

Foundations and Trends® in Systems and Control
Vol. XX, No. XX (2016) 1–247
© 2016 V. Reppa, M. M. Polycarpou and
C. G. Panayiotou
DOI: 10.1561/XXXXXXXXXX



Sensor Fault Diagnosis

Vasso Reppa
Laboratory of Signals and Systems (L2S, UMR CNRS 8506)
CentraleSupélec-CNRS-University Paris-Sud
University Paris-Saclay
reppavasso@gmail.com

Marios M. Polycarpou
KIOS Research Center for Intelligent Systems and Networks
Electrical & Computer Engineering Department
University of Cyprus
mpolycar@ucy.ac.cy

Christos G. Panayiotou
KIOS Research Center for Intelligent Systems and Networks
Electrical & Computer Engineering Department
University of Cyprus
christosp@ucy.ac.cy

Contents

1	Introduction to Sensor Fault Diagnosis	2
2	Modeling for Sensor Fault Diagnosis	12
2.1	Cyber-physical Modeling Framework	12
2.2	Fault Modeling	20
2.3	Summary and Discussion	22
3	Sensor Fault Diagnosis Architecture	24
3.1	Centralized Architecture	24
3.2	Decentralized Architecture	38
3.3	Distributed Architecture	43
3.4	Summary and Discussion	45
4	Sensor Fault Detection	47
4.1	Residual Generation	48
4.2	Computation of Adaptive Thresholds	60
4.3	Sensor Fault Detection Decision Logic	66
4.4	Other Established Fault Detection Schemes	70
4.5	Illustrative Example	86
4.6	Summary and Discussion	106
5	Sensor Fault Isolation	111

5.1	Decentralized Multiple Sensor Fault Isolation	112
5.2	Distributed Multiple Sensor Fault Isolation	118
5.3	Other Established Fault Isolation schemes	123
5.4	Illustrative Example	135
5.5	Summary and Discussion	142
6	Performance Analysis	145
6.1	Sensor Fault Detectability	147
6.2	Multiple Sensor Fault Isolability	167
6.3	Illustrative Example	168
6.4	Summary and Discussion	183
7	Learning Approaches for Sensor Fault Diagnosis	184
7.1	Parametric Model	186
7.2	Adaptive Nonlinear Estimation Scheme	187
7.3	Stability Analysis	188
7.4	Distributed Sensor Fault Detection using Nonlinear Uncertainty Approximators	192
7.5	Illustrative Example	198
7.6	Summary and Discussion	218
8	Concluding Remarks and Open Issues	224
	Acknowledgements	231
	Appendices	232
A	Illustrative Example: Design Parameters	233
	References	237

Abstract

This tutorial investigates the problem of the occurrence of multiple faults in the sensors used to monitor and control a network of cyber-physical systems. The goal is to formulate a general methodology, which will be used for designing sensor fault diagnosis schemes with emphasis on the isolation of multiple sensor faults, and for analyzing the performance of these schemes with respect to the design parameters and system characteristics. The backbone of the proposed methodology is the design of several monitoring and aggregation cyber agents (modules) with specific properties and tasks. The monitoring agents check the healthy operation of sets of sensors and infer the occurrence of faults in these sensor sets based on structured robustness and sensitivity properties. These properties are obtained by deriving analytical redundancy relations of observer-based residuals sensitive to specific subsets of sensor faults, and adaptive thresholds that bound the residuals under healthy conditions, assuming bounded modeling uncertainty and measurement noise. The aggregation agents are employed to collect and process the decisions of the agents, while they apply diagnostic reasoning to isolate combinations of sensor faults that have possibly occurred. The design and performance analysis methodology is presented in the context of three different architectures: for cyber-physical systems that consist of a set of interconnected systems, a distributed architecture and a decentralized architecture, and for cyber-physical systems that are treated as monolithic, a centralized architecture. For all three architectures, the decomposition of the sensor set into subsets of sensors plays a key role in their ability to isolate multiple sensor faults. A discussion of the challenges and benefits of the three architecture is provided, based on the system scale, the type of system nonlinearities, the number of sensors and the communication needs. Lastly, this tutorial concludes with a discussion of open problems in fault diagnosis.

V. Reppa, M. M. Polycarpou and C. G. Panayiotou. *Sensor Fault Diagnosis*. Foundations and Trends[®] in Systems and Control, vol. XX, no. XX, pp. 1–247, 2016.

DOI: 10.1561/XXXXXXXXXX.

1

Introduction to Sensor Fault Diagnosis

Recent advances in information and communication technologies, embedded systems and sensor networks have generated significant research activity in the development of so-called cyber-physical systems (CPS). CPS consist of two components: (i) physical, biological or engineered systems and (ii) a cyber core, comprised of communication networks and computational availability that monitors, coordinates and controls the physical part [Antsaklis et al. [2013]]. In this tutorial, we will consider a network of interconnected CPS, where each subsystem may be characterized by simple dynamics, but the overall dynamics can be large-scale and complex. The focus of research on CPS is to improve the collaborative link between physical and computational (cyber) elements for increased adaptability, efficiency and autonomy. The key motivation for the advancement of CPS is the need to better coordinate the interactions between the software and hardware designs by facilitating self-awareness in evolving environments and the handling of a huge amount of data of different time and space characteristics. However, reaching such a level of system intelligence necessitates the development of mechanisms capable to assess the reliability of information acquired by distributed deployed sensors and sensor networks

through wired and wireless links [Ding et al. [2006]], or by internet-of-things devices.

A representative example of a large network of cyber-physical systems is a smart city with intelligent infrastructures for supporting the environment, energy and water distribution, transportation, telecommunication, health care, home automation and many more [Chourabi et al. [2012]]. Each of these critical infrastructures consists of a large number of distributed, interconnected subsystems, which need to be monitored and controlled using a large number of sensing/actuation devices and feedback control algorithms.

Although the benefits of the use of automated monitoring and control procedures are widely accepted, this use has made critical infrastructures more susceptible to faults [Kröger and Zio [2011]]. Thus, supervision schemes capable of diagnosing and accommodating faults are applied for ensuring system reliability and safety. From a systems point of view, safety, reliability and fault tolerance become key challenges in designing cyber-physical systems. For meeting these challenges, the cyber core should be empowered with supervision capabilities for diagnosing faults in the physical part and compensating their effects by taking appropriate remedial actions [Blanke et al. [2016], Isermann [2006]].

Fault detection addresses the problem of determining the presence of faults in a system and estimating their instant of occurrence [Isermann [2006], Gertler [1998], Chen and Patton [1999], Blanke et al. [2016], Ding [2008]]. Fault detection is followed by *fault isolation*, which deals with finding which ones are the faulty components in the system, or the type of fault. *Fault identification* is described as the procedure of determining the size and the time variant behavior of the fault. In some cases, during the fault identification procedure, we also seek to assess the extend of the fault and the risks associated with it [Chen and Patton [1999], Ding [2008]]. The result of fault identification is essential for performing *fault accommodation* by either changing the control law or using virtual sensors or actuators in response to a fault, without switching off any system component [Blanke et al. [2016]]. In this tutorial, we will consider mainly the fault detection and fault isolation problems, which, for simplicity, together we will refer to as fault

diagnosis.

Various methodologies have been developed for the fault diagnosis problem in general [Isermann [2006]], but the detection and isolation of sensor faults has become a key challenging problem in the last few years. This is due to the large number of sensors and sensor networks, used for (i) monitoring and controlling large-scale CPS; (ii) providing rich and redundant information for executing safety-critical tasks; and (iii) offering information to citizens and governmental agencies for resolving problems promptly in emergency situations. For instance, in intelligent transportation, vehicles may be equipped with odometers, lasers, frontal camera video-sensors, GPS, speed or object tracking sensors, in order to be able to acquire and broadcast information relevant to performing tasks such as cooperative or fully autonomous driving, avoiding lane departure and collision, etc. In smart buildings, multiple sensors are installed in different zones (measuring quantities such as temperature, humidity, CO₂, contaminant concentration, occupancy), as well as in heating, ventilation and air-conditioning systems for measuring supply/return/mixed air temperature, supply/return air differential pressure, return air humidity, etc. Such sensing information may be used for reducing the energy consumption of a building and maintaining the desired living conditions, as well as for executing evacuation plans in safety-critical situations (e.g. fire). Undetected sensor faults can severely impact automation and supervision schemes [Sherry and Mauro [2014]], possibly leading to system instability, loss of information fidelity, incorrect decisions and disorientation of remedial actions [BEA [2012]].

Sensor fault detection and isolation (FDI) methods are classified into physical redundancy-based and model-based methods [Betta and Pietrosanto [2000]]. In many applications, the physical redundancy approach is not used due to the high cost of installation and maintenance, as well as due to space restrictions. However, the evolution of microtechnology in recent years has contributed to the reduction of the size and fabrication cost of sensors, making physical-redundancy methods more cost effective. Current technological advances are geared towards the use of multiple, possibly heterogeneous sensors, which are not neces-

sarily co-located, however the measured variables may have redundant information, which is useful for fault diagnosis purposes. For example, in a smart building, there may exist two sensors measuring the temperature in adjacent rooms; in such a case, the relation (either known apriori or learned during operation) between the two measured quantities maybe used to determine if one of the two sensors is faulty [Alippi et al. [2013]]. With the current trend towards utilizing larger and larger numbers of sensors, there is also a higher probability of multiple sensor faults occurring, which is an issue that has not been well studied in the fault diagnosis literature. The majority of sensor FDI techniques rely on the utilization of models only [Isermann [2006]]. These techniques are further categorized as quantitative or qualitative methods [Venkatasubramanian et al. [2003a,b]]; the first category relies on a nominal mathematical model describing the system, while the second one uses symbolic and qualitative system representations. Additionally, some research work has been done on the combination of the two sensor FDI approaches [Issury and Henry [2009]].

Qualitative model-based techniques are typically used by the artificial intelligence diagnostic (DX) community [Travé-Massuyès [2012]]. The design of these techniques is based on the utilization of either causal models, such as signed digraphs, bond graphs, fault trees, etc. [Vedam and Venkatasubramanian [1997], Bregon et al. [2012]], or functional or structural abstraction hierarchies [Daigle et al. [2012], Blanke et al. [2016], Monteriu et al. [2007]]. The nature of these models facilitates especially the fault isolation procedure. Moreover, the qualitative approach treats fault detection and isolation as a unified problem, and exploits reasoning techniques, thereby providing by design more straightforward methods for multiple sensor fault isolation [Nyberg [2006], De Kleer and Williams [1987], Daigle et al. [2012], Frisk et al. [2012]]. The equivalence and the differences between methods developed by both communities, as well as the design of a unified framework taking advantage of the benefits of each approach have been studied by several researchers [Cordier et al. [2004], Pulido and González [2004], Gentil et al. [2004]].

While qualitative model-based approaches have mostly been

adopted by the DX community, quantitative model-based approaches such as parity equations and observers are widely used for sensor FDI by the control-oriented FDI community [Gertler [1998], Chen and Patton [1999]]. Among the quantitative methods, observer-based approaches have been applied to nonlinear systems, using a single nonlinear observer [Rajamani and Ganguli [2004], Narasimhan et al. [2008], Yan and Edwards [2007], Talebi et al. [2009]], or a bank of observers [Mattone and De Luca [2006], Rajaraman et al. [2006], Samy et al. [2011], Reppa et al. [2014b, 2012]]. Several researchers have developed sensor FDI methods, which treat sensor faults as actuator faults and apply observer-based approaches for nonlinear systems [Kabore and Wang [2001], De Persis and Isidori [2001]], as well as methodologies for tackling the problem of actuator and sensor faults in a unified framework [Du et al. [2013]]. One of the common characteristics of the majority of observer-based methods is the use of the open-loop system model and the input and output data related to the open-loop system. Recently, observer-based sensor FDI techniques have been proposed, that take advantage of the information about the closed-loop operation of the system (i.e. reference signals and controller's structure), when this is available [Olaru et al. [2010], Seron et al. [2012, 2013]]. The control-oriented FDI community focuses mostly on making methods robust against modeling uncertainties and views the fault detection and fault isolation as two different tasks.

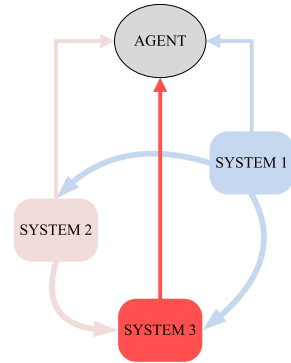
Clearly, mathematical models never capture the real behavior of the modeled system, due to the presence of uncertainties including parametric uncertainty, unmodeled high-order system dynamics, or faults occurring in the system, which can be function of the system state and input. A powerful approach to robust FDI for nonlinear uncertain systems is based on the use of learning techniques for reducing modeling uncertainty [Polycarpou and Helmicki [1995], Trunov and Polycarpou [2000]]. The main concept behind the learning approach for FDI is the approximation of the unknown system behavior using adaptive approximation models (e.g. sigmoidal neural networks, radial basis functions, support vector machines) and nonlinear estimation schemes [Trunov and Polycarpou [2000], Caccavale et al. [2008], Talebi et al.

[2009], Thumati and Jagannathan [2010]]. Under healthy conditions, adaptive approximation based schemes can be used to learn the modeling uncertainty during the initial stage of nonlinear system operation (training period) [Caccavale et al. [2009]]. Then, the nonlinear functional approximator of modeling uncertainty is used for optimizing the adaptive thresholds, thus enhancing the fault detectability and isolability of a quantitative model-based scheme [Reppa et al. [2014b]]. Another approach for reducing the modeling error is offline identification of uncertainties. In general, adaptive approximation schemes provide a flexible methodology for learning the uncertainties in the sense that the training time can be adjusted online based on some criterion involving the estimation error. Under faulty conditions, adaptive approximation based schemes can be applied for learning the faults for isolation and identification purposes initially [Zhang et al. [2005, 2008]], and then for accommodating the fault effects [Zhang et al. [2004], Reppa et al. [2014a]].

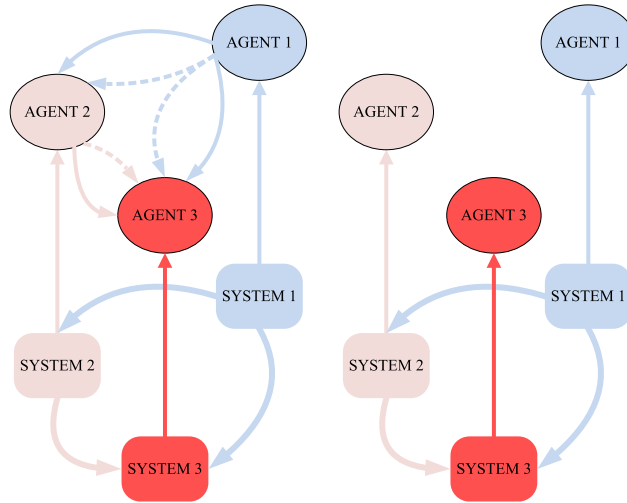
The majority of model-based sensor FDI methods are deployed in a centralized framework (see Fig 1.1-1), but these approaches are less suitable for large-scale and complex systems such as a network of interconnected CPS. In this context, centralized approaches have the following disadvantages: (i) increased computational complexity of the FDI algorithms, since centralized architectures are tailored to handle (multiple) faults globally, (ii) increased communication requirements due to the transmission of information to a central point, (iii) vulnerability to security threats, because the central cyber core in which the sensor FDI algorithm resides is a single-point of failure, and (iv) reduced scalability in case of system expansion, due to the utilization of a global physical model or black-box. A common design characteristic of non-centralized methods is that they handle the large-scale and complex system as a set of interconnected subsystems and they employ local agents that perform diagnosis based on local subsystems' models. The local agents are commonly deployed in either a distributed (Fig 1.1-2) or decentralized (Fig 1.1-3) architecture. The classification of these architectures is based on the type of system interconnections, the cyber levels of diagnosis, the task of the local diagnosers, as well as the

type of communication and information exchanged between the local and high-level diagnosers.

In [Yan and Edwards [2008], Zhang and Zhang [2013], Zhang et al. [2014], Klinkhio et al. [2008], Ferrari et al. [2012], Boem et al. [2013a], Ferdowsi et al. [2012], Indra et al. [2012], Reppa et al. [2015b,c]], decentralized and distributed FDI methods are developed for physically interconnected subsystems. Distributed architectures have also been designed for systems with interconnections in the control law [Shames et al. [2011]], interconnected inputs [Daigle et al. [2007]] or sensing interconnections (i.e. relative output measurements) [Davoodi et al. [2014]]. For enhancing fault isolation, multiple levels of diagnosis in a hierarchical architecture have been designed. In particular, while the single level diagnosis is realized by the local diagnosers [Yan and Edwards [2008], Zhang and Zhang [2013], Klinkhio et al. [2008], Shames et al. [2011], Davoodi et al. [2014], Daigle et al. [2007]], additional FDI units are developed in a hierarchical architecture, aggregating and processing the outputs of the local diagnosers [Ferrari et al. [2012], Boem et al. [2013a], Ferdowsi et al. [2012], Indra et al. [2012]]. The decentralized or distributed nature of the FDI process is related either to the task executed by the local diagnosers or the communication between the local diagnosers. In decentralized schemes, a local diagnoser is commonly designed to detect and isolate faults only in its underlying system [Yan and Edwards [2008], Klinkhio et al. [2008], Reppa et al. [2015b]], while it may not exchange any information with other local diagnosers [Ferdowsi et al. [2012], Indra et al. [2012]]. On the contrary, in distributed schemes, there is communication between the local diagnosers and every local diagnoser can detect and isolate faults in neighboring systems [Zhang and Zhang [2013], Shames et al. [2011], Davoodi et al. [2014], Daigle et al. [2007], Ferrari et al. [2012], Boem et al. [2013a], Reppa et al. [2015c]]. The design of distributed FDI architectures may also differ in the type of exchanged information. Specifically, the local diagnosers may exchange estimations [Zhang and Zhang [2013], Yan and Edwards [2008]] [Daigle et al. [2007]], or measurements of the interconnected states [Shames et al. [2011], Ferrari et al. [2012], Boem et al. [2013a]], or fault signatures [Daigle et al. [2007]]. In multi-level FDI



(1) Centralized architecture.



(2) Distributed architecture. (3) Decentralized architecture.

Figure 1.1: Typical architectures for interconnected systems. (1) In a centralized approach, input/output information of all systems is transmitted to one agent. (2) In a distributed architecture, input/output information of each subsystem is transmitted to its dedicated agent, and the agents are allowed to exchange information (input/output information, decisions). (3) In a decentralized architecture, input/output information of each subsystem is transmitted to its dedicated agent, but the agents do not exchange information.

schemes, the communication between levels is commonly sporadic and event-driven, while the information transmitted to higher levels can be the decisions of the local diagnosers [Ferrari et al. [2012], Boem et al. [2013a], Reppa et al. [2015c]], the time instances of fault detection of the local diagnosers [Ferdowsi et al. [2012]] or the calculated analytical redundancy relations [Indra et al. [2012]].

It is worth to note that a set of interconnected systems can also be viewed as a monolithic system. In this case, non-centralized fault diagnosis techniques can be applied to the monolithic system. The use of multiple processing units (agents) and an aggregation unit that fuses the information from these units, can be found in the well-developed FDI methods for stochastic systems based on interacting multiple models (IMM)[Zhang and Li [1998]], multiple sensor fusion (MSF) for stochastic systems [Salahshoor et al. [2008], Reece et al. [2009]] or hidden Markov models (HMM) [Alippi et al. [2013]]. In IMM-based techniques, the multiple models describe the system in healthy and various faulty system modes and are designed using the a priori knowledge of the possible system faults. Fault diagnosis using MSF-based techniques can be conducted by using local filters that generate local estimates and local decisions, and a global filter that combines the local state estimates to derive an improved global estimate and/or fuse the local decisions for obtaining a global decision. In HMM-based methods, spatial and temporal relationships among sensor datastreams is exploited, and for each pair of sensors a HMM-based module is designed. The lower processing layer detects variations in the relationships between pairs of sensors, while the upper processing (cognitive) level aggregates the information coming from all sensor units to distinguish faults from changes in the environment and false positives.

The main goal of this tutorial is to provide a cyber-physical methodology for designing and analyzing quantitative model-based sensor FDI techniques for large-scale nonlinear systems, which are monitored and controlled by a large number of sensors or sensor networks. To this end, Chapter 2 presents models that describe the system behavior under healthy and faulty conditions, along with the underlying assumptions that are commonly used for the design of sensor FDI techniques and

the formulation of the sensor fault diagnosis problem. Then, Chapter 3 surveys various architectures (centralized, decentralized, distributed) for solving the sensor FDI problem, taking into account the system scale and the number of sensors, as well as the communication needs. Chapter 4 describes the stages for designing observer-based fault detection methods, taking into account the nonlinear system nature, while Chapter 5 details the isolation steps with emphasis on multiple sensor faults. The performance of the observer-based sensor FDI techniques is analyzed in Chapter 6 with respect to robustness against modeling uncertainties, sensor fault detectability and isolability. Chapter 7 presents learning techniques that can be used for enhancing the performance of sensor FDI methods under healthy and faulty conditions.

2

Modeling for Sensor Fault Diagnosis

in this chapter, we will present some quantitative mathematical models that can be used to describe the operation of the physical part of a large-scale cyber-physical systems. By considering a network of interconnected nonlinear subsystems of lower dimension, we will propose some models for describing the subsystem's behavior under healthy and faulty conditions.

2.1 Cyber-physical Modeling Framework

Consider a network of M interconnected CPS. The I -th CPS, $I \in \{1, \dots, M\}$, is described by the pair $(\mathcal{P}^{(I)}, \mathcal{C}^{(I)})$, where $\mathcal{P}^{(I)}$ corresponds to the physical part, while $\mathcal{C}^{(I)}$ denotes the cyber part. In this section, we will initially present several mathematical models for describing the physical part $\mathcal{P}^{(I)}$ of the network of cyber-physical systems, which may be characterized by nonlinear or linear dynamics.

2.1.1 Nonlinear Modeling of Cyber-Physical Systems

The physical part $\mathcal{P}^{(I)}$ is modeled by a nonlinear dynamical subsystem, denoted by $\Sigma^{(I)}$, and a set of sensors, denoted by $\mathcal{S}^{(I)}$, as shown in Fig.

2.1; i.e.,

$$\begin{aligned} \Sigma^{(I)} : \quad \dot{x}^{(I)}(t) = & A^{(I)}x^{(I)}(t) + \gamma^{(I)}(x^{(I)}(t), u^{(I)}(t)) \\ & + h^{(I)}(x^{(I)}(t), u^{(I)}(t), z^{(I)}(t)) \\ & + \eta^{(I)}(x^{(I)}(t), u^{(I)}(t), z^{(I)}(t), t), \end{aligned} \quad (2.1)$$

where $x^{(I)} \in \mathbb{R}^{n_I}$, $u^{(I)} \in \mathbb{R}^{\ell_I}$ are the state and input vector of $\Sigma^{(I)}$, respectively, while $z^{(I)} \in \mathbb{R}^{p_I}$ is the interconnection state vector, containing the states of the neighboring (interconnected) subsystems of $\Sigma^{(I)}$ [Klinkhieo et al. [2008], Ferrari et al. [2012], Boem et al. [2013a]].

The constant matrix $A^{(I)} \in \mathbb{R}^{n_I \times n_I}$ is the linearized part of the state equation and $\gamma^{(I)} : \mathbb{R}^{n_I} \times \mathbb{R}^{\ell_I} \mapsto \mathbb{R}^{n_I}$ represents the known nonlinear dynamics. The term $A^{(I)}x^{(I)} + \gamma^{(I)}(x^{(I)}, u^{(I)})$ represents the known local dynamics, while $h^{(I)} : \mathbb{R}^{n_I} \times \mathbb{R}^{\ell_I} \times \mathbb{R}^{p_I} \mapsto \mathbb{R}^{n_I}$ represents the known interconnection dynamics. The last term $\eta^{(I)} : \mathbb{R}^{n_I} \times \mathbb{R}^{\ell_I} \times \mathbb{R}^{p_I} \times \mathbb{R} \mapsto \mathbb{R}^{n_I}$ denotes the modeling uncertainty associated with $\Sigma^{(I)}$, representing various possible sources of uncertainty such as linearization error, uncertainty in the model's parameters, system disturbances etc. The input vector $u^{(I)}$ is generated by a control agent, denoted by $\mathcal{K}^{(I)}$, which is implemented in the cyber part $\mathcal{C}^{(I)}$ of the I -th CPS, based on some desired reference signals $r^{(I)}(t)$ (see Example 2.1 for an illustration of three interconnected subsystems). It is noted that, based on the assumed model described by (2.1), the system is input-decoupled, meaning that the I -th subsystem is affected only by the I -th control signal $u^{(I)}$, not by other control inputs.

The sensor set of $\mathcal{P}^{(I)}$ is characterized by

$$\mathcal{S}^{(I)} : \quad y^{(I)}(t) = C^{(I)}x^{(I)}(t) + d^{(I)}(t) + f^{(I)}(t), \quad (2.2)$$

where $y^{(I)} \in \mathbb{R}^{m_I}$ is the output vector, $d^{(I)} \in \mathbb{R}^{m_I}$ denotes the noise vector corrupting the measurements of sensors in $\mathcal{S}^{(I)}$ and $f^{(I)} \in \mathbb{R}^{m_I}$ represents the possible sensor fault vector. The j -th sensor $\mathcal{S}^{(I)}\{j\}$, $j \in \{1, \dots, m_I\}$ is described by

$$\mathcal{S}^{(I)}\{j\} : \quad y_j^{(I)}(t) = C_j^{(I)}x^{(I)}(t) + d_j^{(I)}(t) + f_j^{(I)}(t), \quad (2.3)$$

where $f_j^{(I)}$ represents the change in the output $y_j^{(I)}$ due to a single, permanent fault in the j -th sensor of the set $\mathcal{S}^{(I)}$. Note that a *perma-*

nent fault implies that after it occurs, it remains active forever. Further discussion on the fault modeling can be found in Section 2.2.

Example 2.1. An example of three interconnected subsystems $\Sigma^{(1)}$, $\Sigma^{(2)}$, and $\Sigma^{(3)}$ are shown in Fig. 2.1. Their associated local sensors sets $\mathcal{S}^{(1)}$, $\mathcal{S}^{(2)}$, and $\mathcal{S}^{(3)}$ are used by three controllers $\mathcal{K}^{(1)}$, $\mathcal{K}^{(2)}$, and $\mathcal{K}^{(3)}$ to generate the control signals based on the reference signals $r^{(1)}$, $r^{(2)}$, and $r^{(3)}$. The subsystems are affected by external disturbances $\eta^{(I)}(t)$, $I = 1, 2, 3$, while the local sensor sets are affected by noise $d^{(I)}(t)$ and faults $f^{(I)}(t)$. In this example, the interconnection vectors are defined as $z^{(1)} = 0$, $z^{(2)} = H^{(2)}x^{(1)}$ and $z^{(3)} = H^{(3)}x^{(2)}$, where $H^{(2)}$, $H^{(3)}$ are known linear matrices.

