1^*,...,\lambda_n^*])$, $V = W^{-1}$;
9: **else**
10: $\phi^l = \phi$;
11: **until** $(\phi^u - \phi^l \leq \varepsilon)$

Alg. 1 summarizes the procedure to solve problem (55) for constant ϕ. To determine the minimum fault offset ϕ_i^{min} that guarantees strong detectability, a bisection approach is followed, outlined in Alg. 2. The algorithm bisects the fault offset parameter ϕ and updates the upper bound on ϕ if the optimal pair Λ, V ensures strong detectability; otherwise the lower bound is updated. The procedure is repeated until the convergence of the lower and upper ϕ bounds.

B. Phase 2

Upon obtaining the minimum fault offsets, ϕ_i^{min}, that guarantee strong detectability for each sensor i, Phase 2 finds a common design, Λ, V, that minimizes β such that strong detectability is guaranteed for all sensors faults with offset $\beta F_0^i(\phi_i^{min})$. To achieve this, we need to find a common design Λ, V for which each of the n single faults are strongly detectable along the direction of at least one eigenvector. Given a particular value of β, our approach to solve this problem is to run Alg. 1 for each sensor to determine the eigenvector dimensions for which the particular fault is strongly detectable. In this way, a matrix C is formed, where $C_{ij} = 0$ indicates that sensor fault i is strongly detectable from eigenvector j, otherwise $C_{ij} = 1$. Then, to obtain the minimum number of strongly non-detectable faults, the following *Linear Sum Assignment Problem* is solved using low complexity algorithms [16].

$$\min_{\Pi} \; C_T(\Pi) = \mathrm{trace}(C^\top \Pi) \tag{58a}$$

$$\text{s.t. } \Pi \mathbf{1} = \mathbf{1}, \quad \Pi^\top \mathbf{1} = \mathbf{1}, \quad \Pi \in \{0,1\}^{n \times n}. \tag{58b}$$

The solution of (58) provides the best assignment of eigenvectors/eigenvalues to sensor faults, Π^*, while the total cost $C_T(\Pi^*)$ indicates the number of strongly non-detectable faults. Hence, all faults are strongly detectable when $C_T(\Pi^*) = 0$. Note that $\Pi_{ij}^* = 1$ implies that the j-th eigenvector obtained from having a single fault at sensor i should be used for the detection of the particular fault in the common design. Then, a bisection procedure is followed, similar to Alg. 2, to determine the minimum value of β that guarantees strong detectability for all faults.

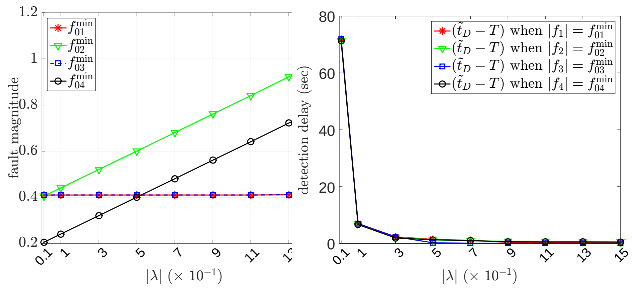
V. SIMULATION EXAMPLE

In this section we illustrate the application of the proposed methodology to the simple example of the flexible link robot presented in [17]. We selected the same parameter values as in [17] with $\gamma(x,u) = [0, 21.6u, 0, -3.33\sin(x_3)]^\top$, while we considered system disturbances and measurement noise such that $\eta = [\eta_1, \eta_2, 0, 0]^\top$ with $\bar{\eta}_i = 10^{-1}$, $i \in \{1,2\}$, $\bar{d}_j = 10^{-1}$, $j \in \{1,2,3,4\}$. The nonlinear observer is structured as shown in (4) with $\gamma(y,u) = [0, 21.6u, 0, -3.33\sin(y_3)]^\top$. Assuming healthy conditions, the Lipschitz condition in this example can be described using (22); $|\gamma(x,u) - \gamma(y_H,u)| \preceq \Gamma \bar{d}_3$, $\Gamma = \begin{bmatrix} 0 & 0 & 0 & 3.33 \end{bmatrix}^\top$. Algorithms 1-3 were used to obtain Λ and V that, for all $i \in \{1,\ldots,4\}$, maximize the set of faults $\mathscr{F}_i = \left\{ f_i = f_{0i} + \tilde{f}_i : |f_{0i}| \geq f_{0i}^{\min}, |\tilde{f}_i| \leq 0 \right\}$ satisfying Corollary 3.4.