In the presence of a single fault or multiple sensor faults, it is important that the monitoring system is able to diagnose the faults as quickly and accurately as possible before they lead to catastrophic failures or propagate to other CPS through the distributed control scheme. The research objectives are to design a methodology for detecting and isolating multiple sensor faults that may occur in one or more CPS and analyze its performance, taking into account the following assumptions:

Assumption 1: For each subsystem $I \in \{1, \dots, M\}$: (i) the state vector $x^{(I)}$ and input vector $u^{(I)}$ generated by a feedback controller, remain bounded before and after the occurrence of multiple sensor faults; i.e., there exist compact regions of stability $\mathcal{U}_I \subset \mathbb{R}^{\ell_I}$, $\mathcal{X}_I \subset \mathbb{R}^{n_I}$ such that $(x^{(I)}(t), u^{(I)}(t)) \in \mathcal{X}_I \times \mathcal{U}_I$, for all $t \geq 0$, leading to the existence of a compact region $\mathcal{Z}_I \subset \mathbb{R}^{p_I}$ within which $z^{(I)}$ resides, and (ii) the corresponding reference signals of the interconnection variables $z^{(I)}(t)$, denoted by $z_r^{(I)}(t)$, are known.

Assumption 2: The nonlinear vector field $\gamma^{(I)}$ is locally Lipschitz in $x^{(I)} \in \mathcal{X}_I$, for all $u^{(I)} \in \mathcal{U}_I$ and $t \geq 0$, while $h^{(I)}$ is locally Lipschitz in $x^{(I)} \in \mathcal{X}_I$ and $z^{(I)} \in \mathcal{Z}_I$, for all $u^{(I)} \in \mathcal{U}_I$ and $t \geq 0$.

Assumption 3: The unknown modeling uncertainty $\eta^{(I)}$, is bounded by a known functional $\bar{\eta}^{(I)}$ for all $x^{(I)} \in \mathcal{X}_I$, $u^{(I)} \in \mathcal{U}_I$, $z^{(I)} \in \mathcal{Z}_I$ and $t \geq 0$; i.e.

$$\left| \eta^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t) \right| \leq \bar{\eta}^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t), \quad (2.4)$$

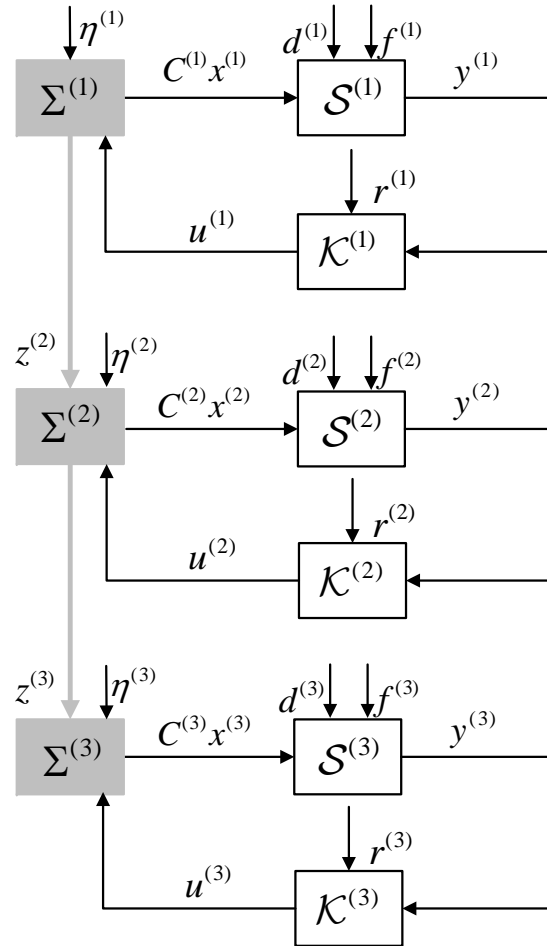


Figure 2.1: An example of interconnected cyber-physical systems.

whereas $\bar{\eta}^{(I)}$ is locally Lipschitz in $x^{(I)} \in \mathcal{X}_I$ and $z^{(I)} \in \mathcal{Z}_I$, for all $u^{(I)} \in \mathcal{U}_I$ and $t \geq 0$.

Assumption 4: The noise corrupting the measurements of sensor $\mathcal{S}^{(I)}\{j\}$ is uniformly bounded:

$$|d_j^{(I)}(t)| \leq \bar{d}_j^{(I)}, \quad j \in \{1, \dots, m_I\}, \quad (2.5)$$

where $\bar{d}_j^{(I)}$ is a known constant bound.

The first part of Assumption 1 is a well-posedness condition, requiring that the feedback controller can retain the boundedness of the state variables in the presence of sensor faults. This assumption is necessary due to the fact that, in this work, we do not address the fault accommodation problem, but only fault detection and isolation issues. The second part of Assumption 1 implies that the design specifications and reference models for the control design of the neighboring subsystems are known. In more detail, it is assumed that the controller $\mathcal{C}^{(J)}$ of a neighboring subsystem of $\Sigma^{(I)}$, $J \neq I$, has been designed to generate a control signal $u^{(J)}$ such that the output $y^{(J)}$ of the subsystem $\Sigma^{(J)}$ tracks a smooth bounded reference trajectory $y_r^{(J)}$, which is derived from a stable local reference model with bounded reference input signal $r^{(J)}$ [Hovakimyan et al. [2005], Chen and Li [2008]]. Based on the reference trajectories $y_r^{(J)}$ or the reference states $x_r^{(J)}$ that generate $y_r^{(J)}$, we can determine the reference signals for the interconnection vector $z^{(I)}$.

Assumption 2 characterizes the class of nonlinear interconnected systems under consideration. Many nonlinearities in practical systems can be considered as Lipschitz, such as power systems and robotic applications [Zhang et al. [2014], Zhu and Han [2002]], in which sinusoidal terms are used in their mathematical models and are globally Lipschitz or car dynamics with locally Lipschitz nonlinearities [Yan and Edwards [2008]].

Assumption 3 provides a bound commonly used for distinguishing between modeling uncertainties and faults. This bound can be obtained either analytically, by explicitly determining the sources of uncertainty and their corresponding bounds, or using off-line identification techniques. The bounding function $\bar{\eta}^{(I)}$ is assumed to be locally Lipschitz,

however this is an easy condition to satisfy from a practical viewpoint since it is straightforward to select a bounding function that satisfies a locally Lipschitz condition. Even if a bounding function is not locally Lipschitz, it can be easily approximated within a region of interest by a Lipschitz function.

Assumption 4 describes a practical representation of the available knowledge for the sensor noise that is typically provided in a given range of operation by sensor manufacturers or introduced when a noise-free analog signal is converted into a digital one with a finite number of digits.

In the sequel, the dependence of the signals on time (e.g. $x^{(t)}$) will be dropped for notation brevity, except for some cases that we want to highlight the time dependence of a signal.

Example 2.2. Consider two interconnected second-order nonlinear systems $\Sigma^{(1)}$ and $\Sigma^{(2)}$; the first subsystem is described by

$$\begin{aligned} \Sigma^{(1)} : \begin{bmatrix} \dot{x}_1^{(1)} \\ \dot{x}_2^{(1)} \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ a_1^{(1)} & a_2^{(1)} \end{bmatrix} \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} + \begin{bmatrix} 0 \\ g_1^{(1)}u^{(1)} + g_2^{(1)}\cos(x_1^{(1)}) \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ h_2^{(1)}(x_1^{(2)}) \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_2^{(1)}(x^{(1)}, u^{(1)}, z^{(1)}) \end{bmatrix}, \end{aligned} \quad (2.6)$$

while the second subsystem is described by

$$\begin{aligned} \Sigma^{(2)} : \begin{bmatrix} \dot{x}_1^{(2)} \\ \dot{x}_2^{(2)} \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ a_1^{(2)} & a_2^{(2)} \end{bmatrix} \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} + \begin{bmatrix} 0 \\ g_1^{(2)}u^{(2)} + g_2^{(2)}\sin(x_1^{(2)}) \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ h_2^{(2)}(x_1^{(1)}) \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_2^{(2)}(x^{(2)}, u^{(2)}, z^{(2)}) \end{bmatrix}, \end{aligned} \quad (2.7)$$

where $a_1^{(I)}, a_2^{(I)}, b^{(I)}, g_1^{(I)}, g_2^{(I)}$ are constants for all $I = 1, 2$. By defining

$x^{(I)} = [x_1^{(I)}, x_2^{(I)}]^\top$, both subsystems satisfy (2.1) with

$$z^{(1)} = x_1^{(2)}, \quad (2.8)$$

$$z^{(2)} = x_1^{(1)}, \quad (2.9)$$

$$A^{(I)} = \begin{bmatrix} 0 & 1 \\ a_1^{(I)} & a_2^{(I)} \end{bmatrix}, \quad (2.10)$$

$$\gamma^{(1)}(x^{(1)}, u^{(1)}) = \begin{bmatrix} 0 \\ g_1^{(1)}u^{(1)} + g_2^{(1)}\cos(x_1^{(1)}) \end{bmatrix} \quad (2.11)$$

$$\gamma^{(2)}(x^{(2)}, u^{(2)}) = \begin{bmatrix} 0 \\ g_1^{(2)}u^{(2)} + g_2^{(2)}\sin(x_1^{(2)}) \end{bmatrix}, \quad (2.12)$$

$$h^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}) = \begin{bmatrix} 0 \\ h_2^{(I)}(z^{(I)}) \end{bmatrix} \quad (2.13)$$

$$\eta^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t) = \begin{bmatrix} 0 \\ \eta_2^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t) \end{bmatrix} \quad (2.14)$$

For all $I = 1, 2$, the input $u^{(I)} \in \mathbb{R}$ is generated taking into account the following control laws

$$u^{(1)} = \frac{1}{g_1^{(1)}} \left[-a_1^{(1)}x_1^{(1)} - a_2^{(1)}x_2^{(1)} - g_2^{(1)}\cos(x_1^{(1)}) - h_2^{(1)}(x_1^{(2)}) + v^{(1)} \right], \quad (2.15)$$

$$u^{(2)} = \frac{1}{g_1^{(2)}} \left[-a_1^{(2)}x_1^{(2)} - a_2^{(2)}x_2^{(2)} - g_2^{(2)}\sin(x_1^{(2)}) - h_2^{(2)}(x_1^{(1)}) + v^{(2)} \right], \quad (2.16)$$

with

$$v^{(I)} = \ddot{x}_{1ref}^{(I)} - \kappa_0 e_1^{(I)} - \kappa_1 e_2^{(I)} \quad (2.17)$$

$$e_1^{(I)} = x_1^{(I)} - x_{1ref}^{(I)} \quad (2.18)$$

$$e_2^{(I)} = x_2^{(I)} - \dot{x}_{1ref}^{(I)} \quad (2.19)$$

where $x_{1ref}^{(I)}(t)$ is the reference signal of $x_1^{(I)}$ and its derivatives $\dot{x}_{1ref}^{(I)}(t)$

and $\ddot{x}_{1ref}^{(I)}(t)$, constituting the desired reference signal vector $r^{(I)}(t)$ shown in Fig. 2.1, i.e. $r^{(I)} = \begin{bmatrix} x_{1ref}^{(I)}(t) & \dot{x}_{1ref}^{(I)}(t) & \ddot{x}_{1ref}^{(I)}(t) \end{bmatrix}^\top$.

The nonlinear local and interconnection dynamics satisfy Assumptions 1-2, while the Lipschitz constant can be determined as shown in (2.20), taking into account the form of local nonlinear dynamics $\gamma_1^{(1)}$ presented in (2.6), we have

$$\begin{aligned} & \left| \gamma^{(1)}(x^{(1)}, u^{(1)}) - \gamma^{(1)}(\hat{x}^{(1)}, u^{(1)}) \right| \leq \left| g_2^{(1)} \right| \left| \cos(x_1^{(1)}) - \cos(\hat{x}_1^{(1)}) \right| \\ & \leq \left| g_2^{(1)} \right| \left| 2 \sin \left(0.5 \left(x_1^{(1)} + \hat{x}_1^{(1)} \right) \right) \sin \left(0.5 \left(x_1^{(1)} - \hat{x}_1^{(1)} \right) \right) \right| \\ & \leq \left| g_2^{(1)} \right| \left| 2 \sin \left(0.5 \left(x_1^{(1)} - \hat{x}_1^{(1)} \right) \right) \right| \leq \left| g_2^{(1)} \right| \left| x_1^{(1)} - \hat{x}_1^{(1)} \right| \end{aligned} \quad (2.20)$$

where the Lipschitz constant is $\lambda_{\gamma_1} = \left| g_2^{(1)} \right|$. The reference signal for the interconnection vector $z^{(I)}$ is

$$z_r^{(I)} = x_{1ref}^{(I)} \quad (2.21)$$

where $x_{1ref}^{(I)}$ is the reference signal of $x_1^{(I)}$.

The term $\eta_2^{(1)}$ represents the modeling uncertainty; for example, the modeling uncertainty may involve parametric uncertainty and unknown dynamics that satisfy Assumption 3; e.g. $\eta_2^{(1)}(x^{(1)}, u^{(1)}, z^{(1)}) = \tilde{a}_1^{(1)} x_1^{(1)} + \tilde{a}_2^{(1)} x_2^{(1)} + \tilde{g}_1^{(1)} u^{(1)} + \tilde{g}_2^{(1)} \cos(x_1^{(1)}) + \tilde{h}_2^{(1)}(x_1^{(2)})$, where $\tilde{a}_1^{(I)}$, $\tilde{a}_2^{(I)}$, $\tilde{g}_1^{(2)}$, $\tilde{g}_2^{(2)}$, $\tilde{b}_1^{(2)}$ are unknown but bounded parameters, while $\tilde{h}_2^{(1)}(x_1^{(2)})$ represents unknown but bounded dynamics. Thus, a bound on $\left| \eta_2^{(1)}(x^{(1)}, u^{(1)}, z^{(1)}) \right|$ can be obtained as

$$\begin{aligned} \left| \eta_2^{(1)}(x^{(1)}, u^{(1)}, z^{(1)}) \right| & \leq \bar{a}_1^{(1)} \left| x_1^{(1)} \right| + \bar{a}_2^{(1)} \left| x_2^{(1)} \right| + \bar{g}_1^{(1)} \left| u^{(1)} \right| \\ & \quad + \bar{g}_2^{(1)} \left| \cos(x_1^{(1)}) \right| + \bar{h}_2^{(1)} \end{aligned} \quad (2.22)$$

where $\bar{a}_1^{(1)}$, $\bar{a}_2^{(1)}$, $\bar{g}_1^{(1)}$, $\bar{g}_2^{(1)}$, $\bar{h}_2^{(1)}$ are known bounds for $\tilde{a}_1^{(1)}$, $\tilde{a}_2^{(1)}$, $\tilde{g}_1^{(1)}$, $\tilde{g}_2^{(1)}$, $\tilde{h}_2^{(1)}(x_1^{(2)})$, respectively.

For controlling each subsystem based on (2.15)-(2.19), two sensors

are used; i.e.

$$\mathcal{S}^{(I)}\{1\} : \quad y_1^{(I)}(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x^{(I)}(t) + d_1^{(I)}(t) + f_1^{(I)}(t), \quad (2.23)$$

$$\mathcal{S}^{(I)}\{2\} : \quad y_2^{(I)}(t) = \begin{bmatrix} 0 & 1 \end{bmatrix} x^{(I)}(t) + d_2^{(I)}(t) + f_2^{(I)}(t), \quad (2.24)$$

while the sensor noise signals $d_1^{(I)}$ and $d_2^{(I)}$ satisfy Assumption 4. It is noted that the interconnection state $x_1^{(I)}$ is measured for all $I = 1, 2$.

2.1.2 Linear Modeling of Cyber-Physical Systems

The physical part $\mathcal{P}^{(I)}$ can be represented by a linear dynamical system, described by

$$\begin{aligned} \Sigma^{(I)} : \quad \dot{x}^{(I)}(t) = & \left(A^{(I)} + \Delta A^{(I)} \right) x^{(I)}(t) + \left(B^{(I)} + \Delta B^{(I)} \right) u^{(I)}(t) \\ & + \left(D_1^{(I)} + \Delta D_1^{(I)} \right) \left(z^{(I)}(t) \right) + \eta_L^{(I)}(t), \end{aligned} \quad (2.25)$$

where $A^{(I)}$, $B^{(I)}$ and $D^{(I)}$ are known system matrices of appropriate dimensions, while $\Delta A^{(I)}$, $\Delta B^{(I)}$ and $\Delta D^{(I)}$ represent parametric uncertainty and the vector $\eta_L^{(I)}$ represents additive system disturbances, satisfying the following assumption:

Assumption 5: The parametric uncertainty matrices are bounded and the additive system disturbance is uniformly bounded; i.e.

$$\left| \Delta A^{(I)} \right| \leq \overline{\Delta A}^{(I)}, \quad (2.26)$$

$$\left| \Delta B^{(I)} \right| \leq \overline{\Delta B}^{(I)}, \quad (2.27)$$

$$\left| \Delta D^{(I)} \right| \leq \overline{\Delta D}^{(I)}, \quad (2.28)$$

$$\left| \Delta \eta_L^{(I)} \right| \leq \bar{\eta}_L^{(I)} \quad (2.29)$$

where $\overline{\Delta A}^{(I)}$, $\overline{\Delta B}^{(I)}$, $\overline{\Delta D}^{(I)}$ and $\bar{\eta}_L^{(I)}$ are known, constant bounds. Compared to the nonlinear case of physical system $\mathcal{P}^{(I)}$, the linear case requires less strict assumptions for the modeling uncertainties. This will be further discussed in Sections 4 and 6.

2.2 Fault Modeling

A fault is defined as an unpermitted deviation of at least one characteristic property or parameter of the system from the accept-

able/usual/standard condition [Blanke et al. [2016], Isermann [2006]]. A fault is represented by its time profile, which includes the time duration and the evolution mode of occurrence or disappearance, and its function (signature). Some examples of fault models with various time duration, time profile and fault function are shown in Fig. 2.2. Based on the time duration, an intermittent fault f can be modeled as [Gertler [1998], Isermann [2006]]:

$$f(t) = \sum_{\nu=1}^{\nu^*} \beta_{\nu}(t_{2\nu-1}, t_{2\nu}) \varphi_{\nu}(t - t_{2\nu-1}), \quad (2.30)$$

where β_{ν} is the time profile and φ_{ν} is the fault function. Particularly, the time profile β_{ν} can be represented by:

$$\beta_{\nu}(T_1, T_2) = \beta_{\nu 1}(t - T_1) - \beta_{\nu 2}(t - T_2), \quad (2.31)$$

where $\beta_{\nu 1}$ is the **evolution mode** of the fault occurrence at T_1 and $\beta_{\nu 2}$ is the evolution mode of disappearance at T_2 . In general, the evolution mode $\beta_{\nu k}$, $k = 1, 2$ can be modeled as

$$\beta_{\nu k}(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\kappa_{\nu k} t}, & t \geq 0 \end{cases} \quad (2.32)$$

where $\kappa_{\nu k} > 0$ is the evolution rate. When $\kappa_{\nu k} \rightarrow \infty$, the occurrence of fault is characterized as **abrupt**. The parameter ν^* is the number of time intervals of the fault presence (e.g. in Fig. 2.2 $\nu^* = 3$ in the case of intermittent faults), $\varphi_{\nu}(t)$ is the sensor fault function within a time interval, and β_{ν} denotes the time profile of the fault that occurs at the time instant $t_{2\nu-1}$ and disappears at the time instant $t_{2\nu}$, with $t_{2\nu-1} < t_{2\nu} < t_{2\nu+1}$. Special cases of intermittent faults are the transient ($\nu^* = 1$) and permanent ($\nu^* = 1, t_2 \rightarrow \infty$) faults. Particularly, transient faults can be modeled as

$$f(t) = \beta_{\nu}(t_1, t_2) \varphi_{\nu}(t - t_1), \quad (2.33)$$

while the permanent faults can be described by

$$f(t) = \beta_{\nu}(t_1, \infty) \varphi_{\nu}(t - t_1), \quad (2.34)$$

The fault function φ_ν is generally time varying. Some special cases of fault functions are: (1) the offset faults, i.e. $\varphi_\nu(t) = \varphi_\nu^\circ$, where φ_ν° is constant magnitude, (2) the drift faults, i.e. $\varphi_\nu(t) = R_\nu t$, where R_ν is the constant slope, and (3) precision degradation, whose fault function $\varphi_\nu(t)$ is a random variate from a given distribution in the case of precision degradation.

In this tutorial, we consider that permanent faults affect the sensors in the set $\mathcal{S}^{(I)}$ with no further assumption on their time profile and evolution rate; i.e. a fault in sensor $\mathcal{S}^{(I)}\{j\}$ is expressed as

$$f_j^{(I)}(t) = \beta_j^{(I)}(t - T_{f_j}^{(I)})\phi_j^{(I)}(t - T_{f_j}^{(I)}), \quad (2.35)$$

where $\phi_j^{(I)}(t) \equiv \varphi_1(t)$ and $T_{f_j}^{(I)}$ is the unknown time instant of fault occurrence and $\beta_j^{(I)}(t - T_{f_j}^{(I)}) \equiv \beta_\nu(T_{f_j}^{(I)}, \infty)$, that is

$$\beta_j^{(I)}(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\kappa_{f_j}^{(I)}t}, & t \geq 0 \end{cases} \quad (2.36)$$

with $\kappa_{f_j}^{(I)} > 0$ be the evolution rate. Multiple sensor faults may occur simultaneously or consecutively; e.g., $T_{f_1}^{(I)} = T_{f_2}^{(I)} = \dots = T_{f_{m_I}}^{(I)}$ in case of **simultaneous** faults (i.e. the onset time of the faults is the same) and $T_{f_1}^{(I)} < T_{f_2}^{(I)} < \dots < T_{f_{m_I}}^{(I)}$ in case of **consecutive** faults (i.e. there is a time gap between the onset time of the faults).

2.3 Summary and Discussion

in this chapter, we initially presented nonlinear and linear models that can be used to describe the behavior of a network of interconnected cyber-physical systems that are affected by multiple sensor faults, taking into account modeling uncertainties such as parametric uncertainty or system disturbances. Moreover, we described how sensor faults affect the output of sensors that are employed for monitoring and controlling each of the interconnected cyber-physical subsystems, as well as several models for sensor faults with various profiles and functions.

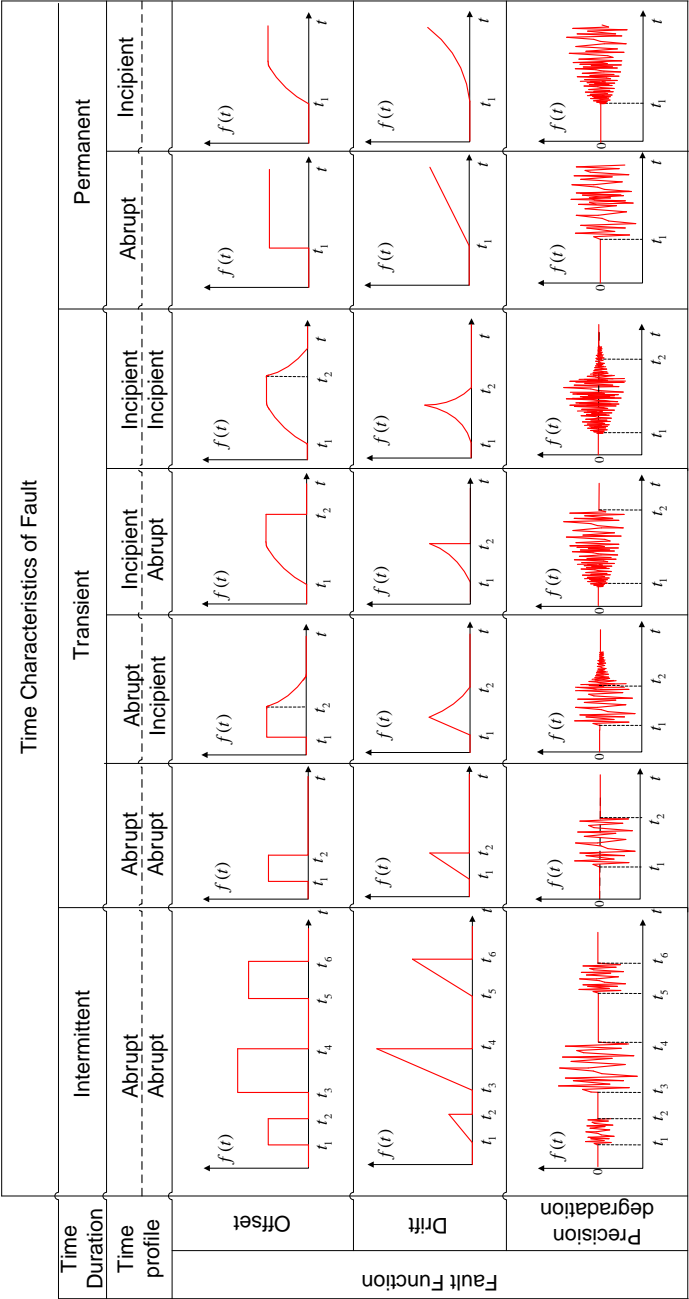


Figure 2.2: Examples of fault models with various time duration (intermittent, transient, permanent), time profile (incipient, abrupt) and fault function (offset, drift, precision degradation). In the case of the intermittent faults, only the abrupt evolution mode of occurrence and disappearance is presented.

3

Sensor Fault Diagnosis Architecture

in this chapter, we present various architectures for designing a sensor fault diagnosis method of a network of cyber-physical systems, where the I -th subsystem $\Sigma^{(I)}$ is described by (2.1), aiming at isolating multiple sensor faults in sensor sets $\mathcal{S}^{(I)}$ determined in (2.2) for all $I \in \{1, \dots, M\}$. The main difference between the architectures relies on the different way of handling the network of M CPS and the available knowledge about the states of every subsystem $\Sigma^{(I)}$, while their common design feature is the decomposition of the set of sensors used for monitoring and control of the set of interconnected CPS or the monolithic system of CPS. It is noted that the goal of this chapter is to present a high level description of the architecture, and the details of the fault detection and fault isolation process will be given in Chapters 4 and 5.

3.1 Centralized Architecture

In a centralized architecture, the network of cyber-physical systems is treated as a monolithic system. Particularly, the physical part, denoted

by \mathcal{P} is described by a nonlinear dynamical system, denoted by Σ ; i.e.,

$$\Sigma: \dot{x}(t) = Ax(t) + \gamma(x(t), u(t)) + \eta(x(t), u(t), t), \quad (3.1)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^\ell$ are the collective state and control input vector of Σ , where x consists of the states $x^{(I)}$ and $z^{(I)}$ and u is comprised of the inputs $u^{(I)}$ for all $I \in \{1, \dots, M\}$. The known system dynamics $Ax + \gamma(x, u)$ collectively represents the known local and interconnection dynamics of the M CPS, $Ax + \gamma(x, u^{(I)}) + h^{(I)}(x^{(I)}, u^{(I)}, C_z^{(I)} z^{(I)}, \zeta^{(I)})$. Assumptions 1-3 presented in Section 2.1.1 are also valid for subsystem Σ . Similarly, the sensor set used for monitoring and controlling the system Σ , denoted by \mathcal{S} , is characterized by the compact output $y \in \mathbb{R}^m$, where $m = \sum_{I=1}^M m_I$; i.e.

$$\mathcal{S}: \quad y(t) = Cx(t) + d(t) + f(t), \quad (3.2)$$

where $y \in \mathbb{R}^m$, $d \in \mathbb{R}^m$ denotes the noise vector corrupting the measurements of sensors in \mathcal{S} and $f \in \mathbb{R}^m$ represents the possible sensor fault vector.

By considering the monolithic system Σ defined in (3.1), the research goal is the isolation of multiple sensor faults that affect the sensor set \mathcal{S} described by (3.2). To this end, we design a monitoring agent, denoted by \mathcal{M} , which has access to the input and output data of the underlying system. The design of \mathcal{M} is realized by decomposing the local sensor set \mathcal{S} into N local sensor sets, as shown in Fig. 3.1. The q -th local sensor set is denoted by $\mathcal{S}^{(q)}$, $q \in \{1, \dots, N\}$, while the corresponding module of the monitoring agent is denoted by $\mathcal{M}^{(q)}$. These local sensor sets may be disjoint or overlapping according to some design specifications related to some conditions for observer stability and multiple sensor fault isolability. The key reason for decomposing the monitoring of the local sensor set \mathcal{S} into N modules is that \mathcal{S} may consist of a large number of sensors, thus making the detection and especially the isolation of multiple sensor faults very difficult and sometimes unfeasible with a single module. The next step is the design of N modules that perform distributed sensor fault detection such that the module $\mathcal{M}^{(q)}$, $q \in \{1, \dots, N\}$, is responsible for monitoring

the healthy operation of the local sensor set $\mathcal{S}^{(q)}$ and detecting sensor faults in it, as presented in Fig. 3.1.

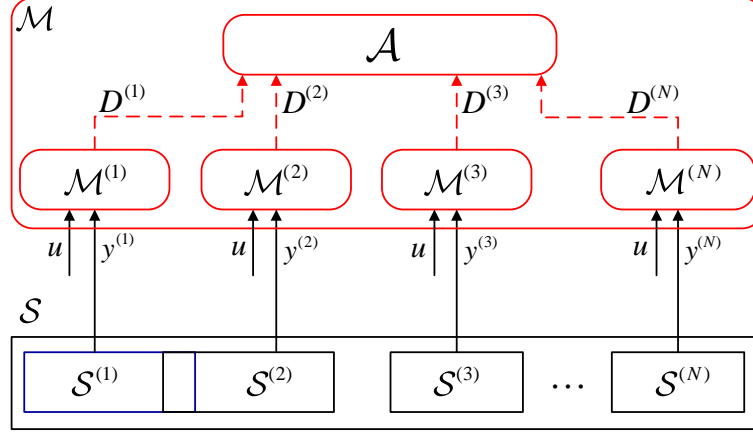


Figure 3.1: Structure of the I -th monitoring agent based on the decomposition of \mathcal{S} into N local sensor sets ($\mathcal{S}^{(q)}$). The module $\mathcal{M}^{(q)}$, $q \in \{1, \dots, N\}$, is responsible for detecting faults in $\mathcal{S}^{(q)}$, while the decisions $D^{(I)}$, $I = \{1, \dots, N\}$ (red, dashed arrows) of the N monitoring modules are integrated and processed by the aggregation module \mathcal{A} , based on a combinatorial decision logic for isolating multiple sensor faults.

Decomposing \mathcal{S} and designing dedicated monitoring modules facilitate the isolation of smaller local sensor sets containing the faulty sensors. Then, the decisions of the monitoring modules, denoted by $D^{(q)}$, are collected and processed by the aggregation module, denoted by \mathcal{A} . The combinatorial process of the decisions is realized using a multiple sensor fault signature matrix and diagnostic reasoning. The utilization of the fault signature matrix [Koscielny et al. [2012], Puig et al. [2006], Meseguer et al. [2010]], along with structured residuals (residuals that are sensitive to a subset of faults) is very efficient in isolating multiple faults in large-scale systems.

The proposed general strategy for decomposing \mathcal{S} starts by considering that the q -th local sensor set $\mathcal{S}^{(q)}$ consists of m_q sensors of \mathcal{S} , and defining the extraction index set $\mathcal{J}^{(q)}$ associated with $\mathcal{S}^{(q)}$; i.e.

$$\mathcal{J}^{(q)} = \{j \in \{1, \dots, m\} : \mathcal{S}\{j\} \in \mathcal{S}^{(q)}\}. \quad (3.3)$$

The set $\mathcal{S}^{(q)}$ is characterized by the output vector $y^{(q)} \in \mathbb{R}^{m_q}$; i.e.,

$$\mathcal{S}^{(q)} : \quad y^{(q)}(t) = C^{(q)}x(t) + d^{(q)}(t) + f^{(q)}(t), \quad (3.4)$$

where $C^{(q)}$ is made up of m_q rows C_j and $y^{(q)}$ represents a column vector made up of m_q elements y_j , for all $j \in \mathcal{J}^{(q)}$ (correspondingly for $d^{(q)}$ and $f^{(q)}$). The decision for the occurrence of $f^{(q)}$ is obtained when a set of analytical redundancy relations, denoted by $\mathcal{E}^{(q)}$ is not satisfied. The set $\mathcal{E}^{(q)}$ is formulated by m_q residuals

$$\varepsilon_j^{(q)}(t) = y_j^{(q)}(t) - C_j \hat{x}^{(q)}(t), \quad j \in \mathcal{J}^{(q)}$$

and their corresponding adaptive thresholds $\bar{\varepsilon}_j^{(q)}(t)$ computed such that

$$|\varepsilon_j^{(q)}(t)| \leq \bar{\varepsilon}_j^{(q)}(t)$$

The set of $\mathcal{E}^{(q)}$ is not satisfied at the first time instant that $|\varepsilon_j^{(q)}(t)| \leq \bar{\varepsilon}_j^{(q)}(t)$, for at least one $j \in \mathcal{J}^{(q)}$. Both the residuals and the adaptive thresholds are designed using a nonlinear observer, which provides the estimation $\hat{x}^{(q)}$. The existence of a stable observer, which uses the measurements of $\mathcal{S}^{(q)}$ (not all the measurements of \mathcal{S}) and estimates the state vector of the nonlinear subsystem Σ , is guaranteed under certain stability conditions, which are derived later on, taking into account Assumptions 1-4 (see Section 4.1).

The use of the monitoring module is the first stage of multiple sensor fault isolation, since we can localize in a distributed manner the subsets of sensors from the global set that contain the faulty sensors. If a local sensor set consists of a single sensor, its dedicated monitoring module is capable of isolating a possible fault in this sensor. When a local sensor set contains more than one sensors and some of the sensors belong also to a few neighboring local sensor sets, then the combinatorial process of module's decisions is necessary for isolating the faulty sensors. Rather than using one aggregation module collecting the decisions of N monitoring modules, the combinatorial logic can be applied to the decisions of N^* monitoring modules, where $N^* \leq N$ (N is the total number of the local sensor sets) denotes the number of a few neighboring and (possibly) overlapping local sensor sets, and be processed by more than one aggregation module.

The problem of decomposing the local sensor set \mathcal{S} into smaller local sensor sets amounts to determining the number of local sensor sets N and the local sensor sets $\mathcal{S}^{(q)}$, for all $q \in \{1, \dots, N\}$. Next, we propose a decomposition procedure, aimed at ensuring the design of stable observers of the monitoring modules and enhancing the multiple sensor fault isolability. The proposed decomposition procedure is summarized in the form of pseudocode in Algorithm 1. The rationale behind the design of the Algorithm 1.

The rationale behind this algorithm is that starting with all possible combinations for grouping sensors that belong to $\mathcal{S}^{(I)}$ (e.g. if $m_I = 3$, $\mathcal{S}^{(I,q)} \leftarrow \{\mathcal{S}^{(I)\{j\}} : j \in \mathcal{K}^{(I)\{q\}}\}$, for all q , where $\mathcal{K}^{(I)} = \{1, 2, 3, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$), we exclude combinations at each of the three main steps, designed as follows. Firstly, we exclude the sensor combinations $\mathcal{S}^{(q)}$ that do not ensure the observability of the pair $(A^{(I)}, C^{(I,q)})$ (lines 1-9), which is related with the linearized part of the nonlinear system Σ . In the following, we exclude the sensor combinations that do not ensure the stability of the nonlinear observer used in the module $\mathcal{M}^{(I,q)}$, taking into account the nonlinear local and interconnection system dynamics (lines 10-21). Finally, we exclude the sensor combinations, aiming at reducing the computational complexity, ensuring the maximum number of isolable sensor fault combinations and the minimum number of mutually non-isolable sensor fault combinations (lines 22-37). The outcome of the algorithm is the number of sensor groups N_I and the sensor groups $\mathcal{S}^{(I,q)}$.

Note that steps 22-40 can be performed before Steps 10-21. Particularly, after step 9, we can perform 22 by replacing \mathcal{J}_s with \mathcal{J}_o ; then, after step 40 we can execute step 10 and use \mathcal{J} instead of \mathcal{J}_o in steps 11-21. In this way, the cardinality of the set \mathcal{J}_o utilized for performing 10-21 may be reduced, leading to the lower complexity for performing the check through step 10-21. In the sequel, Example 3.1 is used to illustrate the application of pseudocode 1, giving emphasis on the applicability of steps 22-40 (the execution of steps 10-21 are omitted for simplicity).

Example 3.1. Let us consider the two subsystems $\Sigma^{(1)}$ and $\Sigma^{(2)}$ described by 2.6 and 2.7 in Example 2.2 as a monolithic system,

Algorithm 1 Decomposition of the sensor set \mathcal{S}

```

1:  $\mathcal{K} \leftarrow$  {all possible  $m^*$  – tuples of indices, where
    $m^* \in \{1, \dots, m-1\}$  {index set of all possible combinations}}
2:  $\mathcal{J}_o \leftarrow \emptyset, \mathcal{J}_s \leftarrow \emptyset$  { $\mathcal{J}_o$ : observability index set,  $\mathcal{J}_s$ : observer stability
   index set}
3: while  $\mathcal{K} \neq \emptyset$  do
4:    $C^* \leftarrow \text{mat}(C_j : j \in \mathcal{K}\{1\})$  {e.g. if  $\mathcal{K}\{1\} = \{1, 3\}$  then
      $\text{mat}(C_j : j \in \{1, 3\})$  corresponds to the 2-row matrix  $[C_1; C_3]$  }
5:   if  $(\mathcal{J}_o \cap \mathcal{K}\{1\} \neq \emptyset$  OR the pair  $(A, C^*)$  is observable) then
6:      $\mathcal{J}_o \leftarrow \mathcal{J}_o \cup \mathcal{K}\{1\}$ 
7:   end if
8:    $\mathcal{K} \leftarrow \mathcal{K} \setminus \mathcal{K}\{1\}$ 
9: end while
10:  $\Lambda \leftarrow (\lambda_\gamma + \lambda_\eta)\{\lambda_\gamma, \lambda_\eta\}$  {Lipschitz constants of  $\gamma$  and  $\bar{\eta}$ }
11: while  $\mathcal{J}_o \neq \emptyset$  do
12:    $C^* \leftarrow \text{mat}(C_j : j \in \mathcal{J}_o\{1\})$ 
13:   Find  $L^* \in \mathbb{R}^{n \times m_q}$  such that  $A - L^*C^*$  is stable
14:   Find  $\rho^*, \xi^* \in \mathbb{R}^+$  such that  $|e^{(A-L^*C^*)t}| \leq \rho^*e^{-\xi^*t}$ 
15:   if  $\xi^* > \rho^*\Lambda$  then
16:      $\mathcal{J}_s \leftarrow \mathcal{J}_s \cup \mathcal{J}_o\{1\}$ 
17:   else
18:     go to 13
19:   end if
20:    $\mathcal{J}_o \leftarrow \mathcal{J}_o \setminus \mathcal{J}_o\{1\}$ 
21: end while
22:  $\mathcal{J} \leftarrow \mathcal{J}_s$  { $\mathcal{J}$ : isolability index set}
23:  $\mathcal{K}_a \leftarrow \{\mathcal{K}, \{1, \dots, m\}\}$ 
24: Create matrix  $F^*$  of dimensions  $\text{card}(\mathcal{J}_s) \times 2^m - 1$ ; the  $q$ -th row
   corresponds to the  $q$ -th element  $\mathcal{J}_s\{q\}$ ; the  $j$ -th column corresponds
   to the  $j$ -th element  $\mathcal{K}_a\{j\}$ ; if  $\mathcal{J}_s\{q\} \cap \mathcal{K}_a\{j\} \neq \emptyset$ , then  $F_{qj} = 1$ , else
    $F_{qj} = 0$ . { $\text{card}()$ : cardinality of a set}
25:  $\nu^* \leftarrow$  number of distinct columns of  $F^*$ 
26:  $\mu^* \leftarrow$  maximum number of mutually identical columns of  $F^*$ 
27:  $\mathcal{J} \leftarrow \mathcal{J}_s$ 

```

```

28: while  $\mathcal{J} \neq \emptyset$  do
29:    $F \leftarrow F^*$  after eliminating the row that corresponds to  $\mathcal{J}\{1\}$ 
30:    $\nu \leftarrow$  number of distinct columns of  $F$ 
31:    $\mu \leftarrow$  maximum number of identical columns of  $F$ 
32:   if  $(\nu = \nu^* \text{ AND } \mu = \mu^*)$  then
33:      $F^* \leftarrow F$ 
34:      $\mathcal{J} \leftarrow \mathcal{J}_s \setminus \mathcal{J}\{1\}$ 
35:   end if
36:    $\mathcal{J} \leftarrow \mathcal{J} \setminus \mathcal{J}\{1\}$ 
37: end while
38:  $N \leftarrow$  number of rows of  $F^*$ 
39:  $\mathcal{J}^{(q)} \leftarrow \mathcal{J}\{q\}$  for all  $q \in \{1, \dots, N\}$ 
40:  $\mathcal{S}^{(q)} \leftarrow \{\mathcal{S}\{j\} : j \in \mathcal{J}^{(q)}\}$  for all  $q \in \{1, \dots, N\}$ 

```

denoted by Σ . By defining the state and input vector of Σ as $x = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_1^{(2)} & x_2^{(2)} \end{bmatrix}^\top$, and $u = \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} u^{(1)} & u^{(2)} \end{bmatrix}^\top$, we have

$$\Sigma : \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ a_1^{(1)} & a_2^{(1)} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & a_1^{(2)} & a_2^{(2)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (3.5)$$

$$+ \begin{bmatrix} 0 \\ g_1^{(1)}u_1 + g_2^{(1)}\cos(x_1) \\ 0 \\ g_1^{(2)}u_2 + g_2^{(2)}\sin(x_3) \end{bmatrix} + \begin{bmatrix} 0 \\ h_2^{(1)}(x_3) \\ 0 \\ h_2^{(2)}(x_1) \end{bmatrix} \\ + \begin{bmatrix} 0 \\ \eta_2^{(1)}(x, u) \\ 0 \\ \eta_2^{(1)}(x, u) \end{bmatrix} \quad (3.6)$$

Taking into account the two local sensors sets described by 2.23 and 2.24, let us denote \mathcal{S} the sensor set associated with Σ , which is comprised by four sensors, that is $\mathcal{S} = \{\mathcal{S}^{(1)}\{1\}, \mathcal{S}^{(1)}\{2\}, \mathcal{S}^{(2)}\{1\}, \mathcal{S}^{(2)}\{2\}\}$ with $\mathcal{S}\{1\} = \mathcal{S}^{(1)}\{1\}$, $\mathcal{S}\{2\} = \mathcal{S}^{(1)}\{2\}$, $\mathcal{S}\{3\} = \mathcal{S}^{(2)}\{1\}$ and $\mathcal{S}\{4\} = \mathcal{S}^{(2)}\{2\}$, and is characterized by the output vector $y = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \end{bmatrix} = \begin{bmatrix} y_1^{(1)} & y_2^{(1)} & y_1^{(2)} & y_2^{(2)} \end{bmatrix}$; i.e.

$$\mathcal{S} : \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} \quad (3.7)$$

where $d = \begin{bmatrix} d_1^{(1)} & d_2^{(1)} & d_1^{(2)} & d_2^{(2)} \end{bmatrix}$, $f = \begin{bmatrix} f_1^{(1)} & f_2^{(1)} & f_1^{(2)} & f_2^{(2)} \end{bmatrix}$. Consider the following system parameters $a_1^{(1)} = 9$, $a_2^{(1)} = -2$, $g_1^{(1)} = 1$, $g_2^{(1)} = 1$, $b^{(1)} = 9$, $a_1^{(2)} = 12$, $a_2^{(2)} = -11$, $g_1^{(2)} = 1$, $g_2^{(2)} = 2$, $b^{(2)} = 12$, and two different cases for $h_2^{(1)}$ and $h_2^{(2)}$:

Case I: $h_2^{(1)} = b^{(1)}x_3$ and $h_2^{(2)} = b^{(2)}x_1$

Case II: $h_2^{(1)} = b^{(1)}\sin(x_3)$ and $h_2^{(2)} = b^{(2)}\cos(x_1)$.

By treating the two subsystems as a monolithic system, the interconnection functions of the two subsystems in the first case are involved in the linearized part of Σ , while in the second case the interconnection functions of the two subsystems are involved in the nonlinear part of Σ . In both cases, given that $m = 4$ we initially define the set $\mathcal{K} = \{1, 2, 3, 4, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}\}$, and then we execute steps 1-9 and 22-40.

Case I: By executing steps 1-9, we obtain the set $\mathcal{J}_o = \{1, 3, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}\}$. Note that we have excluded the index tuples $2, 4, \{2, 4\}$, because the sensor set that contain either sensor $\mathcal{S}\{2\}$, or $\mathcal{S}\{4\}$, or $\mathcal{S}\{2\}$, and $\mathcal{S}\{4\}$ does not ensure the observability of the pairs (A, C_2) , or (A, C_4) , or $(A, [C_2; C_4])$, respectively. The auxiliary matrix F_1^* is defined as shown in Table 3.1. Specifically, based on Table 3.1, the number of distinct

columns of F_1^* is 12, while the maximum number of mutually identical columns is 4 since the tuples $\{1, 3\}$, $\{1, 2, 3\}$, $\{1, 3, 4\}$, $\{1, 2, 3, 4\}$ are associated with the same column-pattern (colored blue in Table 3.1). Next, we perform steps 22-40 with $\mathcal{J}_s = \mathcal{J}_o$. By performing steps 22-40, we observe the following; for instance, if we eliminate row 1 (i.e. exclude the tuple $\{1\}$), we obtain the matrix shown in Table 3.2. The number of distinct columns of this matrix is 11, while the maximum number of identical columns is 5, which is higher than the maximum number of the corresponding number of the initial F^* . If we eliminate row 3 (i.e. exclude the tuple $\{1, 2\}$), we obtain the matrix shown in Table 3.3. The number of distinct columns of this matrix is 11, which is lower than the corresponding number of the initial F^* , while the tuples $\{3, 4\}$ and $\{2, 3, 4\}$ are associated with identical column-patterns (colored red in Table 3.3). The maximum number of identical columns is 4. On the contrary if we eliminate row 4 (i.e. exclude the tuple $\{1, 3\}$), we obtain the matrix presented in Table 3.4), whose number of distinct columns of this matrix and maximum number of identical columns are the same the corresponding number of the initial F^* , shown in Table 3.1. Since the the number of distinct columns and maximum number of identical columns do not change, we can eliminate row 4 and continue by eliminating the remainder rows of the updated matrix F_1^* given in Table 3.4, starting with the row associated with the tuple $\{1, 4\}$. The final version of the matrix is illustrated in Table 3.5.

Case II: By executing steps 1-9, we obtain the set $\mathcal{J}_o = \{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$, $\{2, 3, 4\}$. Note that we have excluded the index tuples $1, 2, 3, 4$, $\{1, 2\}$ and $\{3, 4\}$, because any sensor set that contains either sensor $\mathcal{S}\{1\}$, or $\mathcal{S}\{2\}$, or $\mathcal{S}\{3\}$, or $\mathcal{S}\{4\}$, or $\mathcal{S}\{1\}$ and $\mathcal{S}\{2\}$, or $\mathcal{S}\{3\}$ and $\mathcal{S}\{4\}$, does not ensure the observability of the pair (A, C_1) , or (A, C_2) , (A, C_3) , or (A, C_4) , or $(A, [C_1; C_2])$, or $(A, [C_3; C_4])$. In contrast to Case I, there are more combinations of sensors, or single sensors that do not guarantee the observability property. Intuitively, this happens because the interconnection functions of the two subsystems are involved in the nonlinear part of the monolithic system Σ . The auxiliary matrix F_2^* is defined as shown in Table 3.6.

3.1. Centralized Architecture

Table 3.1: Auxiliary matrix F_1^* .

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$	$\{1,2,3\}$	$\{1,2,4\}$	$\{1,3,4\}$	$\{2,3,4\}$	$\{1,2,3,4\}$
$\{1\}$	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
$\{3\}$	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
$\{1,2\}$	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
$\{1,3\}$	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1
$\{1,4\}$	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1
$\{2,3\}$	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
$\{3,4\}$	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1
$\{1,2,3\}$	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
$\{1,2,4\}$	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
$\{1,3,4\}$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$\{2,3,4\}$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.2: Auxiliary matrix after eliminating row 1 of F_1^* shown in Table 3.1..

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\{3\}$	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
$\{1, 2\}$	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
$\{1, 3\}$	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1
$\{1, 4\}$	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1
$\{2, 3\}$	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
$\{3, 4\}$	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1
$\{1, 2, 3\}$	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
$\{1, 2, 4\}$	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
$\{1, 3, 4\}$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$\{2, 3, 4\}$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.3: Auxiliary matrix arisen after eliminating row 3 of F_1^* shown in Table 3.1.

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\{1\}$	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
$\{3\}$	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
$\{1, 3\}$	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1
$\{1, 4\}$	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1
$\{2, 3\}$	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
$\{3, 4\}$	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1
$\{1, 2, 3\}$	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
$\{1, 2, 4\}$	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
$\{1, 3, 4\}$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$\{2, 3, 4\}$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.4: Auxiliary matrix arisen after eliminating row 4 of F_1^* shown in Table 3.1.

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\{1\}$	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
$\{3\}$	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
$\{1, 2\}$	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
$\{1, 4\}$	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1
$\{2, 3\}$	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
$\{3, 4\}$	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1
$\{1, 2, 3\}$	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
$\{1, 2, 4\}$	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
$\{1, 3, 4\}$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$\{2, 3, 4\}$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.5: Auxiliary matrix arisen after eliminating rows 4 and 8-11 of F_1^* shown in Table 3.1.

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$	$\{1,2,3\}$	$\{1,2,4\}$	$\{1,3,4\}$	$\{2,3,4\}$	$\{1,2,3,4\}$
$\{1\}$	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
$\{3\}$	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
$\{1,2\}$	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
$\{1,4\}$	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1
$\{2,3\}$	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
$\{3,4\}$	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1

As observed the number of distinct columns of F_1^* is 9, while the maximum number of mutually identical columns is 7 colored blue in Table 3.6. Next, we execute steps 22-40 with $\mathcal{J}_s = \mathcal{J}_o$, aiming at reducing the number of rows (i.e. number of sensor sets and corresponding monitoring modules). The final version of F_2^* is presented in Table 3.7, which results from the elimination of the rows associated with the tuples $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$, $\{2, 3, 4\}$ that does not modify the number of distinct columns, and maximum number of identical columns. By comparing F_1^* and F_2^* shown in Tables 3.5 and 3.7, respectively, we can observe that in case II the sensor set \mathcal{S} is decomposed in local sensor sets, which are less than the local sensor sets resulted in case I. The number of distinct columns in case I is larger than in case II, and the maximum number of identical columns in case I is smaller than in case II.

3.2 Decentralized Architecture

This Section presents a decentralized architecture for detecting and isolating multiple sensor faults in a network of interconnected CPS, as shown in Fig. 3.2, where the decentralized process is realized in the cyber superstratum. The backbone of the proposed decentralized methodology is the design of a local sensor fault diagnosis agent dedicated to each interconnected subsystem as illustrated in Fig. 3.2. Particularly, the cyber part $\mathcal{C}^{(I)}$ of the interconnected CPS consists of the monitoring agent, denoted by $\mathcal{M}^{(I)}$, associated with each interconnected subsystem and the control agent $\mathcal{K}^{(I)}$. The monitoring agent $\mathcal{M}^{(I)}$ does not communicate with other agents, but uses *a priori known reference signals related to the system interconnections* as stated in Assumption 1, taking advantage of the available knowledge that stems from the control design. The agent $\mathcal{M}^{(I)}$, shown in Fig. 3.3, is designed to detect and isolate multiple sensor faults that may affect directly some sensors in $\mathcal{S}^{(I)}$ associated with the underlying system $\Sigma^{(I)}$.

For isolating multiple sensor faults, the monitoring of the sensor set is decomposed into several modules dedicated to smaller sets of sensors, as shown in Fig. 3.1. The key reason for decomposing the monitoring of

Table 3.6: Auxiliary matrix F_2^* .

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\{1, 3\}$	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1
$\{1, 4\}$	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1
$\{2, 3\}$	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
$\{2, 4\}$	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1
$\{1, 2, 3\}$	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
$\{1, 2, 4\}$	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
$\{1, 3, 4\}$	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
$\{2, 3, 4\}$	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.7: Auxiliary matrix arisen from the elimination of rows 5-8 of F_2^* in Table 3.6.