Figure 1a illustrates the minimum offset magnitude $f_{0j}^{\min}$, $j \in \{1,2,3,4\}$ of sensor faults that are guaranteed to be strongly detectable by an observer-based FD scheme designed using $\Lambda = \mathrm{diag}([\lambda,\lambda,\lambda,\lambda])$ and V resulted from the optimization process in Section IV. As observed, the minimum offset magnitude of sensor faults f_2 and f_4 that are strongly detectable reduces significantly by reducing the magnitude of the eigenvalue. In the case of f_1 and f_3, this reduction is very small. Using the information presented in Fig. 1a, we performed 72 simulations with the following characteristics: For every value of $|\lambda|$ and the corresponding V, we applied the observer-based FD scheme for detecting a bias sensor fault f_j (i.e. $\tilde{f}_j = \bar{f}_j = 0$) for all $j \in \{1,2,3,4\}$ that occurs at $T = 300$ sec, carrying out 2 simulations; one with $f_j = f_{0j}^{\min}$ and one with $f_j = -f_{0j}^{\min}$. In both cases the time of detection t_D, which is defined in (15) and is equivalent to $t_D = \min_j t_{D_j}$ with $t_{D_j} = \min \left\{ t : \left| V_j^{-1} \varepsilon_y(t) \right| > \bar{\varepsilon}_{y_j}(t) \right\}$, $j \in \{1,2,3,4\}$ is almost the same (maximum deviation 0.6 sec). We computed the average detection time as $\tilde{t}_d = 0.5 * \left(t_d^{(-f_j)} + t_d^{(+f_j)} \right)$. Figure 1b shows the delay of detection, defined as $\tilde{t}_d - T$, when the observer-based scheme using a specific value of $|\lambda|$ is applied. Based on Fig. 1b, we can infer that the delay of detection reduces significantly by increasing the magnitude of the eigenvalue λ. This reduction can be justified by the fact that the delay of detection is associated with the transient response of the residual and the adaptive set $\mathscr{P}_H(t)$ defined in (13) (i.e. associated with the weak fault detectability conditions), which becomes faster with larger values of $|\lambda|$.

Figure 2 shows the simulation results of the application of the observer-based FD scheme with $\lambda = -0.7$ and $V = [\;6.8803 \;\; 0 \;\; 0 \;\; 0; \; -3.0962 \;\; 0.1 \;\; 0 \;\; 0; \; 0 \;\; 0 \;\; 5.2675 \;\; 0;$

$64.692 \quad -9.4024 \quad 4.7408 \quad 9.4024\ ]$, to the system affected by the single sensor fault $f_2 = f_{02}^{\min} = 0.6804$ at $T = 300$ sec. The fault detection criterion (14) is equivalent to check if *for at least one* j there is a time instant at which $\left|V_j^{-1}\varepsilon_y(t)\right| > \bar{\varepsilon}_{y_j}(t)$, where $V_j^{-1}\varepsilon_y(t)$ and $\bar{\varepsilon}_{y_j}(t)$ are respectively depicted with red and blue lines. Under faulty conditions, the residual $\varepsilon_y(t)$ will asymptotically lie in $\mathscr{P}(\infty)$ defined in (50), implying that $\left|V_j^{-1}\varepsilon_y(t)\right| \in [\underline{\upsilon}_j, \overline{\upsilon}_j] = \lambda_j^{-1}V_j^{-1}Af + [-\chi_{Fj} - \upsilon_{Hj}, \chi_{Fj} + \upsilon_{Hj}]$ for all $j \in \{1,2,3,4\}$. As observed, the sensor fault $f_2 = f_{02}^{\min} = 0.6804$ is strongly detectable, since the limit sets $\mathscr{P}(\infty)$ and $\mathscr{P}_H(\infty)$ are separated along the direction of the first eigenvector, with the minimum positive distance (the distance equals $1.038\times\ 10^{-4}$). This leads to the persistence excitation of the residual along the direction of the first eigenvector after the occurrence of the sensor fault.



(a) (b)

Fig. 1: Offset magnitude of strongly detectable sensor faults and average delay of detection versus the eigenvalue λ