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\{1, 3\}$	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1
$\{1, 4\}$	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1
$\{2, 3\}$	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1
$\{2, 4\}$	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1

the local sensor set $\mathcal{S}^{(I)}$ into N_I modules is that although a large-scale system may be already decomposed into smaller subsystems, $\mathcal{S}^{(I)}$ may still consist of a large number of sensors, thus making the detection and especially the isolation of multiple sensor faults very difficult and sometimes unfeasible with a single module. The decomposition of the sensor set $\mathcal{S}^{(I)}$ for all $I \in \{1, \dots, M\}$ is realized based on the Algorithm 1 by replacing A , C , Λ with $A^{(I)}$, $C^{(I)}$, Λ_I , with $\Lambda_I = \lambda_{\gamma_I} + \lambda_{h_I} + \lambda_{\eta_I}$, where λ_{γ_I} , λ_{h_I} , λ_{η_I} are Lipschitz constants related to γ_I , h_I and $\bar{\eta}$, defined in (2.1). The local set is denoted by $\mathcal{S}^{(I,q)}$ and is characterized by the output vector $y^{(I,q)} \in \mathbb{R}^{m_{I,q}}$; i.e.,

$$\mathcal{S}^{(I,q)} : \quad y^{(I,q)}(t) = C^{(I,q)}x^{(I)}(t) + d^{(I,q)}(t) + f^{(I,q)}(t), \quad (3.8)$$

where $C^{(I,q)}$ is made up of $m_{I,q}$ rows $C_j^{(I)}$ and $y^{(I,q)}$ represents a column vector made up of $m_{I,q}$ elements $y_j^{(I)}$, for all $j \in \mathcal{J}^{(q)}$ (correspondingly for $d^{(I,q)}$ and $f^{(I,q)}$), while the associated index set is defined as

$$\mathcal{J}^{(I,q)} = \{j \in \{1, \dots, m\} : \mathcal{S}^{(I)}\{j\} \in \mathcal{S}^{(I,q)}\}. \quad (3.9)$$

Then, the multiple sensor fault isolation decision logic is obtained by the aggregation module $\mathcal{A}^{(I)}$ that collects and processes combinatorially the decisions of the monitoring modules $\mathcal{M}^{(I,q)}$ (as shown in Fig. 3.1).

Among the advantages of using a decentralized diagnosis scheme is the fact that its implementation does not require additional cost and communication capacity, since the monitoring agents do not exchange any information, while there is no need for designing and applying appropriate mechanisms for handling time-delays, such as network access and transmission delays, packet dropouts due to network congestion or transmission errors in physical links, or network drop-out. Under the assumption of multiple sensor faults, the performance of a decentralized scheme is not affected by the propagation of sensor faults through the exchange of information between monitoring agents. Moreover, the decentralized approach can be less vulnerable to deceptive cyber attacks in the communication channels, or communication failures, forming a complement to distributed architectures, especially in safety-critical systems [Mohajerin Esfahani et al. [2010], Amin et al. [2013]].

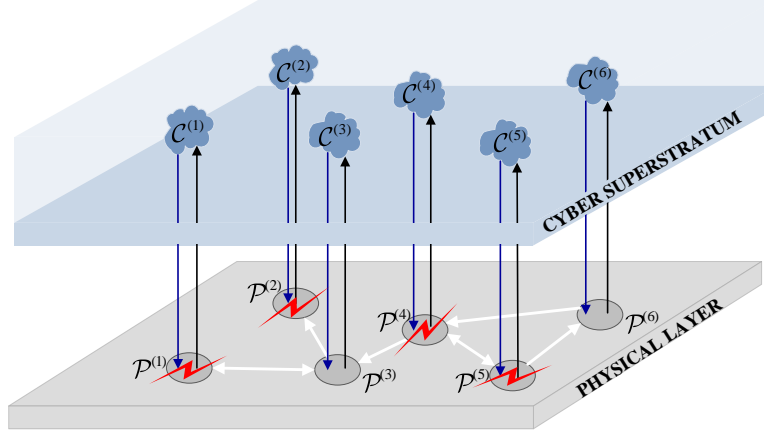


Figure 3.2: Network of interconnected cyber-physical systems in a decentralized sensor FDI scheme.

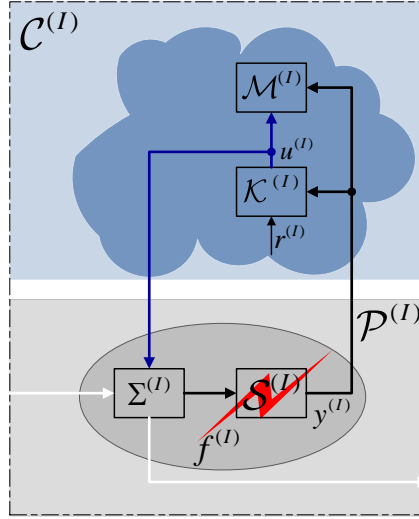


Figure 3.3: Description of the I -th CPS affected by sensor faults, where the monitoring agent operates in a decentralized sensor FDI scheme.

3.3 Distributed Architecture

This Section provides an high-level description of a distributed architecture for detecting and isolating sensor faults in a network of interconnected cyber-physical systems, where the distributed sensor FDI process is realized in the cyber superstratum, in two levels, as presented in Fig. 3.4. The first-level diagnosis $\mathcal{M}^{(I)}$, shown in Fig. 3.5, is designed to detect and isolate multiple sensor faults that may affect directly some sensors in $\mathcal{S}^{(I)}$ associated with the underlying system $\Sigma^{(I)}$. The second-level diagnosis, denoted by \mathcal{G} , utilizes information (depicted as dashed arrows in Fig. 3.4) from the monitoring agents $\mathcal{M}^{(I)}$ for isolating sensor faults propagated in the cyber superstratum due to the communication of the monitoring agents $\mathcal{M}^{(I)}$.

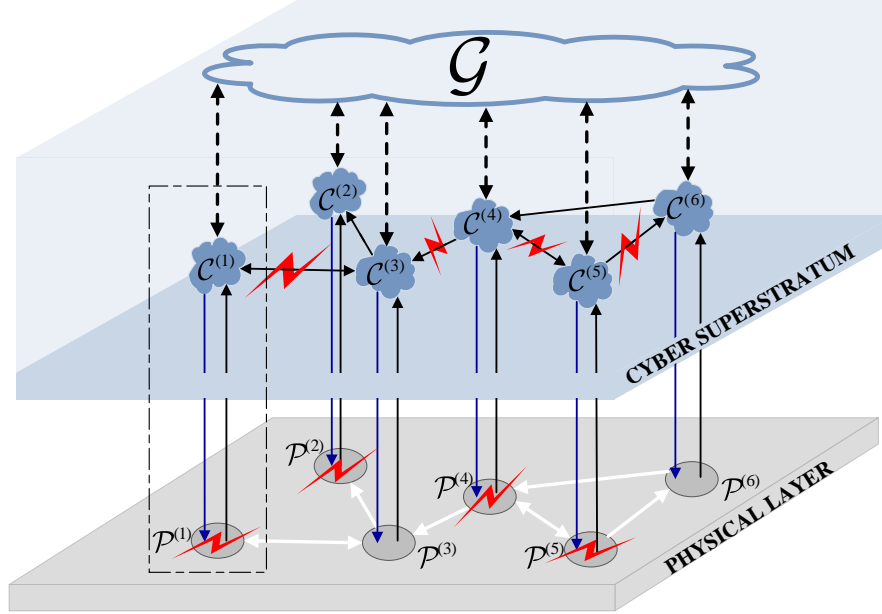


Figure 3.4: Network of interconnected cyber-physical systems deployed in a distributed sensor FDI scheme.

The exchanged information corresponds to the form of the physical interconnections, i.e. unidirectional or bidirectional interactions be-

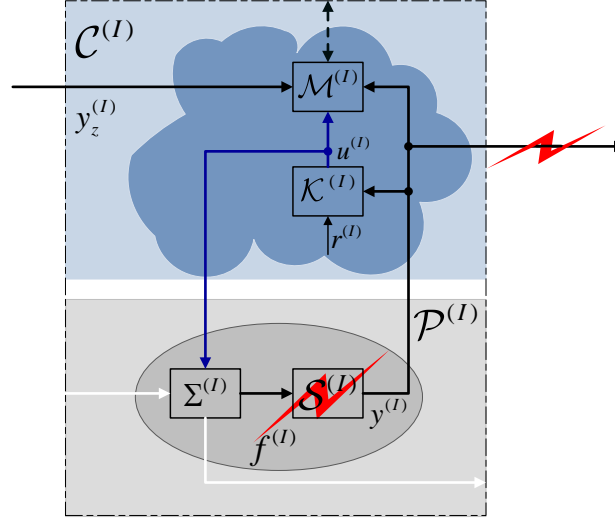


Figure 3.5: Description of the I -th CPS affected by sensor faults, where the monitoring agent operates in a distributed sensor FDI scheme.

tween interconnected subsystems (white, solid arrows in Fig. 3.4). In particular, neighboring monitoring agents can exchange data provided by the sensors that measure some interconnection states $z^{(I)}$. Then, the monitoring agent $\mathcal{M}^{(I)}$ is designed as described in Section 3.1, taking into account the following state space representation of subsystem $\Sigma^{(I)}$

$$\begin{aligned} \Sigma^{(I)} : \quad \dot{x}^{(I)}(t) = & A^{(I)}x^{(I)}(t) + \gamma^{(I)}(x^{(I)}(t), u^{(I)}(t)) \\ & + h^{(I)}(x^{(I)}(t), u^{(I)}(t), \zeta^{(I)}(t), C_z^{(I)}z^{(I)}(t)) \\ & + \eta^{(I)}(x^{(I)}(t), u^{(I)}(t), \zeta^{(I)}(t), C_z^{(I)}z^{(I)}(t), t), \end{aligned} \quad (3.10)$$

where the vector $C_z^{(I)}z^{(I)}$ represents a linear combination of measurable interconnection states with $C_z^{(I)} \in \mathbb{R}^{q_I \times p_I}$ and $\zeta^{(I)} \in \mathbb{R}^{r_I}$ collects a set of interconnection states $z^{(I)}$ that are not measured, while the reference signals of $\zeta^{(I)}$ are assumed to be known, according to Assumption 1 in Section 2.1.1. Then, it is considered that the sensor information transmitted to the agent $\mathcal{M}^{(I)}$, denoted by $\mathcal{S}_z^{(I)}$, is characterized by the output vector $y_z^{(I)} \in \mathbb{R}^{p_I}$; i.e.

$$\mathcal{S}_z^{(I)} : y_z^{(I)}(t) = C_z^{(I)}z^{(I)}(t) + d_z^{(I)}(t) + f_z^{(I)}(t), \quad (3.11)$$

where $d_z^{(I)}, f_z^{(I)} \in \mathbb{R}^{p_I}$ are the noise and sensor fault vector, respectively. The monitoring agent is designed as presented in Section 3.1, where the sensor set $\mathcal{S}^{(I)}$ is decomposed according to the Algorithm 1 by replacing A, C, Λ with $A^{(I)}, C^{(I)}, \Lambda_I$, with $\Lambda_I = \lambda_{\gamma_I} + \lambda_{h_I} + \lambda_{\eta_I}$ where $\lambda_{\gamma_I}, \lambda_{h_I}, \lambda_{\eta_I}$ are Lipschitz constants related to γ_I, h_I and $\bar{\eta}$, defined in (2.1), while the local sensor set $\mathcal{S}^{(I,q)}$ is defined in (3.8). Moreover, the observer in the monitoring module $\mathcal{M}^{(I,q)}$ is designed based on system representation (3.10). Note that model (3.10) is equivalent to model (2.1), but the first model highlights the fact that some of interconnection states are measurable.

A fault in sensor $\mathcal{S}^{(I)}$ can be propagated to neighboring cyber modules due to the information exchange of $\mathcal{M}^{(I)}$ with neighboring monitoring agents, as depicted in Fig. 3.5. Allowing the exchange of sensor information between monitoring agents can enhance sensor fault detectability, compared to a decentralized architecture with no communication between the monitoring agents, since it can reduce the estimation error associated with the residuals and adaptive thresholds, based on which the the monitoring modules and consequently the monitoring agents obtain a decision. On the other hand, the exchange of information may cause fault propagation, which complicates the isolation of faulty sensors. In this case, a local agent would have difficulty in distinguishing between local and propagated sensor faults. This issue is addressed in Section 5 with the design of a global decision logic. It is important to note that the propagation of sensor faults may occur not only in the monitoring agents but also in the control agents in the case of a distributed control architecture, however the design of the control and fault accommodation scheme is beyond the scope of this paper.

3.4 Summary and Discussion

in this chapter, we discussed three architectures, i.e. centralized, decentralized and distributed, that can be adopted for designing a sensor fault diagnosis technique aiming at detecting and isolating multiple sensor faults that may occur in a network of cyber-physical systems. In a centralized scheme, in order to isolate multiple sensor faults, we

proposed the decomposition of the monitoring of the sensor set into several modules that monitor smaller sensor sets, and the collection and combinatorial process of the decisions of the monitoring modules by an aggregation module. In a decentralized architecture, we propose the design of a monitoring agent dedicated to each interconnected CPS, while the agents do not exchange any information but use a priori known reference signals related to the system interconnections, taking advantage of the available knowledge that stems from the control design. Following a distributed approach, we proposed the deployment of monitoring agents, where every agent is dedicated to a corresponding interconnected subsystem, while it is allowed to exchange information with its neighboring monitoring agents. For the isolation of the sensor faults that can be propagated due to the exchange of information, a global aggregation module is designed that collects and processes the decisions of the monitoring agents. The monitoring agents and modules, as well as the aggregation modules are part of the cyber part of CPS.

The three architectures vary with respect to the way of handling the network of M CPS, i.e. as a set of interconnected physical subsystems or a monolithic system, while their common characteristic is the decomposition of the monitoring of the sensor set into subsets of sensors, based on the pseudocode in Algorithm 1. Following a centralized approach, the process of determining the local sensor sets $\mathcal{S}^{(q)}$ can be more complex compared to a distributed or decentralized approach, due to the larger number of sensors in \mathcal{S} and the larger scale of the system that makes the observability of the pair $A, C^{(q)}$ more difficult and increases the values of λ_γ and λ_h .

4

Sensor Fault Detection

The process of sensor fault detection is executed by every monitoring agent, though the monitoring modules resulted from the decomposition of the sensor set as presented in Section 3.1. This chapter describes the design of the module that monitors a specific subset of sensors. The first stage of the decision-making process of the monitoring module is the generation of residuals. Residuals are features that portray the status of the monitoring subsystem. Any unusual change in these features may imply the presence of faults. In this tutorial, residuals represent the deviations of the sensor data (observed behavior) from the estimated sensor outputs (expected behavior). Due to the presence of disturbances and sensor measurement noise, the observed behavior is typically not identical to the expected behavior even during the healthy operation of the sensors. For this reason, the residuals are compared to thresholds that are designed to bound the residuals under healthy conditions, ensuring the robustness of the monitoring modules (and consequently the monitoring agent) with respect to various sources of uncertainties. Finally, a monitoring module infers the presence of sensor faults in the local sensor set, when a set of analytical redundancy relations of observer-based residuals and adaptive thresholds is not satisfied.

The stages of sensor fault detection are the same for all diagnostic architectures: (i) residual generation, (ii) computation of adaptive thresholds, and (iii) sensor fault detection decision logic.

4.1 Residual Generation

In a non-centralized fault diagnostic scheme, the j -th residual, denoted by $\varepsilon_{y_j}^{(I,q)}$ with $j \in \mathcal{J}^{(I,q)}$, $I \in \{1, \dots, M\}$, $q \in \{1, \dots, N_I\}$, is defined as

$$\varepsilon_{y_j}^{(I,q)} = y_j^{(I)} - C_j^{(I)} \hat{x}^{(I,q)}, \quad (4.1)$$

where $y_j^{(I)}$ is the output of the sensor $\mathcal{S}^{(I)}\{j\}$ that belongs to the local sensor set $\mathcal{S}^{(I,q)}$, $\mathcal{J}^{(I,q)}$ is given in (3.9), and $\hat{x}^{(I,q)}$ is generated by a distributed or decentralized nonlinear observer that is associated with the monitoring module $\mathcal{M}^{(I,q)}$. In a centralized fault diagnostic scheme, where the network of CPS is treated as a monolithic system by (3.1), the j -th residual is defined by

$$\varepsilon_{y_j}^{(q)} = y_j - C_j \hat{x}^{(q)}, \quad j \in \mathcal{J}^{(q)}, \quad (4.2)$$

where $\mathcal{J}^{(q)}$ is defined in (3.3) and $\hat{x}^{(q)}$ is generated by a centralized observer associated with the module $\mathcal{M}^{(q)}$.

4.1.1 Distributed Observer-based Estimation Model

in this section we provide the structure of the nonlinear distributed observer implemented in the module $\mathcal{M}^{(I,q)}$ for generating the estimation $\hat{x}^{(q)}$ of the state vector of $\Sigma^{(I)}$ defined in (3.10). We also present the structure of a linear distributed observer in order to illustrate the difference on the conditions for ensuring the stability of the state estimation error dynamics.

Nonlinear Observer

In a distributed architecture presented in Section 3.3, the estimation model associated with the monitoring module $\mathcal{M}^{(I,q)}$ is generated by

the nonlinear observer, described by

$$\begin{aligned} \mathcal{O}^{(I,q)} : \quad \dot{\hat{x}}^{(I,q)} = & A^{(I)}\hat{x}^{(I,q)} + \gamma^{(I)}(\hat{x}^{(I,q)}, u^{(I)}) + h^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_z^{(I)}) \\ & + L^{(I,q)} \left(y^{(I,q)} - C^{(I,q)}\hat{x}^{(I,q)} \right), \end{aligned} \quad (4.3)$$

where $\hat{x}^{(I,q)} \in \mathbb{R}^{n_I}$ is the estimation of $x^{(I)}$ ($x^{(I)}$ is the state vector of subsystem $\Sigma^{(I)}$ described by (3.10)) based on the sensor measurements $y^{(I,q)}$ with initial conditions $\hat{x}^{(I,q)}(0) \in \mathcal{X}_I$, $L^{(I,q)} \in \mathbb{R}^{n_I \times m_{I,q}}$ is the observer gain matrix, $\zeta_r^{(I)}$ are reference signals of the unmeasured interconnection vector $\zeta^{(I)}$ (see Assumption 1 in Section 2.1.1) and $y_z^{(I)}$ is the transmitted sensor information, defined in (3.11). It is noted that the observer used in this work is based on the structure of observers for Lipschitz nonlinear systems (see [Rajamani [1998], Zhu and Han [2002]] and references therein), which is modified appropriately for nonlinear interconnected subsystems. Also, the generation of the estimation model $\hat{x}^{(I,q)}$ can be affected by the occurrence of faults in $\mathcal{S}^{(I,q)}$ and $\mathcal{S}_z^{(I)}$ since it is realized using measurements $y^{(I,q)}$ and $y_z^{(I)}$.

As long as the local sensor set $\mathcal{S}^{(I,q)}$ and the transmitted information $y_z^{(I)}$ are not affected by sensor faults, the estimation model of $\mathcal{M}^{(I,q)}$ under healthy conditions, denoted by $\hat{x}_H^{(I,q)}$, satisfies

$$\begin{aligned} \dot{\hat{x}}_H^{(I,q)} = & A_L^{(I,q)}\hat{x}_H^{(I,q)} + \gamma^{(I)}(\hat{x}_H^{(I,q)} + h^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_{z_H}^{(I)}, u^{(I)}) \\ & + L^{(I,q)}y_H^{(I,q)}, \end{aligned} \quad (4.4)$$

$$A_L^{(I,q)} = A^{(I)} - L^{(I,q)}C^{(I,q)} \quad (4.5)$$

where $y_H^{(I,q)}$ and $y_{z_H}^{(I)}$ are respectively defined in the following equations

$$y_H^{(I,q)}(t) = C^{(I,q)}x^{(I)}(t) + d^{(I,q)}(t), \quad (4.6)$$

$$y_{z_H}^{(I)}(t) = C_z^{(I)}z^{(I)}(t) + d_z^{(I)}(t). \quad (4.7)$$

Let us define $\varepsilon_{x_H}^{(I,q)} \triangleq x^{(I)} - \hat{x}_H^{(I,q)}$ as the state estimation error under

healthy conditions; taking into account (2.1) and (4.4), we obtain:

$$\begin{aligned} \dot{\varepsilon}_{x_H}^{(I,q)} &= A_L^{(I,q)} \varepsilon_{x_H}^{(I,q)} + \tilde{\gamma}_H^{(I,q)} + \tilde{h}_H^{(I,q)} - L^{(I,q)} d^{(I,q)} \\ &\quad + \eta^{(I)}(x^{(I)}, u^{(I)}, \zeta^{(I)}, C_z^{(I)} z^{(I)}, t), \end{aligned} \quad (4.8)$$

$$\tilde{\gamma}_H^{(I,q)} = \gamma^{(I)}(x^{(I)}, u^{(I)}) - \gamma^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}), \quad (4.9)$$

$$\tilde{h}_H^{(I,q)} = h^{(I)}(x^{(I)}, u^{(I)}, \zeta^{(I)}, C_z^{(I)} z^{(I)}) - h^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_{z_H}^{(I)}) \quad (4.10)$$

If the pair $(A^{(I)}, C^{(I,q)})$ is observable, we can find $L^{(I,q)}$ such that the matrix $A_L^{(I,q)}$ is stable. Consequently, we can choose positive constants $\rho^{(I,q)}$, $\xi^{(I,q)}$ and $\rho_d^{(I,q)}$, $\xi_d^{(I,q)}$ such that

$$\left| e^{A_L^{(I,q)} t} \right| \leq \rho^{(I,q)} e^{-\xi^{(I,q)} t} \quad (4.11)$$

$$\left| e^{A_L^{(I,q)} t} L^{(I,q)} \right| \leq \rho_d^{(I,q)} e^{-\xi_d^{(I,q)} t}. \quad (4.12)$$

Let us define the function

$$\Phi^{(I,q)}(t) = \rho^{(I,q)} e^{-\xi^{(I,q)} t} \quad (4.13)$$

and the constants

$$\nu^{(I,q)} = \xi^{(I,q)} - \rho^{(I,q)} \Lambda_I \quad (4.14)$$

$$\Lambda_I = \lambda_{\gamma_I} + \lambda_{h_I} + \lambda_{\eta_I} \quad (4.15)$$

where the parameters λ_{γ_I} , λ_{h_I} , λ_{η_I} denote respectively the Lipschitz constants of $\gamma^{(I)}$, $h^{(I)}$ and $\bar{\eta}^{(I)}$ that satisfy Assumptions 2 and 3.

The stability of the distributed estimation error dynamics under healthy conditions is analyzed in the Theorem 4.1, taking into account that $\bar{d}^{(I,q)}$ and $\bar{d}_z^{(I)}$ are noise bounds for $d^{(I,q)}(t)$ and $d_z^{(I)}$ (respectively defined in (3.8) and (3.11)) such that $|d^{(I,q)}(t)| \leq \bar{d}^{(I,q)}$ and $|d_z^{(I)}(t)| \leq \bar{d}_z^{(I)}$, $\bar{x}^{(I,q)}$ is a bound on $|\varepsilon_{x_H}^{(I,q)}(0)|$ and $\bar{\zeta}^{(I)}$ is defined as a bound on the error between the actual interconnections and their reference signals such that $|\zeta^{(I)}(t) - \zeta_r^{(I)}(t)| \leq \bar{\zeta}^{(I)}$, $\forall t \geq 0$.