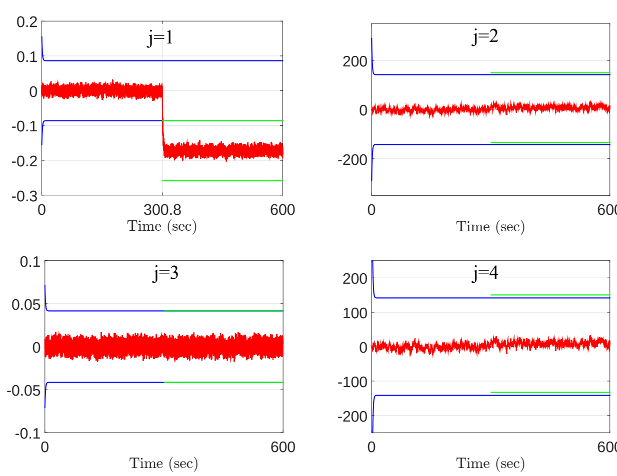


Fig. 2: Eigenstructure 1: $V_j^{-1}\varepsilon_{y_j}$ (red line), $[-\bar{\varepsilon}_{y_j}(t), \bar{\varepsilon}_{y_j}(t)]$ (blue lines), $[\underline{\upsilon}_j, \overline{\upsilon}_j]$ (green lines) for $j \in \{1,2,3,4\}$ in the case that $f_2 = f_2^{\min} = 0.6804$ at $T = 300$ sec.

VI. CONCLUSIONS

In this paper, we presented a methodology for optimizing the design of an observer-based sensor fault detection (FD) scheme for nonlinear systems. Under the assumption of bounded system disturbance and measurement noise, it is proved that the observer-based residual vector under healthy and faulty conditions lies within adaptive convex sets, which asymptotically converge to limit sets. These limit sets are generated based on robust positively invariant sets that enclose the state estimation error dynamics under both healthy and faulty conditions. The goal of the optimization methodology was to find the eigenstructure of the observer that maximizes the set of faults that are guaranteed to be strongly detectable. The strong detectability conditions were obtained based on the separation of the limit sets under healthy and faulty conditions. A simple example of a flexible link robot was used to show the application of the proposed optimization-based methodology. Future work will involve the optimization of the observer sensor FD scheme using partial measurements [6], [18] and assuming multiple sensor faults, as well as the comparison with alternative techniques on fault detectability.

REFERENCES

[1] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2010.

[2] X. Zhang, M. M. Polycarpou, and T. Parisini, "Design and analysis of a fault isolation scheme for a class of uncertain nonlinear systems," *Annual Reviews in Control*, vol. 32, no. 1, pp. 107–121, 2008.

[3] X.-G. Yan, S. K. Spurgeon, and C. Edwards, "Sliding mode observer-based fault detection and isolation," in *Variable Structure Control of Complex Systems*. Springer, 2017, pp. 263–296.

[4] T. Raïssi, G. Videau, and A. Zolghadri, "Interval observer design for consistency checks of nonlinear continuous-time systems," *Automatica*, vol. 46, no. 3, pp. 518–527, 2010.

[5] S. Montes de Oca, V. Puig, and J. Blesa, "Robust fault detection based on adaptive threshold generation using interval lpv observers," *International Journal of Adaptive Control and Signal Processing*, vol. 26, no. 3, pp. 258–283, 2012.

[6] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Sensor fault diagnosis," *Foundations and Trends in Systems and Control*, vol. 3, no. 1-2, pp. 1–248, 2016.

[7] J. Chen and R. J. Patton, *Robust Model-based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers, 1999.

[8] S. X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer-Verlag London, 2013.

[9] A. Varga, *Solving Fault Diagnosis Problems: Linear Synthesis Techniques*. Springer, 2017, vol. 84.

[10] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and fault-tolerant control*. Springer-Verlag Berlin Heidelberg, 2016.

[11] M. M. Polycarpou and A. B. Trunov, "Learning approach to nonlinear fault diagnosis: detectability analysis," *IEEE Transactions on Automatic Control*, vol. 45, no. 4, pp. 806–812, 2000.

[12] V. Reppa, S. Olaru, and M. M. Polycarpou, "Structural detectability analysis of a distributed sensor fault diagnosis scheme for a class of nonlinear systems," in *the 9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, 2015, pp. 1485–1490.

[13] E. Kofman, H. Haimovich, and M. M. Seron, "A systematic method to obtain ultimate bounds for perturbed systems," *International Journal of Control*, vol. 80, no. 2, pp. 167–178, 2007.

[14] S. Olaru, J. A. De Doná, M. M. Seron, and F. Stoican, "Positive invariant sets for fault tolerant multisensor control schemes," *International Journal of Control*, vol. 83, pp. 2622–2640, 2010.

[15] S. S. Rao, *Engineering optimization: theory and practice*. John Wiley & Sons, 2009.

[16] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment problems*. SIAM, 2012.

[17] R. Rajamani and Y. Cho, "Existence and design of observers for nonlinear systems: relation to distance to unobservability," *International Journal of Control*, vol. 69, no. 5, pp. 717–731, 1998.

[18] K. Vijayaraghavan, R. Rajamani, and J. Bokor, "Quantitative fault estimation for a class of non-linear systems," *International Journal of Control*, vol. 80, no. 1, pp. 64–74, 2007.