Theorem 4.1. Suppose that the observer gain $L^{(I,q)}$ is chosen such that: (a) the matrix $A_L^{(I,q)} = A^{(I)} - L^{(I,q)} C^{(I,q)}$ is stable, and (b) the

positive constants $\rho^{(I,q)}$, $\xi^{(I,q)}$ satisfy $\xi^{(I,q)} > \Lambda_I \rho^{(I,q)}$; then the state estimation error under healthy conditions, $\varepsilon_{x_H}^{(I,q)}(t)$, is uniformly bounded and satisfies

$$|\varepsilon_{x_H}^{(I,q)}(t)| \leq Z_H^{(I,q)}(t) \quad (4.16)$$

where

$$Z_H^{(I,q)}(t) = E_H^{(I,q)}(t) + \rho^{(I,q)} \Lambda_I \int_0^t E_H^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau, \quad (4.17)$$

$$\begin{aligned} E_H^{(I,q)}(t) = & \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\ & + \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) d\tau \\ & + \frac{(\lambda_{h_I} + \lambda_{\eta_I})(\bar{d}_z^{(I)} + \bar{\zeta}^{(I)})}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right). \end{aligned} \quad (4.18)$$

Proof. Solving (4.8)

$$\begin{aligned} \varepsilon_{x_H}^{(I,q)} = & e^{A_L^{(I,q)} t} \varepsilon_{x_H}^{(I,q)}(0) + \int_0^t e^{A_L^{(I,q)}(t-\tau)} \tilde{\gamma}_H^{(I,q)}(\tau) d\tau \\ & + \int_0^t e^{A_L^{(I,q)}(t-\tau)} \tilde{h}_H^{(I,q)}(\tau) d\tau + \int_0^t e^{A_L^{(I,q)}(t-\tau)} L^{(I,q)} d^{(I,q)}(\tau) d\tau \\ & + \int_0^t e^{A_L^{(I,q)}(t-\tau)} \eta^{(I)}(x^{(I)}(\tau), u^{(I)}(\tau), \zeta^{(I)}(\tau), C_z^{(I)} z^{(I)}(\tau), \tau) d\tau \end{aligned} \quad (4.19)$$

A bound on $|\varepsilon_{x_H}^{(I,q)}|$ satisfies

$$\begin{aligned} |\varepsilon_{x_H}^{(I,q)}| \leq & \left| e^{A_L^{(I,q)} t} \varepsilon_{x_H}^{(I,q)}(0) \right| + \int_0^t \left| e^{A_L^{(I,q)}(t-\tau)} L^{(I,q)} \right| |d^{(I,q)}(\tau)| d\tau \\ & + \int_0^t \left| e^{A_L^{(I,q)}(t-\tau)} \right| \left(\left| \tilde{\gamma}_H^{(I,q)}(\tau) \right| + \left| \tilde{h}_H^{(I,q)}(\tau) \right| \right. \\ & \left. + \left| \eta^{(I)}(x^{(I)}(\tau), u^{(I)}(\tau), \zeta^{(I)}(\tau), C_z^{(I)} z^{(I)}(\tau), \tau) \right| \right) d\tau \end{aligned} \quad (4.20)$$

Taking into account the conditions of Theorem 4.1, the bound on

$\left| \varepsilon_{x_H}^{(I,q)}(t) \right|$ satisfies:

$$\begin{aligned} \left| \varepsilon_{x_H}^{(I,q)}(t) \right| &\leq \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \int_0^t \rho_d^{(I,q)} e^{-\xi_d^{(I,q)}(t-\tau)} \bar{d}^{(I,q)} d\tau \\ &\quad + \int_0^t \Phi^{(I,q)}(t-\tau) \left(\left| \tilde{\gamma}_H^{(I,q)}(\tau) \right| + \left| \tilde{h}_H^{(I,q)}(\tau) \right| \right. \\ &\quad \left. + \left| \eta^{(I)}(x^{(I)}(\tau), u^{(I)}(\tau), \zeta^{(I)}(\tau), C_z^{(I)} z^{(I)}(\tau), \tau) \right| \right) d\tau, \end{aligned} \quad (4.21)$$

Given Assumptions 1-4, we have

$$\left| \tilde{\gamma}_H^{(I,q)} \right| \leq \lambda_{\gamma_I} \left| \varepsilon_{x_H}^{(I,q)} \right|, \quad (4.22)$$

$$\left| \tilde{h}_H^{(I,q)} \right| \leq \lambda_{h_I} \left\| \begin{bmatrix} \varepsilon_{x_H}^{(I,q)} \\ C_z^{(I)} z^{(I)} - y_{z_H}^{(I)} \\ \zeta^{(I)} - \zeta_r^{(I)} \end{bmatrix} \right\| \leq \lambda_{h_I} \left(\left| \varepsilon_{x_H}^{(I,q)} \right| + \bar{\zeta}^{(I)} + \bar{d}_z^{(I)} \right). \quad (4.23)$$

By setting

$$\bar{\eta}_{\Delta}^{(I,q)} = \bar{\eta}^{(I)}(x^{(I)}, u^{(I)}, \zeta^{(I)}, C_z^{(I)} z^{(I)}, t) - \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_{z_H}^{(I)}, t) \quad (4.24)$$

and taking into account Assumption 3, we obtain

$$\begin{aligned} \left| \eta^{(I)}(x^{(I)}, u^{(I)}, \zeta^{(I)}, C_z^{(I)} z^{(I)}, t) \right| &\leq \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_{z_H}^{(I)}, t) \\ &\quad + \left| \bar{\eta}_{\Delta}^{(I,q)} \right|, \end{aligned} \quad (4.25)$$

Given that $\bar{\eta}^{(I)}$ is locally Lipschitz, we have

$$\left| \bar{\eta}_{\Delta}^{(I,q)} \right| \leq \lambda_{\eta_I} \left(\left| \varepsilon_{x_H}^{(I,q)} \right| + \bar{\zeta}^{(I)} + \bar{d}_z^{(I)} \right) \quad (4.26)$$

Based on (4.22)-(4.26), the bound on $\left| \varepsilon_{x_H}^{(I,q)}(t) \right|$ can be computed as

$$\left| \varepsilon_{x_H}^{(I,q)}(t) \right| \leq E_H^{(I,q)}(t) + \int_0^t \Lambda_I \Phi^{(I,q)}(t-\tau) \left| \varepsilon_{x_H}^{(I,q)}(\tau) \right| d\tau \quad (4.27)$$

where $E_H^{(I,q)}$ is defined in (4.18). Applying the Bellman-Gronwall Lemma [Ioannou and Sun [1995]] results in (4.16). \square

Note that the conditions for the stability of the state error dynamics are obtained taking into account (i) the observability of the pair $(A^{(I)}, C^{(I,q)})$, (ii) the Lipschitz nonlinearities, represented by the Lipschitz constants λ_{γ_I} and λ_{h_I} , and (iii) the bound-function of the modeling uncertainty, represented by the Lipschitz constant λ_{η_I} .

Assuming that there are no modeling uncertainty and noise, i.e. $\eta^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t) = 0$, $d^{(I,q)}(t) = 0$ and $d_z^{(I)} = 0$, the bound on $|\varepsilon_{x_H}^{(I,q)}(t)|$ is expressed as

$$\begin{aligned} Z^{(I,q)}(t) = & \rho^{(I,q)} \left(\bar{x}^{(I,q)} - \frac{\lambda_{h_I} \bar{\zeta}^{(I)}}{\xi^{(I,q)} - \rho^{(I,q)} (\lambda_{\gamma_I} + \lambda_{h_I})} \right) \Phi_z^{(I,q)}(t) \\ & + \frac{\rho^{(I,q)} \lambda_{h_I} \bar{\zeta}^{(I)}}{\xi^{(I,q)} - \rho^{(I,q)} (\lambda_{\gamma_I} + \lambda_{h_I})}. \end{aligned} \quad (4.28)$$

where $\Phi_z^{(I,q)}(t) = e^{-(\xi^{(I,q)} - \rho^{(I,q)} (\lambda_{\gamma_I} + \lambda_{h_I}))t}$. Hence, as $t \rightarrow \infty$, $Z^{(I,q)}(t) \rightarrow \frac{\rho^{(I,q)} \lambda_{h_I} \bar{\zeta}^{(I)}}{\xi^{(I,q)} - \rho^{(I,q)} (\lambda_{\gamma_I} + \lambda_{h_I})}$, where $\bar{\zeta}^{(I)}$ is a bound on the error between some of the actual interconnections and their reference signals. Therefore, the bound $Z^{(I,q)}$ becomes proportional to the error bound $\bar{\zeta}^{(I)}$ as $t \rightarrow \infty$.

Remark 4.1. If the interconnection dynamics of subsystem $\Sigma^{(I)}$ is a function of measurable interconnection states only, i.e. $\zeta^{(I)} = 0$, the nonlinear observer can be structured as in (4.3) with $h^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, y_z^{(I)}) \equiv h^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, 0, y_z^{(I)})$. The stability of the state interconnection dynamics under healthy conditions are ensured under the conditions stated in Theorem 4.1, while the bound on the state estimation error is defined through (4.17) and (4.18) with $\bar{\zeta}^{(I)} = 0$. Thus, in the absence of modeling uncertainty and measurement noise, the bound on the state estimation error, $Z^{(I,q)}(t)$, is described by (4.28) with $\bar{\zeta}^{(I)} = 0$, implying that $Z^{(I,q)}(t) \rightarrow 0$ and $\varepsilon_{x_H}^{(I,q)}(t) \rightarrow 0$ as $t \rightarrow \infty$.

Linear Observer

Similarly to model (3.10), which is arisen from (2.1) after distinguishing the measured and unmeasured interconnection states, we formulate the

linear counterpart as follows:

$$\begin{aligned} \dot{x}^{(I)} = & \left(A^{(I)} + \Delta A^{(I)} \right) x^{(I)} + \left(B^{(I)} + \Delta B^{(I)} \right) u^{(I)} + \eta_L^{(I)} \\ & + \left(D_1^{(I)} + \Delta D_1^{(I)} \right) \left(C_z^{(I)} z^{(I)} \right) + \left(D_2^{(I)} + \Delta D_2^{(I)} \right) \zeta^{(I)}, \end{aligned} \quad (4.29)$$

where $D_1^{(I)}$ and $D_2^{(I)}$ are known system matrices of appropriate dimensions, while $\Delta D_1^{(I)}$ and $\Delta D_2^{(I)}$ represent bounded parametric uncertainty. When the dynamics of subsystem $\Sigma^{(I)}$ of the physical part $\mathcal{P}^{(I)}$ are described by (4.29), the estimation model associated with the monitoring module $\mathcal{M}^{(I,q)}$ can be generated by the following linear observer:

$$\begin{aligned} \mathcal{O}^{(I,q)} : \quad \dot{\hat{x}}^{(I,q)} = & A^{(I)} x^{(I)}(t) + B^{(I)} u^{(I)}(t) + D_1^{(I)} y_z^{(I)}(t) \\ & + D_2^{(I)} \zeta_r^{(I)}(t) + L^{(I,q)} \left(y^{(I,q)} - C^{(I,q)} \hat{x}^{(I,q)} \right). \end{aligned} \quad (4.30)$$

where $\hat{x}^{(I,q)} \in \mathbb{R}^{n_I}$ is the estimation of $x^{(I)}$ ($x^{(I)}$ is the state vector of subsystem $\Sigma^{(I)}$ described by (4.29)) based on the sensor measurements $y^{(I,q)}$ with initial conditions $\hat{x}^{(I,q)}(0) \in \mathcal{X}_I$.

Under healthy conditions, the estimation error dynamics $\varepsilon_{x_H}^{(I,q)} = x^{(I)} - \hat{x}_H^{(I,q)}$, where $\hat{x}_H^{(I,q)}$ is generated by (4.30) taking into account healthy local and transmitted sensor measurements, defined in (4.6) and (4.7), are described by (4.8) with

$$\tilde{\gamma}_H^{(I,q)} = 0, \quad (4.31)$$

$$\tilde{h}_H^{(I,q)} = D_1^{(I)} \left(C_z^{(I)} z^{(I)} - y_z^{(I)} \right) + D_2^{(I)} \left(\zeta^{(I)} - \zeta_r^{(I)} \right), \quad (4.32)$$

$$\begin{aligned} \eta^{(I)} = & \Delta A^{(I)} x^{(I)} + \Delta B^{(I)} u^{(I)}(t) + \Delta D_1^{(I)} \left(C_z^{(I)} z^{(I)}(t) \right) \\ & + \Delta D_2^{(I)} \zeta^{(I)}(t) + \eta_L^{(I)} \end{aligned} \quad (4.33)$$

Taking into account Assumption 5 and that $|\Delta D_1^{(I)}| \leq \overline{\Delta D}_1^{(I)}$, $|\Delta D_2^{(I)}| \leq \overline{\Delta D}_2^{(I)}$, the modeling uncertainty $\eta^{(I)}$ satisfies the following condition

$$\begin{aligned} \left| \eta^{(I)}(x^{(I)}, u^{(I)}, \zeta^{(I)}, C_z^{(I)} z^{(I)}, t) \right| \leq & \bar{\eta}_L^{(I)} + \overline{\Delta A}^{(I)} |x^{(I)}| + \overline{\Delta B}^{(I)} |u^{(I)}| \\ & + \overline{\Delta D}_1^{(I)} |C_z^{(I)} z^{(I)}| + \overline{\Delta D}_2^{(I)} |\zeta^{(I)}| \end{aligned} \quad (4.34)$$

Similarly to (4.25), we have

$$\begin{aligned} \left| \eta^{(I)}(x^{(I)}, u^{(I)}, \zeta^{(I)}, C_z^{(I)} z^{(I)}, t) \right| &\leq \overline{\Delta A}^{(I)} \left| \hat{x}_H^{(I,q)} \right| + \overline{\Delta B}^{(I)} \left| u^{(I)} \right| \\ &\quad + \overline{\Delta D_1}^{(I)} \left| y_z^{(I)} \right| + \overline{\Delta D_2}^{(I)} \left| \zeta_r^{(I)} \right| + \bar{\eta}_L^{(I)} \\ &\quad + \left| \bar{\eta}_\Delta^{(I,q)} \right| \end{aligned} \quad (4.35)$$

where

$$\left| \bar{\eta}_\Delta^{(I,q)} \right| = \overline{\Delta A}^{(I)} \left| \varepsilon_{x_H}^{(I,q)} \right| + \overline{\Delta D_1}^{(I)} \bar{d}_z^{(I)} + \overline{\Delta D_2}^{(I)} \bar{\zeta}^{(I)} \quad (4.36)$$

Given that

$$\left| \tilde{h}_H^{(I,q)} \right| \leq \overline{\Delta D_1}^{(I)} \bar{d}_z^{(I)} + \overline{\Delta D_2}^{(I)} \bar{\zeta}^{(I)}, \quad (4.37)$$

and following the proof of Theorem 4.1, a bound on the state estimation error under healthy conditions satisfies (4.16) with

$$Z_H^{(I,q)}(t) = E_H^{(I,q)}(t) + \rho^{(I,q)} \overline{\Delta A}^{(I)} \int_0^t E_H^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau, \quad (4.38)$$

$$\nu^{(I,q)} = \xi^{(I,q)} - \rho^{(I,q)} \overline{\Delta A}^{(I)} \quad (4.39)$$

$$\begin{aligned} E_H^{(I,q)}(t) &= \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\ &\quad + \frac{2 \left(\overline{\Delta D_1}^{(I)} + \overline{\Delta D_2}^{(I)} \right) (\bar{d}_z^{(I)} + \bar{\zeta}^{(I)}) + \bar{\eta}_L^{(I)}}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right) \\ &\quad + \int_0^t \Phi^{(I,q)}(t-\tau) \left(\overline{\Delta A}^{(I)} \left| \hat{x}_H^{(I,q)} \right| + \overline{\Delta B}^{(I)} \left| u^{(I)} \right| \right) d\tau \\ &\quad + \int_0^t \Phi^{(I,q)}(t-\tau) \left(\overline{\Delta D_1}^{(I)} \left| y_z^{(I)} \right| + \overline{\Delta D_2}^{(I)} \left| \zeta_r^{(I)} \right| \right) d\tau. \end{aligned} \quad (4.40)$$

where $\Phi^{(I,q)}$ defined in (4.13).

In the linear case, the stability of the distributed estimation error dynamics under healthy conditions is analyzed in the following theorem.

Theorem 4.2. Suppose that the observer gain $L^{(I,q)}$ is chosen such that: (a) the matrix $A_L^{(I,q)} = A^{(I)} - L^{(I,q)} C^{(I,q)}$ is stable, and (b) there exist positive constants $\rho^{(I,q)}$, $\xi^{(I,q)}$ such that $|e^{A_L^{(I,q)} t}| \leq \rho^{(I,q)} e^{-\xi^{(I,q)} t}$

and $\xi^{(I,q)} > \overline{\Delta A}^{(I)} \rho^{(I,q)}$, where $\overline{\Delta A}^{(I)}$ is defined in (2.27); then the state estimation error under healthy conditions, $\varepsilon_{x_H}^{(I,q)}(t)$, is uniformly bounded and satisfies (4.16) with (4.38) and (4.40).

Proof. The proof is similar to the proof of Theorem 4.1. \square

Note that in the linear case, the conditions for the stability of the state error dynamics are obtained taking into account the observability of the pair $(A^{(I)}, C^{(I,q)})$ and the upper bound on the parametric uncertainty $\Delta A^{(I)}$ related to the input matrix $A^{(I)}$.

4.1.2 Decentralized Observer-based Estimation Model

In a decentralized deployment of the monitoring agents $\mathcal{M}^{(I)}$ (see Section 3.2), the estimation model associated with the monitoring module $\mathcal{M}^{(I,q)}$ is generated by the nonlinear observer, described by

$$\begin{aligned} \mathcal{O}^{(I,q)} : \quad \dot{\hat{x}}^{(I,q)} = & A^{(I)} \hat{x}^{(I,q)} + \gamma^{(I)}(\hat{x}^{(I,q)}, u^{(I)}) + h^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, z_r^{(I)}) \\ & + L^{(I,q)} (y^{(I,q)} - C^{(I,q)} \hat{x}^{(I,q)}), \end{aligned} \quad (4.41)$$

where $\hat{x}^{(I,q)} \in \mathbb{R}^{n_I}$ is the estimation of $x^{(I)}$ ($x^{(I)}$ is the state vector of subsystem $\Sigma^{(I)}$ described by (2.1)) based on the sensor measurements $y^{(I,q)}$ with initial conditions $\hat{x}^{(I,q)}(0) \in \mathcal{X}_I$, $L^{(I,q)} \in \mathbb{R}^{n_I \times m_{I,q}}$ is the observer gain matrix, $z_r^{(I)}$ are reference signals of the unmeasured interconnection vector $z^{(I)}$ (see Assumption 1 in Section 2.1.1). It is noted that the generation of the estimation model $\hat{x}^{(I,q)}$ can be affected by the occurrence of faults in $\mathcal{S}^{(I,q)}$, since it is realized using measurements $y^{(I,q)}$.

Similarly to the distributed estimation procedure described in Section 4.1.1, the estimation error dynamics under healthy conditions $\varepsilon_{x_H}^{(I,q)} = x^{(I)} - \hat{x}_H^{(I,q)}$, where $\hat{x}_H^{(I,q)}$ is generated by (4.41) assuming that $y^{(I,q)} \equiv y_H^{(I,q)}$ define in (4.6), are described by

$$\begin{aligned} \dot{\varepsilon}_{x_H}^{(I,q)} = & A_L^{(I,q)} \varepsilon_{x_H}^{(I,q)} + \tilde{\gamma}_H^{(I,q)} + \tilde{h}_H^{(I,q)} - L^{(I,q)} d^{(I,q)} \\ & + \eta^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t) \end{aligned} \quad (4.42)$$

with $\tilde{\gamma}_H^{(I,q)}$ defined in (4.9) and

$$\tilde{h}_H^{(I,q)} = h^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}) - h^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, z_r^{(I)}). \quad (4.43)$$

The stability of the decentralized state estimation error dynamics is ensured taking into account the conditions in Theorem 4.1, while the state estimation error dynamics $\varepsilon_{x_H}^{(I,q)}$ satisfies (4.16) and (4.17) with

$$E_H^{(I,q)}(t) = \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \quad (4.44)$$

$$\begin{aligned} & + \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) d\tau \\ & + \frac{(\lambda_{h_I} + \lambda_{\eta_I}) \bar{z}^{(I)}}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right). \end{aligned} \quad (4.45)$$

where $\bar{z}^{(I)}$ is defined as a bound on the error between the actual interconnections and their reference signals i.e.

$$\left| z^{(I)}(t) - z_r^{(I)}(t) \right| \leq \bar{z}^{(I)}, \quad \forall t \geq 0 \quad (4.46)$$

In a decentralized approach, the conditions for a stable observer $\mathcal{O}^{(I,q)}$ are obtained, taking into account that possible faults occurring in neighboring subsystems may affect subsystem $\Sigma^{(I)}$ (i.e. $z^{(I)}$ can deviate from $z_r^{(I)}$ due to sensor faults in neighboring subsystems), but these effects are assumed to be unknown but bounded. In the state estimation error defined in (4.8), these effects are manifested through the terms $\tilde{h}^{(I,q)}$ defined in (4.10) and $\eta^{(I)}$, which are bounded; i.e. given Assumption 1-3,

$$\left| \tilde{h}_H^{(I,q)} \right| \leq \lambda_{h_I} \left(\left| \varepsilon_{x_H}^{(I,q)} \right| + \bar{z}^{(I)} \right) \quad (4.47)$$

$$\begin{aligned} \bar{\eta}^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t) & \leq \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, z_r^{(I)}, t) \\ & + \lambda_{\eta_I} \left(\left| \varepsilon_{x_H}^{(I,q)} \right| + \bar{z}^{(I)} \right) \end{aligned} \quad (4.48)$$

where $\bar{z}^{(I)}$ is defined in Theorem 4.1. This bound can be determined based on Assumption 1, which implies that $z^{(I)}(t) \in \mathcal{Z}_I$, i.e. $z^{(I)}(t)$ belongs to an invariant set before and after the occurrence of multiple sensor faults, and applying set membership manipulations. In the

case that the bounding condition is violated then this may lead to a false alarm. However, in the presence of “small” faults, which are the most difficult to detect, a robust controller will typically maintain a reasonably good tracking performance.

4.1.3 Centralized Observer-based Estimation Model

In a centralized architecture presented in Section 3.1, there is a single monitoring agent \mathcal{M} that consists of N monitoring modules $\mathcal{M}^{(q)}$. The estimation model associated with the module $\mathcal{M}^{(q)}$ is generated by the nonlinear observer, described by

$$\mathcal{O}^{(q)} : \dot{\hat{x}}^{(q)} = A\hat{x}^{(q)} + \gamma(\hat{x}^{(q)}, u) + L^{(q)} \left(y^{(q)} - C^{(q)}\hat{x}^{(q)} \right), \quad (4.49)$$

where $\hat{x}^{(q)} \in \mathbb{R}^n$ is the estimation of x (x is the state vector of system Σ described by (3.1)) based on the sensor measurements $y^{(q)}$ with initial conditions $\hat{x}^{(q)}(0) \in \mathcal{X}$, and $L^{(q)} \in \mathbb{R}^{n_I \times m_{I,q}}$ is the observer gain matrix. It is noted that the generation of the estimation model $\hat{x}^{(q)}$ can be affected by the occurrence of faults in $\mathcal{S}^{(q)}$ (without being affected by any sensor fault that occurs in $S \setminus \mathcal{S}^{(q)}$), since it is realized using the measurements $y^{(q)}$.

Under healthy conditions, the estimation error dynamics $\varepsilon_{x_H}^{(q)} = x - \hat{x}_H^{(q)}$, where $\hat{x}_H^{(q)}$ is generated by (4.49) taking into account healthy measurements, i.e.

$$y_H^{(q)}(t) = C^{(q)}x(t) + d^{(q)}(t) \quad (4.50)$$

and is described by

$$\dot{\varepsilon}_{x_H}^{(q)} = A_L^{(q)}\varepsilon_{x_H}^{(q)} + \tilde{\gamma}_H^{(q)} + \eta(x, u, t) - L^{(q)}d^{(q)} \quad (4.51)$$

$$\tilde{\gamma}_H^{(q)} = \gamma(x, u) - \gamma(\hat{x}_H^{(q)}, u) \quad (4.52)$$

$$A_L^{(q)} = A - L^{(q)}C^{(q)} \quad (4.53)$$

If the pair $(A, C^{(q)})$ is observable, then $L^{(q)}$ can be chosen such that $A_L^{(q)}$ is stable. Consequently, we can choose positive constants $\rho^{(q)}, \xi^{(q)}$ such that $|e^{A_L^{(q)}t}| \leq \rho^{(q)}e^{-\xi^{(q)}t}$, as well as $\rho_d^{(q)}, \xi_d^{(q)}$ such that $\left| e^{A_L^{(q)}t}L^{(q)} \right| \leq$

$\rho_d^{(q)} e^{-\xi_d^{(q)} t}$. Let us define the function

$$\Phi^{(q)}(t) = \rho^{(q)} e^{-\xi^{(q)} t} \quad (4.54)$$

and the constant

$$\nu^{(q)} = \xi^{(q)} - \rho^{(q)} \Lambda \quad (4.55)$$

$$\Lambda = \lambda_\gamma + \lambda_\eta \quad (4.56)$$

where the parameters λ_γ and λ_η denote the Lipschitz constants of γ and $\bar{\eta}$, respectively.

The stability of the centralized estimation error dynamics under healthy conditions is analyzed in the following theorem, taking into account that $\bar{d}^{(q)}$ is a noise bound for $d^{(q)}(t)$ such that $|d^{(q)}(t)| \leq \bar{d}^{(q)}$, and $\bar{x}^{(q)}$ is a bound on $|\varepsilon_{x_H}^{(q)}(0)|$.

Theorem 4.3. Suppose that the observer gain $L^{(q)}$ is chosen such that: (a) the matrix $A_L^{(q)} = A - L^{(q)}C^{(q)}$ is stable, and (b) the positive constants $\rho^{(q)}$, $\xi^{(q)}$ satisfy $\xi^{(q)} > \Lambda\rho^{(q)}$; then the state estimation error under healthy conditions, $\varepsilon_{x_H}^{(q)}(t)$, is uniformly bounded and satisfies

$$|\varepsilon_{x_H}^{(q)}(t)| \leq Z_H^{(q)}(t) \quad (4.57)$$

where

$$Z_H^{(q)}(t) = E_H^{(q)}(t) + \rho^{(q)} \Lambda \int_0^t E_H^{(q)}(\tau) e^{-\nu^{(q)}(t-\tau)} d\tau, \quad (4.58)$$

$$\begin{aligned} E_H^{(q)}(t) = & \Phi^{(q)}(t) \bar{x}^{(q)} + \frac{\rho_d^{(q)} \bar{d}^{(q)}}{\xi_d^{(q)}} \left(1 - e^{-\xi_d^{(q)} t} \right) \\ & + \int_0^t \Phi^{(q)}(t-\tau) \bar{\eta}(\hat{x}_H^{(q)}(\tau), u(\tau), \tau) d\tau \end{aligned} \quad (4.59)$$

Proof. The proof is similar to the proof of Theorem 4.1. \square

Note that the conditions for the stability of the state error dynamics are obtained taking into account (i) the observability of the pair $(A, C^{(q)})$, (ii) the Lipschitz nonlinearity, represented by the Lipschitz constant λ_γ , and (iii) the bound-function of the modeling uncertainty, represented by the Lipschitz constant λ_η . Assuming that there is no

modeling uncertainty and measurement noise, the bound on the state estimation error is defined as $Z^{(q)}(t) = \rho^{(q)} \bar{x}^{(I,q)} e^{-(\xi^{(q)} - \rho^{(q)} \lambda_\gamma)t}$. Therefore, $Z^{(q)}(t) \rightarrow 0$ and $\varepsilon_{x_H}^{(q)}(t) \rightarrow 0$ as $t \rightarrow \infty$.

If the nonlinear dynamics represented by γ and the modeling uncertainty η are functions of a linear combination of states that are measured by the local sensor set $\mathcal{S}^{(q)}$, i.e. $\gamma(C^{(q)}x, u)$ and $\eta(C^{(q)}x, u, t)$, the nonlinear observer can be structured as

$$\mathcal{O}^{(q)} : \dot{\hat{x}}^{(q)} = A\hat{x}^{(q)} + \gamma(y^{(q)}, u) + L^{(q)}(y^{(q)} - C^{(q)}\hat{x}^{(q)}), \quad (4.60)$$

Then, the state estimation error under healthy conditions is described by

$$\dot{\varepsilon}_{x_H}^{(q)} = A_L^{(q)} \varepsilon_{x_H}^{(q)} + \tilde{\gamma}_H^{(q)} + \eta(C^{(q)}x, u, t) - L^{(q)}d^{(q)} \quad (4.61)$$

$$\tilde{\gamma}_H^{(q)} = \gamma(x, u) - \gamma(C^{(q)}x, u). \quad (4.62)$$

The solution of (4.61) satisfies the following equation:

$$\begin{aligned} \varepsilon_{x_H}^{(q)}(t) = & e^{A_L^{(q)}t} \varepsilon_{x_H}^{(q)}(0) + \int_0^t e^{A_L^{(q)}(t-\tau)} \left(\tilde{\gamma}_H^{(q)}(\tau) + \eta(C^{(q)}x(\tau), u(\tau), \tau) \right. \\ & \left. - L^{(q)}d^{(q)}(\tau) \right) d\tau \end{aligned} \quad (4.63)$$

A bound on $|\varepsilon_{x_H}^{(q)}(t)|$ satisfying (4.57) is determined as

$$\begin{aligned} Z_H^{(q)}(t) = & \frac{\rho_d^{(q)} \bar{d}^{(q)}}{\xi_d^{(q)}} \left(1 - e^{-\xi_d^{(q)}t} \right) + \frac{(\lambda_\gamma + \lambda_\eta) \bar{d}^{(q)}}{\xi^{(q)}} \left(\rho^{(q)} - \Phi^{(q)}(t) \right) \\ & + \Phi^{(q)}(t) \bar{x} + \int_0^t \Phi^{(q)}(t-\tau) \bar{\eta}(y^{(q)}(\tau), u(\tau), \tau) d\tau \end{aligned} \quad (4.64)$$

where $\Phi^{(q)}(t) = \rho^{(q)} e^{-\xi^{(q)}t}$. It is noted that the stability of the state estimation dynamics is ensured if the observer gain $L^{(q)}$ is selected such that the matrix $A_L^{(q)} = A - L^{(q)}C^{(q)}$ is stable, irrespective of the Lipschitz nonlinearities represented by the constants λ_γ and λ_η .

4.2 Computation of Adaptive Thresholds

The state estimation error under healthy conditions $\varepsilon_{x_H}^{(I,q)}(t)$ (correspondingly for $\varepsilon_{x_H}^{(q)}(t)$ in a centralized architecture) may represent the

effects of system disturbances, local and transmitted measurement noise and unknown initial conditions on the estimation model associated with $\mathcal{M}^{(I,q)}$. The boundedness of $\varepsilon_{x_H}^{(I,q)}(t)$ is exploited in order to design the adaptive thresholds. Particularly, the j -th adaptive threshold, denoted by $\bar{\varepsilon}_{y_j}^{(I,q)}(t)$ with $j \in \mathcal{J}^{(I,q)}$, $I \in \{1, \dots, M\}$, $q \in \{1, \dots, N_I\}$, is designed to bound the residual under healthy conditions, defined as

$$\varepsilon_{y_{Hj}}^{(I,q)} = y_{Hj}^{(I)} - C_j^{(I)} \hat{x}_H^{(I,q)}, \quad j \in \mathcal{J}^{(I,q)}, \quad (4.65)$$

where $\mathcal{J}^{(I,q)}$ is given in (3.9), $y_{Hj}^{(I)}$ is j -th component of $y_H^{(I)}$ defined in (4.6) and $\hat{x}_H^{(I,q)}$ is the observer-based estimation model under healthy conditions associated with the monitoring module $\mathcal{M}^{(I,q)}$. In a centralized scheme ($M = 1$), i.e. the j -th adaptive threshold is designed to bound the residual defined by

$$\varepsilon_{y_{Hj}}^{(q)} = y_{Hj} - C_j \hat{x}_H^{(q)}, \quad j \in \mathcal{J}^{(q)}, \quad (4.66)$$

where $\mathcal{J}^{(q)}$ is defined in (3.3), y_{Hj} is the j -th component of y_H given in (4.50) and $\hat{x}_H^{(q)}$ is the estimation model under healthy conditions, described by (4.51). The computation of $\bar{\varepsilon}_{y_j}^{(I,q)}$ (similarly for $\bar{\varepsilon}_{y_j}^{(q)}$) is carried out by first expressing the residual $\varepsilon_{y_{Hj}}^{(I,q)}$ as a function of the state estimation error under healthy conditions $\varepsilon_{x_H}^{(I,q)}(t)$; i.e.,

$$\varepsilon_{y_{Hj}}^{(I,q)} = C_j^{(I)} \varepsilon_{x_H}^{(I,q)} + d_j^{(I)}, \quad j \in \mathcal{J}^{(I,q)}, \quad (4.67)$$

where $\varepsilon_{x_H}^{(I,q)}$ varies with respect to the sensor fault diagnosis architecture (see Section 4.1), while $C_j^{(I)}$ and $d_j^{(I)}$ are respectively the output matrix and noise vector, associated with the sensor $\mathcal{S}^{(I)}\{j\}$ defined in (2.3). Following a centralized approach we have

$$\varepsilon_{y_{Hj}}^{(q)} = C_j \varepsilon_{x_H}^{(q)} + d_j, \quad j \in \mathcal{J}^{(q)}, \quad (4.68)$$

The second step is to find a bound on the magnitude of the residual taking into account Assumptions 1-4 presented in Section 2, based on the following conditions: in a non-centralized diagnostic scheme,

$$\left| \varepsilon_{y_{Hj}}^{(I,q)} \right| \leq \left| C_j^{(I)} \varepsilon_{x_H}^{(I,q)} \right| + \bar{d}_j^{(I)} \quad (4.69)$$

for $M > 1$, while in a centralized diagnostic scheme

$$\left| \varepsilon_{y_{Hj}}^{(q)} \right| \leq \left| C_j \varepsilon_{x_H}^{(q)} \right| + \bar{d}_j. \quad (4.70)$$

4.2.1 Distributed Adaptive Thresholds

Taking into account the state estimation error under healthy conditions $\varepsilon_{x_H}^{(I,q)}$ defined through (4.8)-(4.10) and the solution (4.19), a bound on $\left| C_j^{(I)} \varepsilon_{x_H}^{(I,q)} \right|$ can be computed as

$$\begin{aligned} \left| C_j^{(I)} \varepsilon_{x_H}^{(I,q)} \right| &\leq \left| C_j^{(I)} e^{A_L^{(I,q)} t} \right| \left| \varepsilon_{x_H}^{(I,q)}(0) \right| \\ &\quad + \int_0^t \left| C_j^{(I)} e^{A_L^{(I,q)}(t-\tau)} L^{(I,q)} \right| \left| d^{(I,q)}(\tau) \right| d\tau \\ &\quad + \int_0^t \left| C_j^{(I)} e^{A_L^{(I,q)}(t-\tau)} \right| \left(\left| \tilde{\gamma}_H^{(I,q)}(\tau) \right| + \left| \tilde{h}_H^{(I,q)}(\tau) \right| \right. \\ &\quad \left. + \left| \eta^{(I)}(x^{(I)}(\tau), u^{(I)}(\tau), \zeta^{(I)}(\tau), C_z^{(I)} z^{(I)}(\tau), \tau) \right| \right) d\tau \end{aligned} \quad (4.71)$$

By choosing positive constants $\alpha_j^{(I,q)}$, $\zeta_j^{(I,q)}$ and $\alpha_{d_j}^{(I,q)}$, $\zeta_{d_j}^{(I,q)}$ such that

$$\left| C_j^{(I,q)} e^{A_L^{(I,q)} t} \right| \leq \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)} t} \quad (4.72)$$

$$\left| C_j^{(I)} e^{A_L^{(I,q)} t} L^{(I,q)} \right| \leq \alpha_{d_j}^{(I,q)} e^{-\zeta_{d_j}^{(I,q)} t}, \quad (4.73)$$

it yields

$$\begin{aligned} \left| C_j^{(I)} \varepsilon_{x_H}^{(I,q)} \right| &\leq \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)} t} \left| \varepsilon_{x_H}^{(I,q)}(0) \right| + \int_0^t \alpha_{d_j}^{(I,q)} e^{-\zeta_{d_j}^{(I,q)}(t-\tau)} \left| d^{(I,q)}(\tau) \right| d\tau \\ &\quad + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\left| \tilde{\gamma}_H^{(I,q)}(\tau) \right| + \left| \tilde{h}_H^{(I,q)}(\tau) \right| \right. \\ &\quad \left. + \left| \eta^{(I)}(x^{(I)}(\tau), u^{(I)}(\tau), \zeta^{(I)}(\tau), C_z^{(I)} z^{(I)}(\tau), \tau) \right| \right) d\tau \end{aligned} \quad (4.74)$$

Given (4.22) and (4.23), we obtain

$$\begin{aligned}
\left| C_j^{(I)} \varepsilon_{x_H}^{(I,q)} \right| &\leq \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)} t} \left| \varepsilon_{x_H}^{(I,q)}(0) \right| + \int_0^t \alpha_{d_j}^{(I,q)} e^{-\zeta_{d_j}^{(I,q)}(t-\tau)} \left| d^{(I,q)}(\tau) \right| d\tau \\
&\quad + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left((\lambda_{\eta_I} + \lambda_{h_I}) \left(\bar{\zeta}^{(I)} + \bar{d}_z^{(I)} \right) \right) d\tau \\
&\quad + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I \left| \varepsilon_{x_H}^{(I,q)}(\tau) \right| \right. \\
&\quad \left. + \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) \right) d\tau \quad (4.75)
\end{aligned}$$

Using (4.16) and combining inequalities (4.69) and (4.75) results in

$$\left| \varepsilon_{y_{Hj}}^{(I,q)} \right| \leq \bar{\varepsilon}_{y_{Hj}}^{(I,q)}, \quad (4.76)$$

where $\bar{\varepsilon}_{y_{Hj}}^{(I,q)}$ denotes the adaptive threshold under healthy conditions (i.e. there is no sensor fault in the local sensor set $\mathcal{S}^{(I,q)}$ and the transmitted sensor information $\mathcal{S}_z^{(I)}$), which is defined as

$$\begin{aligned}
\bar{\varepsilon}_{y_{Hj}}^{(I,q)} &= Y_j^{(I,q)}(t) + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I Z_H^{(I,q)}(\tau) \right. \\
&\quad \left. + \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) \right) d\tau \quad (4.77) \\
Y_j^{(I,q)}(t) &= \alpha_j^{(I,q)} \bar{x}^{(I,q)} e^{-\zeta_j^{(I,q)} t} + \frac{\alpha_{d_j}^{(I,q)} \bar{d}^{(I,q)}}{\zeta_{d_j}^{(I,q)}} \left(1 - e^{-\zeta_{d_j}^{(I,q)} t} \right) \\
&\quad + \frac{\alpha_j^{(I,q)} (\lambda_{h_I} + \lambda_{\eta_I}) \left(\bar{\zeta}^{(I)} + \bar{d}_z^{(I)} \right)}{\zeta_j^{(I,q)}} \left(1 - e^{-\zeta_j^{(I,q)} t} \right) + \bar{d}_j^{(I)}, \quad (4.78)
\end{aligned}$$

where $Z_H^{(I,q)}$ is determined through (4.17) with $E_H^{(I,q)}$ defined in (4.18) and $\hat{x}_H^{(I,q)}$ and $y_{z_H}^{(I)}$ defined in (4.4) and (4.7), respectively.

The mathematical analysis realized through (4.71)-(4.76) corresponds to the design methodology for computing the adaptive thresholds that ensure the robustness of the monitoring module $\mathcal{M}^{(I,q)}$. At every time instant, the j -th adaptive threshold associated with $\mathcal{M}^{(I,q)}$

is computed as

$$\begin{aligned} \bar{\varepsilon}_{y_j}^{(I,q)}(t) = & Y_j^{(I,q)}(t) + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I Z^{(I,q)}(\tau) \right. \\ & \left. + \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_z^{(I)}(\tau), \tau) \right) d\tau, \end{aligned} \quad (4.79)$$

where

$$\begin{aligned} Z^{(I,q)}(t) = & E^{(I,q)}(t) + \rho^{(I,q)} \Lambda_I \int_0^t E^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau, \\ E^{(I,q)}(t) = & \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\ & + \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_z^{(I)}(\tau), \tau) d\tau \\ & + \frac{(\lambda_{h_I} + \lambda_{\eta_I})(\bar{d}_z^{(I)} + \bar{\zeta}^{(I)})}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right). \end{aligned} \quad (4.81)$$

with $\Phi^{(I,q)}(t) = \rho^{(I,q)} e^{-\xi^{(I,q)} t}$ and $\nu^{(I,q)} = \xi^{(I,q)} - \rho^{(I,q)} \Lambda_I$. It is noted that $\nu^{(I,q)} = \xi^{(I,q)} - \rho^{(I,q)} \Lambda_I$ is positive according to Theorem 4.1, making the adaptive threshold finite for all t . Moreover, the adaptive thresholds can be affected by the occurrence of faults in $\mathcal{S}^{(I,q)}$ and $\mathcal{S}_z^{(I)}$, since their measurements ($y^{(I,q)}$ and $y_z^{(I)}$) are used to generate the estimation model $\hat{x}^{(I,q)}$ and to compute the bound on the modeling uncertainty $\bar{\eta}^{(I)}$.

The j -th adaptive threshold can be implemented using linear filters [Reppa et al. [2014b]]; i.e.,

$$\begin{aligned} \bar{\varepsilon}_{y_j}^{(I,q)}(t) = & W(s) \left[\bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(t), u^{(I)}(t), \zeta_r^{(I)}(t), y_{z_H}^{(I)}(t), t) + \Lambda_I Z^{(I,q)}(t) \right] \\ & + Y_j^{(I,q)}(t), \end{aligned} \quad (4.82)$$

$$W(s) = \frac{\alpha_j^{(I,q)}}{s + \zeta_j^{(I,q)}}. \quad (4.83)$$

The notation $W(s)[z(t)]$ denotes the output of the filter $W(s)$ defined in Laplace domain with $z(t)$ as input, for any signal $z(t)$. Similarly, the signals $Z^{(I,q)}$ and $E^{(I,q)}$ can be implemented using linear filtering techniques.

4.2.2 Decentralized Adaptive Thresholds

Following a decentralized approach, the computation of adaptive thresholds is realized based on the design methodology shown in Section 4.2.1. Thus, the adaptive threshold associated with $\mathcal{M}^{(I,q)}$ is computed at every time instant according to:

$$\begin{aligned} \bar{\varepsilon}_{y_j}^{(I,q)}(t) = & Y_j^{(I,q)}(t) + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I Z^{(I,q)}(\tau) \right. \\ & \left. + \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) \right) d\tau \end{aligned} \quad (4.84)$$

where

$$\begin{aligned} Y_j^{(I,q)}(t) = & \alpha_j^{(I,q)} \bar{x}^{(I)} e^{-\zeta_j^{(I,q)} t} + \frac{\alpha_{d_j}^{(I,q)} \bar{d}^{(I,q)}}{\zeta_{d_j}^{(I,q)}} \left(1 - e^{-\zeta_{d_j}^{(I,q)} t} \right) \\ & + \frac{\alpha_j^{(I,q)} (\lambda_{h_I} + \lambda_{\eta_I}) \bar{z}^{(I)}}{\zeta_j^{(I,q)}} \left(1 - e^{-\zeta_j^{(I,q)} t} \right) + \bar{d}_j^{(I)}. \end{aligned} \quad (4.85)$$

where $Z^{(I,q)}$ is defined in (4.80) with

$$\begin{aligned} E^{(I,q)}(t) = & \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\ & + \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) d\tau \\ & + \frac{(\lambda_{h_I} + \lambda_{\eta_I}) \bar{z}^{(I)}}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right). \end{aligned} \quad (4.87)$$

where $\bar{z}^{(I)}$ is defined in (4.46).

It is noted that the adaptive thresholds can be affected by the occurrence of faults in $\mathcal{S}^{(I,q)}$ (without being affected by any sensor fault in the neighboring subsystems), since its measurements ($y^{(I,q)}$) are used to generate the estimation model $\hat{x}^{(I,q)}$ and compute the bound on the modeling uncertainty $\bar{\eta}^{(I)}$.

4.2.3 Centralized Adaptive Thresholds

Following a centralized approach, the computation of adaptive thresholds is realized based on the state estimation error defined through

(4.51)-(4.53) and the design methodology presented in Section 4.2.1. Thus, the j -th adaptive threshold is computed at every time instant as:

$$\begin{aligned} \bar{\varepsilon}_{y_j}^{(q)}(t) = & \int_0^t \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-\tau)} \left(\Lambda Z^{(q)}(\tau) + \bar{\eta}(\hat{x}^{(q)}(\tau), u^{(I)}(\tau), \tau) \right) d\tau \\ & + Y_j^{(q)}(t) \end{aligned} \quad (4.88)$$

where

$$Y_j^{(q)}(t) = \alpha_j^{(q)} \bar{x}^{(q)} e^{-\zeta_j^{(q)} t} + \frac{\alpha_{d_j}^{(q)} \bar{d}^{(q)}}{\zeta_{d_j}^{(q)}} \left(1 - e^{-\zeta_{d_j}^{(q)} t} \right) + \bar{d}_j \quad (4.89)$$

$$Z^{(q)}(t) = E^{(q)}(t) + \rho^{(q)} \Lambda \int_0^t E^{(q)}(\tau) e^{-\nu^{(q)}(t-\tau)} d\tau, \quad (4.90)$$

$$\begin{aligned} E^{(q)}(t) = & \Phi^{(q)}(t) \bar{x}^{(q)} + \frac{\rho_d^{(q)} \bar{d}^{(q)}}{\xi_d^{(q)}} \left(1 - e^{-\xi_d^{(q)} t} \right) \\ & + \int_0^t \Phi^{(q)}(t-\tau) \bar{\eta}(\hat{x}^{(q)}(\tau), u(\tau), \tau) d\tau \end{aligned} \quad (4.91)$$

It is noted that the adaptive threshold is affected by the occurrence of sensor faults in $\mathcal{S}^{(q)}$ (without being affected by any sensor fault that occurs in $S \setminus \mathcal{S}^{(q)}$), since its measurements ($y^{(q)}$) are used to generate the estimation model $\hat{x}^{(q)}$ and compute the bound on the modeling uncertainty $\bar{\eta}$.

4.3 Sensor Fault Detection Decision Logic

The primary goal of the monitoring module $\mathcal{M}^{(I,q)}$ (correspondingly for $\mathcal{M}^{(q)}$ in a centralized architecture) is to infer about the presence of sensor faults in the local sensor set $\mathcal{S}^{(I,q)}$ (correspondingly for $\mathcal{S}^{(q)}$ in a centralized architecture). If the monitoring modules have been deployed in a distributed sensor FDI architecture, then the decision of the module $\mathcal{M}^{(I,q)}$ can be affected by the sensor faults in the transmitted sensor information $y_z^{(I)}$.

For any sensor fault diagnosis architecture presented in Section 3, the decision logic implemented in $\mathcal{M}^{(I,q)}$ relies on checking the satisfaction of a set of analytical redundancy relations (ARRs) of residuals

and adaptive thresholds presented in Sections 4.1 and 4.2 [Blanke et al. [2016], Cordier et al. [2004], Puig et al. [2005]]. In this tutorial, the ARRs are dynamical constraints, formulated using the residuals and adaptive thresholds, which are designed based on input and output data and the estimated state vector.

The set of ARRs, based on which the module $\mathcal{M}^{(I,q)}$ with $I \in \{1, \dots, M\}$, $q \in \{1, \dots, N_I\}$, obtains a decision, is defined as

$$\mathcal{E}^{(I,q)} = \bigcup_{j \in \mathcal{J}^{(I,q)}} \mathcal{E}_j^{(I,q)}, \quad (4.92)$$

where for all $j \in \mathcal{J}^{(I,q)}$

$$\mathcal{E}_j^{(I,q)} : \quad \left| \varepsilon_{y_j}^{(I,q)}(t) \right| - \bar{\varepsilon}_{y_j}^{(I,q)}(t) \leq 0. \quad (4.93)$$

When inequality in (4.93) is true, it is inferred that the ARR $\mathcal{E}_j^{(I,q)}$ is satisfied, while the set $\mathcal{E}^{(I,q)}$ is said to be satisfied when $\mathcal{E}_j^{(I,q)}$ is satisfied for all $j \in \mathcal{J}^{(I,q)}$. Note that (4.92) and (4.93) are valid in a centralized approach with $I = M = 1$, where the set of ARRs is denoted as $\mathcal{E}^{(q)}$ and the j -th ARR is denoted as $\mathcal{E}_j^{(q)}$, $j \in \mathcal{J}^{(q)}$.

The output of $\mathcal{M}^{(I,q)}$ is a boolean function, defined as

$$D^{(I,q)}(t) = \begin{cases} 0, & \text{if } D_j^{(I,q)}(t) = 0, \forall j \in \mathcal{J}^{(I,q)} \\ 1, & \text{otherwise} \end{cases} \quad (4.94)$$

$$D_j^{(I,q)}(t) = \begin{cases} 0, & \text{if } t < T_j^{(I,q)} \\ 1, & \text{otherwise} \end{cases} \quad (4.95)$$

$$T_j^{(I,q)} = \min \left\{ t : \left| \varepsilon_{y_j}^{(I,q)}(t) \right| - \bar{\varepsilon}_{y_j}^{(I,q)}(t) > 0 \right\} \quad (4.96)$$

If $\mathcal{E}_j^{(I,q)}$ is always satisfied, then the detection time is defined as $T_j^{(I,q)} = \infty$. The time instant of sensor fault detection is defined as

$$T_{FD}^{(I,q)} = \min \left\{ T_j^{(I,q)} : j \in \mathcal{J}^{(I,q)} \right\} \quad (4.97)$$

The sensor fault detection decision logic relies on two basic properties of the set of ARRS $\mathcal{E}^{(I,q)}$ (correspondingly for $\mathcal{E}^{(q)}$ in a centralized architecture), that is the structured robustness and structured sensor fault sensitivity that depends on the sensor fault diagnosis architecture.

The structured robustness property implies that every set of ARRs $\mathcal{E}^{(I,q)}$ is insensitive to modeling uncertainties and noise affecting the measurements of the sensors involved in the generation of the set $\mathcal{E}^{(I,q)}$, thus prohibiting false alarms. Based on the structured sensor fault sensitivity property, every set of ARRs $\mathcal{E}^{(I,q)}$ is implicitly sensitive only to combinations of faults that may affect the sensors involved in the generation of the set $\mathcal{E}^{(I,q)}$ and not to sensor faults affecting sensors that are not involved in the set $\mathcal{E}^{(I,q)}$.

In the following subsections, the structured robustness and structured sensor fault sensitivity are analyzed in relation to the sensor fault diagnosis architecture.

Distributed Approach

Following a distributed sensor fault diagnosis architecture (see Section 3.3) and the methodology for computing the residuals and adaptive thresholds associated with the monitoring module $\mathcal{M}^{(I,q)}$ presented in Sections 4.1.1 and 4.2.1, the properties of the set of ARRs $\mathcal{E}^{(I,q)}$ are described in Lemma 4.4.

Lemma 4.4. Taking into account the conditions of Theorem 4.1, it is ensured that:

1. **Structured robustness:** If neither the local sensor set $\mathcal{S}^{(I,q)}$ nor the transmitted sensor information $y_z^{(I)}$ are affected by sensor faults, then the set of ARRs $\mathcal{E}^{(I,q)}$ is always satisfied.
2. **Structured sensor fault sensitivity:** If there is a time instant at which $\mathcal{E}^{(I,q)}$ is not satisfied, then the occurrence of at least one sensor fault in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$ is guaranteed.

Proof. (1) If both $\mathcal{S}^{(I,q)}$ and $\mathcal{S}_z^{(I)}$ are healthy, the j -th residual is described by (4.65); i.e., $\varepsilon_{y_j}^{(I,q)} = \varepsilon_{y_{Hj}}^{(I,q)}$, and the j -th adaptive threshold is defined by (4.77), i.e., $\bar{\varepsilon}_{y_j}^{(I,q)} = \bar{\varepsilon}_{y_{Hj}}^{(I,q)}$, implying that (4.76) is valid and consequently $\mathcal{E}^{(I,q)}$ is satisfied.

(2) The second part of Lemma 4.4 can be proved by *reductio ad absurdum*. Suppose that no sensor fault has occurred in $\mathcal{S}^{(I,q)}$ and $\mathcal{S}_z^{(I)}$. Then, $\mathcal{E}^{(I,q)}$ is satisfied, according to part (1) of Lemma 4.4. This contradicts our assumption of part (2). \square

The inference mechanism based these properties and the value of the boolean decision function, is realized in an ‘exoneration’ framework described by the following assertion:

Exoneration framework: Given a set of observations, the sensors that are used for the generation of the set of ARRS $\mathcal{E}^{(I,q)}$ (correspondingly for $\mathcal{E}^{(q)}$ in a centralized architecture) necessarily reveal their faulty operation by provoking the violation of the set $\mathcal{E}^{(I,q)}$, or equivalently, all sensors that are used for the generation of the set of ARRS $\mathcal{E}^{(I,q)}$ are exonerated, i.e. are considered as functioning properly, if the set $\mathcal{E}^{(I,q)}$ is satisfied.

When $D^{(I,q)} = 1$, it is guaranteed that sensor faults have occurred in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$ and it is not a false alarm, since the set $\mathcal{E}^{(I,q)}$ are structurally robust. As long as $D^{(I,q)} = 0$, either there is no fault occurrence in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$, or sensor faults have occurred but they have not been detected yet. Thus, when $D^{(I,q)} = 0$, we ‘exonerate’ the set $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$ from being faulty and we characterize it as *non-faulty*. However, it is not guaranteed that have not occurred in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$. Ensuring that the local sensor set is not affected by faults is equivalent to ensuring that there is no delay in detection (we detect faults at the ‘same’ time that they occur [Seron et al. [2012]]). Assuming modeling uncertainties, the delay in detection depends on the sensitivity of the set $\mathcal{E}^{(q)}$ to the faults in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$. This can be done during the design and after analyzing the performance of the modules with respect to fault detectability as presented in Chapter 6.

Due to the possible overlapping between some local sensor sets $\mathcal{S}^{(I,q)}$ and possible propagation of sensor faults transmitted through the sensor information $y_z^{(I)}$, when $D^{(I,q)}(t) = 1$ the decision of the module $\mathcal{M}^{(I,q)}$ is sent to the aggregation module $\mathcal{A}^{(I)}$ for isolating multiple sensor faults in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$.

Decentralized Approach

In a decentralized architecture (see Section 3.2), where the residuals and adaptive thresholds are computed as described in Sections 4.1.2 and 4.2.2, the properties of the set of ARRs $\mathcal{E}^{(I,q)}$ are described in Lemma 4.5.

Lemma 4.5. Taking into account the conditions of Theorem 4.1, it is ensured that

1. **Structured robustness:** If there is no sensor fault in the local sensor set $\mathcal{S}^{(I,q)}$, the set of ARR_s $\mathcal{E}^{(I,q)}$ is always satisfied.
2. **Structured sensor fault sensitivity:** If there is a time instant at which $\mathcal{E}^{(I,q)}$ is not satisfied, then the occurrence of at least one sensor fault in the local sensor set $\mathcal{S}^{(I,q)}$ is guaranteed.

Proof. The proof is similar to the proof of Lemma 4.4. □

When $D^{(I,q)}(t) = 1$, the module $\mathcal{M}^{(I,q)}$ ensures that the set $\mathcal{S}^{(I,q)}$ is faulty, based on the structured sensor fault sensitivity (Lemma 4.5). As long as $D^{(I,q)}(t) = 0$ either there is no sensor fault in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$ or sensor faults have occurred, but have not been detected by the module $\mathcal{M}^{(I,q)}$ until the time $T_{FD}^{(I,q)}$. However, within an exoneration framework (see Section 4.3) the module $\mathcal{M}^{(I,q)}$ characterizes the sensor set $\mathcal{S}^{(I,q)}$ as *non-faulty*. Due to the possible overlapping between some local sensor sets $\mathcal{S}^{(I,q)}$, when $D^{(I,q)}(t) = 1$ the decision of the module $\mathcal{M}^{(I,q)}$ is collected by the aggregation module $\mathcal{A}^{(I)}$ and processed in combination with the decisions of the other modules for isolating multiple sensor faults in $\mathcal{S}^{(I,q)}$.

Centralized Approach

In a centralized architecture presented in Section 3.1, where the residuals and adaptive thresholds are computed as described in Sections 4.1.3 and 4.2.3, the properties of the set of ARR_s $\mathcal{E}^{(q)}$ are described in a similar way as in Lemma 4.5 taking into account that $\mathcal{E}^{(q)} \equiv \mathcal{E}^{(I,q)}$ and $\mathcal{S}^{(q)} \equiv \mathcal{S}^{(I,q)}$.

4.4 Other Established Fault Detection Schemes

This section presents several established observer schemes used in robust fault detection methods for linear, nonlinear and interconnected systems. An overview of their design along with the underlying assumptions are provided, followed by a brief discussion on their common

characteristics and differences with respect to the observer schemes proposed in this tutorial.

4.4.1 Unknown Input Observer-based Scheme

One of the well-developed robust fault detection schemes relies on the use of unknown input observers (UIOs). A UIO can be applied to a class of systems described by:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + E\eta(t) \\ y(t) &= Cx(t)\end{aligned}\tag{4.98}$$

where x , u and y are respectively the state, input and output vector of the system, and η represents additive unknown disturbances. The matrices A , B , E , C are known. According to [Chen and Patton [1999]], *an observer is defined as an unknown input observer for the system in (4.98), if its state estimation error vector approaches zero asymptotically, regardless of the presence of the unknown input (disturbances) in the system.* Thus, the rationale behind the design of the UIOs is the decoupling of the disturbances, given that the structure of disturbances, represented by the matrix E , is a priori known.

The model of the UIO for the linear system (4.98) is shown below:

$$\begin{aligned}\dot{\hat{w}}(t) &= F\hat{w}(t) + TBu(t) + Ky(t) \\ \hat{x}(t) &= \hat{w}(t) + Hy(t)\end{aligned}\tag{4.99}$$

where \hat{x} is the estimated state vector, w is the state of the full-order observer and the matrices F , T , K and H are selected such that the following equations are satisfied [Chen and Patton [1999]]:

$$\begin{aligned}K &= K_1 + K_2 \\ (HC - I)E &= 0 \\ T &= I - HC \\ F &= A - HCA - K_1C \\ K_2 &= FH\end{aligned}\tag{4.100}$$

Taking into account (4.98)-(4.100), the state estimation error, defined

as $e = x - \hat{x}$, fulfils the following equation

$$\dot{e}(t) = Fe(t), \quad (4.101)$$

If all the eigenvalues of F are stable, the state estimation error e approaches zero asymptotically.

In order to detect additive sensor faults that affect the output of the linear system (4.98), that is

$$y(t) = Cx(t) + f(t) \quad (4.102)$$

where f represents the fault vector, the residual vector $r(t) = y(t) - C\hat{x}(t)$, or equivalently

$$r(t) = (I - CH)y(t) - Cw(t) \quad (4.103)$$

is generated and the fault detection decision logic is [Chen and Patton [1999]]

For the fault free case: $\|r(t)\| < \text{Threshold}$

For faulty cases: $\|r(t)\| \geq \text{Threshold}$

In order to isolate additive sensor faults, a bank of UIO-based residual generators is designed. The j -th UIO, $j \in \{1, \dots, m\}$ (m is the number of sensors) is designed taking into account the following output model

$$\begin{aligned} y^j(t) &= C^j x(t) + f^j(t) \\ y_j(t) &= c_j x(t) + f_j(t) \end{aligned} \quad (4.104)$$

where x is given in (4.98), c_j is the j -th row of the matrix C , C^j is made up from the matrix C by deleting the j -th row c_j , y_j is the j -th component of y (correspondingly for f_j) and y^j is obtained from the vector y by deleting the j -th component y_j (correspondingly for f^j). The structure of the j -th UIO is the following:

$$\begin{aligned} \dot{w}^j(t) &= F^j w(t) + T^j Bu(t) + K^j y^j(t) \\ \hat{x}_j(t) &= w^j(t) + H^j y^j(t) \end{aligned} \quad (4.105)$$

where \hat{x}_j is the estimated state vector, w_j is the state of the j -th full-order observer and the matrices F^j , T^j , K^j and H^j are selected in order to satisfy the following equations

$$\begin{aligned} K^j &= K_1^j + K_2^j \\ (H^j C^j - I)E &= 0 \\ T^j &= I - H^j C^j \\ F^j &= T^j A - K_1^j C^j \\ K_2^j &= F^j H^j \end{aligned} \tag{4.106}$$

The residual vector generated by the j -th UIO is defined as

$$r^j(t) = (I - C^j H^j) y^j(t) - C^j w^j(t) \tag{4.107}$$

Note that the UIO is driven by all outputs except for y_j , i.e. is driven by y^j . Hence, the residual r^j is sensitive to faults f^j except for fault f_j . Assuming the occurrence of a single sensor fault, the decision logic applied by the bank of UIO-residual generators is the following:

- if $\|r^j(t)\| < T_{sf}^j$ and $\|r^k(t)\| \geq T_{sf}^k$ for all $k \in \{1, \dots, j-1\} \cup \{j+1, \dots, m\}$, then sensor fault f^j is isolated.

The rationale behind this decision logic is that the j -th UIO observer is designed such that the disturbances η and the j -th sensor fault f_j are decoupled and do not affect the state estimation error, which converges asymptotically to zero.

Remark 4.2. The use of UIO in a fault detection scheme ensures the robustness of the scheme in an active way by decoupling the additive disturbances in the residual generation stage, and not in the evaluation stage i.e. when the residual is compared to a threshold that is computed without assuming any prior information about the disturbances. The decoupling of uncertainties in a fault detection scheme is feasible when the uncertainties are additive disturbances with known structure (i.e. matrix E in (4.98) is known), which is different from the structure of faults. When the uncertainties are multiplicative such as parametric uncertainty, decoupling may affect the performance of the fault detection scheme (i.e. the residual generation), because it may

decouple part of the system dynamics or faults. The fault detection schemes presented in this tutorial ensure robustness passively by treating the presence of the uncertainties in the evaluation stage, i.e. when the residual is compared to the threshold that is computed based on the available information for any type of uncertainties such as modeling errors, linearization errors, measurement noise, parametric uncertainty.

UIOs for Interconnected Systems

The UIO-based fault diagnosis scheme has been applied to interconnected, nonlinear systems [Klinkhieo et al. [2008]]. The design of the UIO scheme is realized by considering that the i -th subsystem is described by

$$\begin{aligned}\dot{x}_i(t) &= g_i(x_i(t), u_i(t)) + G_i(z_i(t)) \\ y_i(t) &= h(x_i(t)) + \alpha_i f_i(t)\end{aligned}\tag{4.108}$$

where x_i , u_i and y_i are the local state, input and output vector, and z_i is the interconnection vector, defined as:

$$z_i(t) = \sum_{j=1}^{N_i} Z_{ji} x_j(t)\tag{4.109}$$

where H_{ji} represents the interconnections between the i -th subsystem and its N_i neighboring subsystems. The function g_i represents the local dynamics, while the function G_i represents the interconnection dynamics. Then, by performing a small signal linearization of subsystem (4.108), the following representation is obtained:

$$\begin{aligned}\dot{x}_i(t) &= A_i x_i(t) + B_i u_i(t) + E_i z_i(t) \\ y_i(t) &= C_i x_i(t) + \alpha_i f_i(t)\end{aligned}\tag{4.110}$$

The linearization is realized assuming that no fault acts on the system, without changing the interconnection function G_i such that $G_i(z_i(t)) = E_i z_i(t)$. The UIO for the i -th subsystem given in (4.110) is designed considering that the “unknown input” is the interconnection vector z_i ; i.e. the UIO is structured as in (4.105) (by replacing the index j with

i), and the matrices F^i , T^i , K^i and H^i are selected in order to satisfy the following equations

$$\begin{aligned} K^i &= K_1^i + K_2^i \\ (H^i C_i - I^i) E_i &= 0 \\ T^i &= I^i - H^i C_i \\ F^i &= T^i A_i - K_1^i C_i \\ K_2^i &= F^i H^i \end{aligned} \tag{4.111}$$

Remark 4.3. In this tutorial, the known interconnection dynamics are modeled as a nonlinear vector field of the local state, input and interconnection vector, while there is modeling uncertainty that is also a function of the interconnection vector (see Section 2). If the interconnection dynamics are modeled as a linear or linearizable function of the interconnection signals only, UIOs can be an alternative approach, since the interconnecting signals can be treated as disturbances that can be decoupled as mentioned before.

4.4.2 Sliding Mode Observer-based Scheme

Several researchers have developed observer-based fault diagnosis scheme based on sliding mode techniques. The main design feature of a fault diagnosis scheme that uses sliding mode observers (SMO) is that it is able under certain conditions to reconstruct the fault functions, if no uncertainty exists in the system, or estimate the fault functions, if there is uncertainty. Next, we present one of the SMO-based sensor fault diagnosis scheme, developed in [Yan and Edwards [2007]].

Consider a nonlinear system of the form

$$\begin{aligned} \dot{x}(t) &= g(x(t), u(t)) + \Delta g(x(t)) \\ y(t) &= h(x(t)) + Df(t) \end{aligned} \tag{4.112}$$

where x , u and y are respectively the state, input and output vector of the system, g and h are smooth known vector fields, and D is a known sensor fault distribution matrix with full column rank. The unknown vector function $\Delta g(x(t))$ represents all the uncertainties and disturbances affecting the system and f is a sensor fault vector that

satisfies

$$\|f(t)\| \leq \bar{f}(t) \quad (4.113)$$

where \bar{f} is a known continuous function.

Under certain observability assumptions [García and D'Attelis [1995]], the authors in [Yan and Edwards [2007]] propose a nonlinear transformation $T : x \mapsto w$, structured as

$$\begin{aligned} w_{i1} &= h_i(x) \\ w_{i2} &= \mathcal{L}_{g(x,u)} h_i(x) \\ &\vdots \\ w_{ir_i} &= \mathcal{L}_{g(x,u)}^{r_i-1} h_i(x) \end{aligned} \quad (4.114)$$

where $w = \begin{bmatrix} w_1^\top & w_2^\top & \dots & w_p^\top \end{bmatrix}^\top$, p is the number of sensors, $w_i = \begin{bmatrix} w_{i1} & w_{i2} & \dots & w_{ir_i} \end{bmatrix}^\top$ for $i \in \{1, \dots, p\}$, and r_i is a uniform observability index of the pair $(g(x, u), h(x))$ in the domain $\Omega \times \mathcal{U}$ with $\sum_{i=1}^p r_i = n$, with $x \in \Omega \subset \mathbb{R}^n$ and $u \in \mathcal{U} \subset \mathbb{R}^m$ (for more details see [Yan and Edwards [2007]]).

Using the nonlinear transformation (4.114), system (4.112) is transformed:

$$\begin{aligned} \dot{w}(t) &= Aw(t) + B\Phi(w(t), u(t)) + \Psi(w(t)) \\ y(t) &= Cw(t) + Df(t) \end{aligned} \quad (4.115)$$

$A = \text{diag}\{A_1, \dots, A_p\}$, $B = \text{diag}\{B_1, \dots, B_p\}$ and $C = \text{diag}\{C_1, \dots, C_p\}$ with $A_i \in \mathbb{R}^{r_i \times r_i}$ and $B_i, C_i^\top \in \mathbb{R}^{r_i}$ defined as

$$A_i = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, B_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, C_i^\top = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad (4.116)$$

The vector fields $\Phi \in \mathbb{R}^p$ and Ψ are defined as:

$$\begin{aligned}\Phi(w, u) &= \begin{bmatrix} \mathcal{L}_{g(x,u)}^{r_1} h_1(x) \\ \mathcal{L}_{g(x,u)}^{r_2} h_2(x) \\ \vdots \\ \mathcal{L}_{g(x,u)}^{r_p} h_p(x) \end{bmatrix}_{x=T^{-1}(w)}, \\ \Psi(w) &= \begin{bmatrix} \psi_1(w) \\ \psi_2(w) \\ \vdots \\ \psi_p(w) \end{bmatrix} = \left[\frac{\partial T(x)}{\partial x} \Delta g(x) \right]_{x=T^{-1}(w)}\end{aligned}\quad (4.117)$$

with $\phi_i \in \mathbb{R}^{r_i}$, $i \in \{1, \dots, p\}$. The nonlinear function $\Phi(w, u)$ can be expressed as

$$\Phi(w, u) = -\Lambda w + \Gamma(w, u) \quad (4.118)$$

if there exists a matrix $\Lambda = \text{diag}\{\Lambda_1, \dots, \Lambda_p\}$ with $\Lambda_i = \begin{bmatrix} a_{i0} & a_{i1} & \dots & a_{i(r_i-1)} \end{bmatrix}$ such that the roots of the polynomial

$$\lambda^{r_i} + a_{i(r_i-1)}\lambda^{r_i-1} + \dots + a_{i1}\lambda + a_{i0} = 0 \quad (4.119)$$

lie in left-half plane for $i \in \{1, \dots, p\}$, and $L(u)$ such that

$$\|\Gamma(w, u) - \Gamma(\hat{w}, u)\| \leq L(u)\|w - \hat{w}\| \quad (4.120)$$

for any $w, \hat{w} \in T(\Omega)$. For the system (4.115), the following linear filter is introduced:

$$\dot{w}_a = A_a w_a + B_a y \quad (4.121)$$

where A_a is selected to be Hurwitz, and B_a is chosen to be a nonsingular matrix such that

$$BD = \begin{bmatrix} 0_{(p-q) \times q} \\ D_1 \end{bmatrix} \quad (4.122)$$

where D_1 is nonsingular. Introducing (4.118) in (4.115), and then combining with (4.121) leads to the augmented system

$$\dot{w} = (A - B\Lambda)w + B\Gamma(w, u) + \Psi(w), \quad (4.123)$$

$$\dot{w}_a = B_a C w + A_a w_a + B_a D f, \quad (4.124)$$

$$y_a = C_a w_a \quad (4.125)$$

where $A - B\Lambda$ is stable due to the design of Λ and C_a is selected to be orthogonal. Note that the sensor fault in (4.115) has been transformed in a pseudo-actuator fault in (4.123)-(4.125). The sliding mode observer is structured as:

$$\dot{\hat{w}} = (A - B\Lambda)\hat{w} + B\Gamma(\hat{w}, u), \quad (4.126)$$

$$\dot{\hat{w}}_a = B_a C \hat{w} + A_a \hat{w}_a + \nu(t, y_a, \hat{y}_a), \quad (4.127)$$

$$\hat{y}_a = C_a \hat{w}_a \quad (4.128)$$

$$\nu(t, y_a(t), \hat{y}_a(t)) = k(t) C_a^\top \frac{y_a(t) - \hat{y}_a(t)}{\|y_a(t) - \hat{y}_a(t)\|} \quad (4.129)$$

where \hat{w} and \hat{w}_a are the estimations of w and w_a , respectively, and the gain $k(t)$ is chosen to satisfy

$$k(t) \geq \|B_a C\|b + \|B_a D\|\bar{f}(t) + \eta \quad (4.130)$$

where \bar{f} is defined in (4.113), η is a positive constant and b is a bound on the state estimation error; i.e.

$$\sup_{t \in [0, \infty)} \{w(t) - w_a(t)\} \leq b \quad (4.131)$$

The bound b can be computed taking into account that after a finite time instant the state estimation error $e \triangleq w - \hat{w}$ will remain within the set \mathcal{B} , defined as

$$\mathcal{B} = \left\{ e : e^\top P e \leq \left(\frac{d + \epsilon_1}{-\lambda_{\max}(P^{-1}Q)} \right)^2 \right\} \quad (4.132)$$

where $\epsilon_1 > 0$, P is a symmetric positive definite matrix, which is the solution of a set of linear matrix inequalities detailed in [Yan and Edwards [2007]]

$$Q = (A - B\Lambda)^\top P + P(A - B\Lambda) + \frac{1}{\epsilon_2} P B B^\top P + \epsilon_2 L(u)^2 I \quad (4.133)$$

with $\epsilon_2 > 0$, and d is a bound on the uncertainty term Ψ ; i.e.,

$$\sqrt{\Psi(w(t))^\top P \Psi(w(t))} \leq 0.5d, \quad w(t) \in T(\Omega) \quad (4.134)$$

The estimation signal $\hat{f}(t)$ (or reconstruction signal in case that $d = 0$) is computed as

$$\hat{f}(t) = \begin{bmatrix} 0_{q \times (p-q)} & D_1^{-1} \end{bmatrix} k(t) C_a^\top \frac{y_a(t) - \hat{y}_a(t)}{\|y_a(t) - \hat{y}_a(t)\| + \sigma} \quad (4.135)$$

where σ is an appropriate positive constant. Assuming no uncertainty, $\hat{f}(t)$ can duplicate the actual fault precisely after some time. Hence, sensor faults are detected if $\hat{f}(t)$ remains non-zero after some time instant. In the presence of uncertainty, the fault detection decision logic is based on checking if $\hat{f}(t)$ exceeds a threshold appropriate for isolation.

Remark 4.4. In order to design a SMO, an upper bound on the sensor faults $\bar{f}(t)$ is needed for selecting the observer gain via (4.130). If there is no available prior knowledge for $\bar{f}(t)$, a larger bound \bar{f}^* can be used such as $\bar{f}^* > \bar{f}(t)$, which may increase the observer gain. In the fault detection (and isolation) schemes described in this tutorial, we assume that the sensor faults are bounded but the design of the schemes does not require the knowledge of the bound.

4.4.3 Reduced-order Nonlinear Observer-based Scheme

in this section we present the design of a reduced-order nonlinear observer-based fault detection scheme developed in [Zhang [2011]], which can be applied to a class of nonlinear systems, where the nonlinearity is a function of state, input and output variables.

Consider the class of nonlinear systems, expressed as:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + D\zeta(x(t), u(t)) + g(y(t), u(t)) + \varphi(x(t), u(t), t) \\ y(t) &= Cx(t) + \beta(t - T_0)F\theta(t) \end{aligned} \quad (4.136)$$

where x , u , y are respectively the state, input and output of the system, ζ and g are known nonlinear functions, while φ represents the modeling uncertainty and θ denotes the time-varying sensor fault. The matrix

F stands for the location of the sensor fault. It is assumed that the matrices $D \in \mathbb{R}^{n \times v}$ and $C \in \mathbb{R}^{p \times n}$ are of full rank with $v \leq p < n$, and the pair (A, C) is observable, while D fulfills the following conditions: (i) $\text{rank}(CD) = v$, (ii) all the invariant zeros of (A, D, C) (if any) lie in the left half plane. Under the aforementioned assumptions, system (4.136) is transformed using a linear change of coordinates, that is $w = \begin{bmatrix} w_1^\top & w_2^\top \end{bmatrix}^\top = Tx$ with $w_1 \in \mathbb{R}^{(n-p)}$ and $w_2 \in \mathbb{R}^p$ such that

- $TAT^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, where the matrix $A_{11} \in \mathbb{R}^{(n-p) \times (n-p)}$ is stable
- $TD = \begin{bmatrix} 0 \\ D_2 \end{bmatrix}$, where $D_2 \in \mathbb{R}^{p \times v}$
- $CT^{-1} = [0 \ I_p]$

The system in the new coordinates is expressed as:

$$\dot{w}_1(t) = A_{11}w_1(t) + A_{12}w_2(t) + \psi_1(y(t), u(t)) + \eta_1(w(t), u(t), t) \quad (4.137)$$

$$\begin{aligned} \dot{w}_2(t) &= A_{21}w_1(t) + A_{12}w_2(t) + \rho(w(t), u(t)) + \psi_2(y(t), u(t)) \\ &\quad + \eta_2(w(t), u(t), t) \end{aligned} \quad (4.138)$$

$$y(t) = \bar{C}w_2(t) + \beta(t - T_0)F\theta(t) \quad (4.139)$$

where $\rho(w, u) = D_2\zeta(w, u)$, $\bar{C} = I_p$, $\begin{bmatrix} \psi_1(y, u)^\top & \psi_2(y, u)^\top \end{bmatrix}^\top = Tg(y, u)$, and $\begin{bmatrix} \eta_1(w, u, t)^\top & \eta_2(w, u, t)^\top \end{bmatrix}^\top = T\varphi(T^{-1}w, u, t)$.

Taking into account (4.137)-(4.139), the proposed observer is structured as:

$$\dot{\hat{w}}_1(t) = A_{11}\hat{w}_1(t) + A_{12}\bar{C}^{-1}y(t) + \psi_1(y(t), u(t)) \quad (4.140)$$

$$\begin{aligned} \dot{\hat{w}}_2(t) &= A_{21}\hat{w}_1(t) + A_{12}w_2(t) + \rho(\hat{w}(t), u(t)) + \psi_2(y(t), u(t)) \\ &\quad + L(y(t) - \hat{y}(t)) \end{aligned} \quad (4.141)$$

$$\hat{y}(t) = \bar{C}\hat{w}_2(t) \quad (4.142)$$

where \hat{w}_1 and \hat{w}_2 are the estimations of w_1 and w_2 , and $\hat{w} = \left[\hat{w}_1^\top (\bar{C}^{-1}y)^\top \right]^\top$. The residual vector corresponds to the output estimation error $\tilde{y} = y - \hat{y}$, where the j -th residual is defined as $\tilde{y}_j = \bar{C}_j \tilde{w}_2$ with $\tilde{w}_2 = w_2 - \hat{w}_2$. The fault detection decision logic relies on checking whether the j -th residual $\tilde{y}_j(t)$ exceeds the adaptive threshold denoted by $\nu_j(t)$; i.e.,

$$|\tilde{y}_j(t)| > \nu_j(t) \quad (4.143)$$

The adaptive threshold $\nu_j(t)$ is computed taking into account that: (i) the modeling uncertainties are bounded, that is $|\eta_q(w, u, t)| \leq \bar{\eta}_q(y, u, t)$, $q = 1, 2$, (ii) the known nonlinear term $\rho(w, u)$ is uniformly Lipschitz, i.e. $|\rho(w, u) - \rho(\hat{w}, u)| \leq \gamma|w - \hat{w}|$, where γ is a Lipschitz constant and (iii) the state w remains bounded before and after the fault occurrence. The adaptive threshold is defined as:

$$\nu_j(t) = k_j \int_0^t e^{-\lambda_j(t-\tau)} ((\|A_{21}\| + \gamma) \chi(\tau) + \bar{\eta}_2(y, u, \tau)) \quad (4.144)$$

$$\chi(t) = k_0 \omega_0 e^{-\lambda_0 t} + k_0 \int_0^t e^{-\lambda_0(t-\tau)} \bar{\eta}_1(y, u, \tau) d\tau \quad (4.145)$$

where $k_0, k_j, \lambda_0, \lambda_j$ are positive, design parameters such that $\|e^{A_{11}t}\| \leq k_0 e^{-\lambda_0 t}$ and $|\bar{C}_j e^{A_{22}t}| \leq k_j e^{-\lambda_j t}$.

Remark 4.5. In [Zhang and Zhang [2011]], the system nonlinearity is decomposed into a structured nonlinear part that depends on the states and input, and an unstructured nonlinear part that is function of measured states and input. Based on this decomposition and exploiting the structured nonlinearity, the system may be transformed in a new coordinate system that facilitates the design of a ‘reduced-order’ nonlinear observer with a square gain matrix L , or else reduces the complexity of the observer scheme. The existence of the change of coordinates is ensured under certain algebraic conditions associated with the system structure.

Nonlinear Observer-based Scheme for Interconnected Systems

The proposed observer-based fault detection scheme has been applied to interconnected systems [Zhang and Zhang [2013]], with the i -th sub-

system modeled as:

$$\begin{aligned} \dot{x}_i(t) = & A_i x_i(t) + D_i \zeta_i(x_i(t), u_i(t)) + g_i(y_i(t), u_i(t)) + \sum_{j=1}^M h_{ij}(x_j, u_j) \\ & + E_i \varphi(x_i(t), u_i(t), t), \end{aligned} \quad (4.146)$$

$$y_i(t) = C_i x_i(t) + \beta_i(t - T_0) \theta_i(t) \quad (4.147)$$

where h_{ij} denotes the interconnection function between the i -th and j -th subsystem. The i -th subsystem is transformed in a new coordinate system. Particularly, by assuming that the matrices $D_i \in \mathbb{R}^{n_i \times v_i}$, $E_i \in \mathbb{R}^{n_i \times q_i}$ and $C_i \in \mathbb{R}^{p_i \times n_i}$ are of full rank with $v_i \leq p_i < n_i$, while D_i and E_i fulfill the following conditions: (i) $\text{rank}(C_i D_i) = v_i$, and (ii) $\text{rank}(C_i E_i) = q_i$. Under these assumptions, there exists a new system of coordinates $w_i = \begin{bmatrix} w_{i1}^\top & w_{i2}^\top \end{bmatrix}^\top = T_i x_i$ with $w_{i1} \in \mathbb{R}^{(n_i - p_i)}$ and $w_{i2} \in \mathbb{R}^{p_i}$ such that: (i) $T_i D_i = \begin{bmatrix} 0 \\ D_{i2} \end{bmatrix}$,

$$T_i E_i = \begin{bmatrix} 0 \\ E_{i2} \end{bmatrix}$$

where $D_{i2} \in \mathbb{R}^{p_i \times v_i}$ and $E_{i2} \in \mathbb{R}^{p_i \times q_i}$, and (ii) $C_i T_i^{-1} = \begin{bmatrix} 0 & \bar{C}_i \end{bmatrix}$, where \bar{C}_i is orthogonal. The system in the new coordinates is described by:

$$\dot{w}_{i1}(t) = A_{i1} w_{i1}(t) + A_{i2} w_{i2}(t) + \psi_{i1}(y_i(t), u_i(t)) + \sum_{j=1}^M H_{ij}^1(w_j(t), u_j(t)) \quad (4.148)$$

$$\begin{aligned} \dot{w}_{i2}(t) = & A_{i3} w_{i1}(t) + A_{i4} w_{i2}(t) + \rho_i(w_i(t), u_i(t)) + \psi_{i2}(y_i(t), u_i(t)) \\ & + \eta_i(w_i(t), u_i(t), t) + \sum_{j=1}^M H_{ij}^2(w_j(t), u_j(t)) \end{aligned} \quad (4.149)$$

$$y_i(t) = \bar{C}_i w_{i2}(t) + \beta_i(t - T_0) \theta_i(t) \quad (4.150)$$

where $\begin{bmatrix} A_{i1} & A_{i2} \\ A_{i3} & A_{i4} \end{bmatrix} = T_i A_i T_i^{-1}$, $\begin{bmatrix} \psi_{i1}^\top & \psi_{i2}^\top \end{bmatrix}^\top = T_i g_i$, and $\begin{bmatrix} (H_{ij}^1)^\top & (H_{ij}^2)^\top \end{bmatrix}^\top = T_i h_{ij}$. The functions ρ_i and η_{i2} represent the known nonlinearity and modeling uncertainty, respectively, in the new coordinate system.

Given the transformed system (4.148)-(4.150), the nonlinear observer associated with the i -th subsystem is structured as:

$$\begin{aligned} \dot{\hat{w}}_{i1}(t) &= A_{i1}\hat{w}_{i1}(t) + A_{i2}\bar{C}_i^{-1}y_i(t) + \psi_{i1}(y_i(t), u_i(t)) \\ &\quad + \sum_{j=1}^M H_{ij}^1(\hat{w}_j(t), u_j(t)) \end{aligned} \quad (4.151)$$

$$\begin{aligned} \dot{\hat{w}}_{i2}(t) &= A_{i3}\hat{w}_{i1}(t) + A_{i4}w_{i2}(t) + \rho_i(\hat{w}_i(t), u_i(t)) + \psi_{i2}(y_i(t), u_i(t)) \\ &\quad + \sum_{j=1}^M H_{ij}^2(\hat{w}_j(t), u_j(t)) + L(y_i(t) - \hat{y}_i(t)) \end{aligned} \quad (4.152)$$

$$\hat{y}_i(t) = \bar{C}_i \hat{w}_{i2}(t) \quad (4.153)$$

where \hat{w}_{i1} and \hat{w}_{i2} are the estimations of w_{i1} and w_{i2} with $\hat{w}_i = \begin{bmatrix} \hat{w}_{i1}^\top & (\bar{C}_i^{-1}y_i)^\top \end{bmatrix}^\top$ (correspondingly for \hat{w}_j). The residual vector related to the i -th subsystem corresponds to the output estimation error $\tilde{y}_i = y_i - \hat{y}_i$, where the p -th residual is defined as $\tilde{y}_{ip} = \bar{C}_{ip}\tilde{w}_{i2}$ with $\tilde{w}_{i2} = w_{i2} - \hat{w}_{i2}$. Sensor faults are detected when the p -th residual $\tilde{y}_{ip}(t)$ exceeds the adaptive threshold denoted by $\nu_{ip}(t)$; i.e.,

$$|\tilde{y}_{ip}(t)| > \nu_{ip}(t) \quad (4.154)$$

The adaptive threshold $\nu_{ip}(t)$ is computed taking into account that: (i) the state w_i of each subsystem remains bounded before and after fault occurrence, and (ii) the known nonlinear term $\rho(w, u)$ is uniformly Lipschitz, i.e. $|\rho_i(w_i, u_i) - \rho_i(\hat{w}_i, u_i)| \leq \gamma_i(y_i, u_i)|w_i - \hat{w}_i|$, where $\gamma_i(y_i, u_i)$ is a known, uniformly bounded function, (iii) the modeling uncertainty is bounded, that is $|\eta_i(w_i, u_i, t)| \leq \bar{\eta}_q(y_i, u_i, t)$, and (iv) the interconnection terms satisfy the following condition $\left| H_{ij}^1(w_j, u_j) - H_{ij}^1(\hat{w}_j, u_j) \right| \leq \gamma_{ij}^1 |w_j - \hat{w}_j|$ and $\left| H_{ij}^2(w_j, u_j) - H_{ij}^2(\hat{w}_j, u_j) \right| \leq \gamma_{ij}^2(y_i, u_i) |w_j - \hat{w}_j|$, where γ_{ij}^1 is a Lipschitz constant, and $\gamma_{ij}^2(y_i, u_i)$ is a known, uniformly

bounded function. The adaptive threshold is defined as:

$$\nu_{ip}(t) = k_{ip} \int_0^t e^{-\lambda_{ip}(t-\tau)} (|\xi_i(y_i, u_i)| \chi(\tau) + \bar{\eta}_i(y_i, u_i, \tau)) d\tau \quad (4.155)$$

$$\xi_i(y_i, u_i) = \left[\gamma_{i1}^2(y_i, u_i), \dots, \gamma_{i(i-1)}^2(y_i, u_i), \|A_{i3}\| + \gamma_i, \gamma_{i(i+1)}^2(y_i, u_i), \dots, \gamma_{iM}^2(y_i, u_i) \right] \quad (4.156)$$

$$\chi(t) = \frac{\bar{V}_0 e^{-t(\lambda_{\min}(\bar{Q})/\lambda_{\max}(\text{diag}\{\bar{P}_1, \dots, \bar{P}_M\}))}}{\lambda_{\min}(\text{diag}\{\bar{P}_1, \dots, \bar{P}_M\})} \quad (4.157)$$

where k_{ip} , λ_{ip} are positive constants such that $|\bar{C}_{ip} e^{A_{i4}t}| \leq k_{ip} e^{-\lambda_{ip}}$, \bar{V}_0 is a positive constant such that $|V(0)| \leq \bar{V}_0$ with $V(t) = \sum_{i=1}^M (w_{i1}(t) - \hat{w}_{i1}(t))^\top \bar{P}_i (w_{i1}(t) - \hat{w}_{i1}(t))$, while the matrices \bar{P}_i and \bar{Q} should satisfy the following conditions:

- the matrix \bar{P}_i is symmetric, positive definite such that $\bar{R}_i = -A_{i1}^\top \bar{P}_i - \bar{P}_i A_{i1} > 0$, for all $i \in \{1, \dots, M\}$
- the matrix \bar{Q} is square with the element i, j is defined as $\bar{Q}_{ij} = \lambda_{\min}(\bar{R}_i)$ if $i = j$, and $\bar{Q}_{ij} = -\|\bar{P}_i\| \gamma_{ij}^1 - \|\bar{P}_j\| \gamma_{ji}^1$ if $i \neq j$.

4.4.4 Interval Observer-based Scheme

in this section, we present the design of an interval observer for a class of continuous time, nonlinear systems developed in [Raïssi et al. [2010]].

Consider the nonlinear system of the form:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), p) + v(t), \quad x(0) \in [x_0] \\ y(t) &= g(x(t), u(t), p) + w(t) \end{aligned} \quad (4.158)$$

where x , u and y are respectively the state, input and output vector of the system with the initial state x_0 restricted within the compact set $[x_0]$ (the brackets $[\cdot]$ denote a set), p is the parameter vector, while v and w represent the process and measurement noise respectively. The functions f and g are assumed to be Lipschitz. In order to design an interval observer, a quasi linear parameter varying (qLPV) representation should be obtained for system (4.158), and the observation error

dynamics should be *cooperative* (see [Raïssi et al. [2010]] for more details about the cooperativity property). A qLPV representation, which guarantees that the nonlinear trajectory remains within the set of its resulting trajectories, is obtained by performing a linearisation around the operational state domain instead of a linearisation throughout the equilibrium points.

Let us consider that $x \in [x]$, $u \in [u]$, $p \in [p]$ and $d = [v^\top \ w^\top]^\top$ with $d \in [d]$ and $[d]$ is computed taking into account that the process and measurement noises are bounded, that is $v \in [v]$ and $w \in [w]$, and denote $\rho(t) = [x(t)^\top, u(t)^\top, d(t)^\top]^\top$ the scheduling parameter. A guaranteed qLPV model is expressed as:

$$\begin{aligned} \dot{x}(t) &= A(\rho(t), p)x(t) + B(\rho(t), p)u(t) + E(\rho(t), p)d(t) \\ y(t) &= C(\rho(t), p)x(t) + D(\rho(t), p)u(t) + G(\rho(t), p)d(t) \end{aligned} \quad (4.159)$$

with $A(\rho(t), p) \in [\underline{A}, \overline{A}]$, $B(\rho(t), p) \in [\underline{B}, \overline{B}]$, $E(\rho(t), p) \in [\underline{E}, \overline{E}]$, $C(\rho(t), p) \in [\underline{C}, \overline{C}]$, $D(\rho(t), p) \in [\underline{D}, \overline{D}]$ and $G(\rho(t), p) \in [\underline{G}, \overline{G}]$. The interval matrices $[\underline{A}, \overline{A}]$, $[\underline{B}, \overline{B}]$, $[\underline{E}, \overline{E}]$, $[\underline{C}, \overline{C}]$, $[\underline{D}, \overline{D}]$, $[\underline{G}, \overline{G}]$, are computed by performing interval analysis, and specifically by using mean value inclusion functions computed and the midpoints of $[x]$, $[u]$, $[v]$, $[w]$ and p (for more details, see [Raïssi et al. [2010]]). By denoting $u = (u, d)^\top$ and $B = (B, E)$, the qLPV representation is re-written as:

$$\begin{aligned} \dot{x}(t) &= A(x(t), u(t))x(t) + B(x(t), u(t))u(t) \\ y(t) &= C(x(t), u(t))x(t) + D(x(t), u(t))u(t) \end{aligned} \quad (4.160)$$

Given the qLPV model (4.160), the interval observer consists of two point observers $\underline{\mathcal{O}}$ and $\overline{\mathcal{O}}$, structured as:

$$\underline{\mathcal{O}}: \begin{cases} \dot{\underline{x}}(t) = (\underline{A} - L\underline{C})\underline{x}(t) + (\underline{B} - L\underline{D})\underline{u}(t) + L(y(t) + \underline{\epsilon}) \\ \underline{y}(t) = \underline{C}\underline{x}(t) + \underline{D}\underline{u}(t) \end{cases} \quad (4.161)$$

$$\overline{\mathcal{O}}: \begin{cases} \dot{\overline{x}}(t) = (\overline{A} - L\overline{C})\overline{x}(t) + (\overline{B} - L\overline{D})\overline{u}(t) + L(y(t) + \overline{\epsilon}) \\ \overline{y}(t) = \overline{C}\overline{x}(t) + \overline{D}\overline{u}(t) \end{cases} \quad (4.162)$$

where $\underline{\epsilon}$ and $\overline{\epsilon}$ are the bounds on the measurement error. The observer gain L is selected such that: (i) the interval matrices $\underline{A} - L\underline{C}$ and $\overline{A} - L\overline{C}$ are cooperative, (ii) the interval matrix $(\Delta_{\underline{A}} - L\Delta_{\underline{C}})\underline{x}(t) +$

$(\Delta_{\underline{B}} - L\Delta_{\underline{D}})\underline{u}(t) + L\underline{\epsilon}$ is negative, while the matrix $(\Delta_{\overline{A}} - L\Delta_{\overline{C}})\overline{x}(t) + (\Delta_{\overline{B}} - L\Delta_{\overline{D}})\overline{u}(t) + L\overline{\epsilon}$ is positive, and (iii) the matrix $\text{mid}[A] - L\text{mid}[C]$, where mid denotes the midpoint of a set, is stable.

In this work, the residual $\underline{r}(t)$ and $\overline{r}(t)$ respectively associated with each observer $\underline{\mathcal{O}}$ and $\overline{\mathcal{O}}$ are defined as

$$\underline{r}(t) = \underline{y}(t) - y(t), \quad \overline{r}(t) = \overline{y}(t) - y(t) \quad (4.163)$$

Taking into account that $y(t) \in [\underline{y}(t), \overline{y}(t)]$ under healthy conditions, the fault detection decision logic is based on the test

$$0 \notin [\underline{y}(t), \overline{y}(t)] - y(t) = [\underline{r}(t), \overline{r}(t)] \quad (4.164)$$

Remark 4.6. The common design characteristic of the observer scheme proposed in this tutorial and the observer schemes presented in Sections 4.4.1-4.4.3 is the use of a single observer that generates an estimation $\hat{x}(t)$ of the state vector $x(t)$, and the observer design parameters are obtained such that the state estimation error $x(t) - \hat{x}(t)$ is bounded, and converges to zero if no uncertainty affects the system. On the contrary, an interval observer consists of a pair of two point-wise observers, which generate a lower and upper bound $\underline{x}(t)$ and $\overline{x}(t)$ for the state vector $x(t)$ such that $x(t) \in [\underline{x}(t), \overline{x}(t)]$. In this case, the interval observer gain is selected such that $\underline{x}(t) - x(t) \leq 0$ and $\overline{x}(t) - x(t) \geq 0$, and the error $\underline{x}(t) - \overline{x}(t)$ converges to zero. Similarly, the fault detection decision logic proposed in this tutorial and in the fault detection schemes presented in Sections 4.4.1-4.4.3 relies on checking if the magnitude of the residual $r(t)$ exceeds a threshold $r^{th}(t)$, or equivalently if $r(t) \notin [-r^{th}(t), r^{th}(t)]$. Using interval observers, the fault detection decision test is $0 \notin [\underline{r}(t), \overline{r}(t)]$.

4.5 Illustrative Example

The objective of this section is to illustrate the design methodology of a distributed and decentralized sensor fault detection method applied to the interconnected nonlinear subsystems presented in Example 2.2. By considering that the interconnection dynamics of subsystem $\Sigma^{(1)}$

are expressed as $h^{(1)}(z^{(1)}) = b^{(1)} \sin(z^{(1)})$, $\Sigma^{(1)}$ is described by

$$\begin{aligned} \Sigma^{(1)} : \dot{x}^{(1)} = & A^{(1)}x^{(1)} + Q_1 \left(g_1^{(1)} u^{(1)} \right) + Q_1 \left(g_2^{(1)} \cos(x_1^{(1)}) \right) \\ & + Q_1 \left(b^{(1)} \sin(z^{(1)}) \right) + Q_1 \left(\eta_2^{(1)}(x^{(1)}, u^{(1)}, z^{(1)}) \right), \end{aligned} \quad (4.165)$$

where $A^{(1)}$ is defined in (2.11) and

$$Q_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top \quad (4.166)$$

The modeling uncertainty is determined as

$$\eta_2^{(1)}(x^{(1)}, u^{(1)}, z^{(1)}) = \tilde{g}_2^{(1)} \cos(x_1^{(1)}) \quad (4.167)$$

satisfying

$$\left| \eta_2^{(1)}(x^{(1)}, u^{(1)}, z^{(1)}) \right| \leq \bar{g}_2^{(1)} \left| \cos(x_1^{(1)}) \right| \quad (4.168)$$

By considering that the interconnection dynamics of subsystem $\Sigma^{(2)}$ are expressed as $h^{(2)}(z^{(2)}) = b^{(2)} \cos(z^{(2)})$, subsystem $\Sigma^{(2)}$ is described by

$$\begin{aligned} \Sigma^{(2)} : \dot{x}^{(2)} = & A^{(2)}x^{(2)} + Q_1 \left(g_1^{(2)} u^{(2)} \right) + Q_1 \left(g_2^{(2)} \sin(x_1^{(2)}) \right) \\ & + Q_1 \left(b^{(2)} \cos(z^{(2)}) \right) + Q_1 \left(\eta_2^{(2)}(x^{(2)}, u^{(2)}, z^{(2)}) \right), \end{aligned} \quad (4.169)$$

where $A^{(2)}$ is defined in (2.11) and the modeling uncertainty is determined as

$$\eta_2^{(2)}(x^{(2)}, u^{(2)}, z^{(2)}, t) = \tilde{g}_2^{(2)} \sin(x_1^{(2)}) \quad (4.170)$$

satisfying

$$\left| \eta_2^{(2)}(x^{(2)}, u^{(2)}, z^{(2)}, t) \right| \leq \bar{g}_2^{(2)} \left| \sin(x_1^{(2)}) \right|. \quad (4.171)$$

The sensors used for the control of $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are characterized by

$$\mathcal{S}^{(I)}\{1\} : \quad y_1^{(I)}(t) = C_1^{(I)} x^{(I)}(t) + d_1^{(I)}(t) + f_1^{(I)}(t), \quad (4.172)$$

$$\mathcal{S}^{(I)}\{2\} : \quad y_2^{(I)}(t) = C_2^{(I)} x^{(I)}(t) + d_2^{(I)}(t) + f_2^{(I)}(t), \quad (4.173)$$

where

$$C_1^{(I)} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad (4.174)$$

$$C_2^{(I)} = \begin{bmatrix} 0 & 1 \end{bmatrix}. \quad (4.175)$$

For every subsystem, we design a monitoring agent, i.e. $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. The design of each monitoring module starts with the decomposition of the sensor set $\mathcal{S}^{(I)}$, based on Algorithm 1. Initially, it is checked if the pairs $(A^{(I)}, C^{(I,q)})$ with

$$C^{(I,1)} = C_1^{(I)}, \quad (4.176)$$

$$C^{(I,2)} = C_2^{(I)} \quad (4.177)$$

are observable. Since all pairs are observable, we can find $L^{(I,q)}$ such that $A^{(I)} - L^{(I,q)}C^{(I,q)}$ is stable and we proceed with checking the satisfaction of the second condition, which is customized according to the model of the subsystem under investigation; i.e., we check if for all $I = 1, 2$ and $q = 1, 2$, there are positive constants $\rho^{(I,q)}, \xi^{(I,q)}$ such that $|Q_2 e^{(A^{(I)} - L^{(I,q)}C^{(I,q)})t} Q_1| \leq \rho^{(I,q)} e^{-\xi^{(I,q)}t}$ and $\xi^{(I,q)} > \rho^{(I,q)}(|g_2^{(I)}| + \bar{g}_2^{(I)})$ as we analyze next (see Section 4.5.1), where

$$Q_2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad (4.178)$$

Since the existence of positive constants $\rho^{(I,q)}, \xi^{(I,q)}$ is ensured, the sensor set $\mathcal{S}^{(I)}$ is decomposed into two singletons, defined as $\mathcal{S}^{(I,1)} = \{\mathcal{S}^{(I)}\{1\}\}$ and $\mathcal{S}^{(I,2)} = \{\mathcal{S}^{(I)}\{2\}\}$, which are characterized by the sensor outputs $y_1^{(I)}$ and $y_2^{(I)}$; i.e.,

$$\mathcal{S}^{(I,1)} : y^{(I,1)} = y_1^{(I)}, \quad (4.179)$$

$$\mathcal{S}^{(I,2)} : y^{(I,2)} = y_2^{(I)}. \quad (4.180)$$

Thus, every agent consists of two modules $\mathcal{M}^{(I,1)}$ and $\mathcal{M}^{(I,2)}$, $I = 1, 2$ that monitor respectively sensor $\mathcal{S}^{(I)}\{1\}$ and $\mathcal{S}^{(I)}\{2\}$, as shown in Fig. 4.1. In this distributed architecture, the transmitted information for each agent is described by

$$\mathcal{S}_z^{(1)} : y_z^{(1)} = y_1^{(2)}, \quad (4.181)$$

$$\mathcal{S}_z^{(2)} : y_z^{(2)} = y_1^{(1)}, \quad (4.182)$$

implying that $\mathcal{S}_z^{(1)} = \{\mathcal{S}_1^{(2)}\}$ and $\mathcal{S}_z^{(2)} = \{\mathcal{S}_1^{(1)}\}$. Note that in a decentralized sensor fault diagnosis scheme, the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ use the reference signals $x_{1ref}^{(2)}$ and $x_{1ref}^{(1)}$ instead of the transmitted measurements $y_1^{(2)}$ and $y_1^{(1)}$.

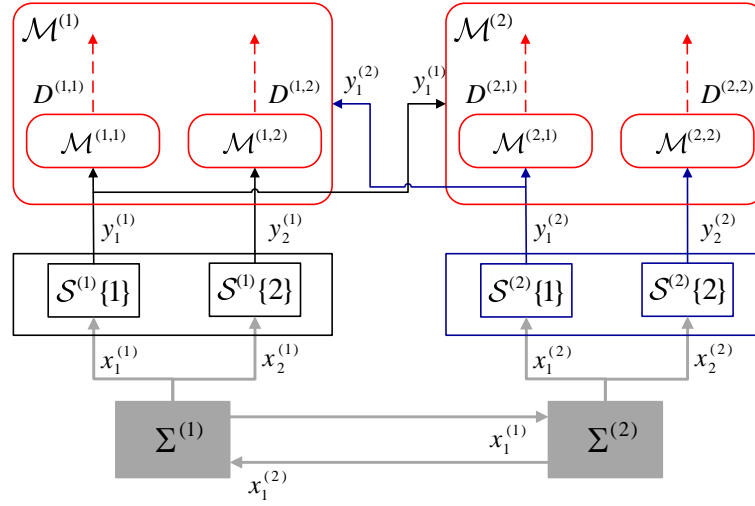


Figure 4.1: Illustrative example: Distributed sensor fault detection scheme for two interconnected systems.

After decomposing the monitoring of the sensor set $\mathcal{S}^{(I)}$, $I = 1, 2$ into two monitoring modules, we proceed with the design of each module. An overview of the distributed design of $\mathcal{M}^{(I,q)}$, $I = 1, 2$ and $q = 1, 2$ is illustrated in Fig. 4.2, taking into account (4.179)-(4.182). A nonlinear observer $\mathcal{O}^{(I,q)}$ is designed to generate the estimated state vector $\hat{x}^{(I,q)}$ used for generating the residual $\varepsilon_{y_q}^{(I,q)} = y_q^{(I)} - C_q^{(I)} \hat{x}^{(I,q)}$ and the adaptive threshold $\bar{\varepsilon}_{y_q}^{(I,q)}$. The module $\mathcal{M}^{(I,q)}$ infers the presence of sensor faults, when the output $D^{(I,q)}$ becomes ‘1’, that is when the analytic redundancy relation $\mathcal{E}^{(I,q)}$ is not satisfied. Note that in a decentralized sensor fault diagnosis scheme, $x_{1ref}^{(I)}$ is used instead of $y_z^{(I)}$.

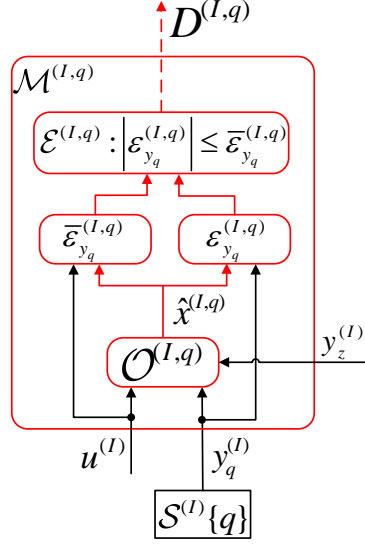


Figure 4.2: Illustrative example: Overview of the design of a distributed monitoring module $\mathcal{M}^{(I,q)}$, $I = 1, 2$ and $q = 1, 2$.

4.5.1 Residual Generation

in this section, we present the distributed and decentralized design of the nonlinear observer that generates the estimation of the state vector of subsystems $\Sigma^{(1)}$ and $\Sigma^{(2)}$, that is $\hat{x}^{(I,q)}$, $I = 1, 2$ and $q = 1, 2$, which is used in the residual, described by

$$\varepsilon_{y_q}^{(I,q)} = y_q^{(I)} - C_q \hat{x}^{(I,q)}, \quad I = 1, 2, \quad q = 1, 2 \quad (4.183)$$

Observer Associated with the Modules $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(1,2)}$

Taking into account the model of subsystem $\Sigma^{(1)}$ given in (4.165), a distributed and a decentralized nonlinear observer associated with the monitoring module $\mathcal{M}^{(1,q)}$ are structured as in (4.3) (omitting $\zeta_r^{(I)}$, since the interconnection state is measured) and (4.41), respectively, where $C^{(1,q)}$ and $y^{(1,q)}$, are respectively defined in (4.176) and (4.179);

i.e.,

$$\begin{aligned} \mathcal{O}^{(1,q)} : \dot{\hat{x}}^{(1,q)} = & A^{(1)}\hat{x}^{(1,q)} + Q_1\left(g_1^{(1)}u^{(1)}\right) + Q_1\left(g_2^{(1)}\cos(\hat{x}_1^{(1,q)})\right) \\ & + Q_1\left(b^{(1)}\sin(z_\star^{(1)})\right) + L^{(1,q)}\left(y_q^{(1)} - C_q^{(1)}\hat{x}^{(1,q)}\right) \end{aligned} \quad (4.184)$$

where Q_1 is given in (4.166), while $z_\star^{(1)}$ varies with the sensor fault diagnosis approach. Particularly, given (2.21) and (4.181), the signal $z_\star^{(1)}$ is determined as

- in a **distributed** configuration of $\mathcal{O}^{(1,q)}$: $z_\star^{(1)} = y_1^{(2)}$
- in a **decentralized** configuration of $\mathcal{O}^{(1,q)}$: $z_\star^{(1)} = x_{1ref}^{(2)}$

In order to obtain the conditions for ensuring the stability of the proposed nonlinear observer, we follow the procedure described in Section 4.1.1. Taking into account (4.165), (4.167) and (4.172)-(4.173), the estimation error under healthy conditions $\varepsilon_{x_H}^{(1,q)}$ is described by

$$\begin{aligned} \varepsilon_{x_H}^{(1,q)} = & e^{A_L^{(1,q)}t}\varepsilon_{x_H}^{(1,q)}(0) + \int_0^t e^{A_L^{(1,q)}(t-\tau)}Q_1\left(g_2^{(1)}\left(\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1,q)})\right)\right. \\ & \left.+ b^{(1)}\left(\sin(x_1^{(2)}) - \sin(z_\star^{(1)})\right) + \tilde{g}_2^{(1)}\cos(x_1^{(1)})\right)d\tau \\ & - \int_0^t e^{A_L^{(1,q)}(t-\tau)}L^{(1,q)}d_q^{(1)}d\tau \end{aligned} \quad (4.185)$$

where $\hat{x}_H^{(1,q)}$ is the state estimation under healthy conditions. A bound on $|\varepsilon_{x_H}^{(1,q)}|$ can be obtained as

$$\begin{aligned} |\varepsilon_{x_H}^{(1,q)}| \leq & \left| e^{A_L^{(1,q)}t} \right| |\varepsilon_{x_H}^{(1,q)}(0)| + \int_0^t \left| e^{A_L^{(1,q)}(t-\tau)}Q_1 \right| \left(\left| g_2^{(1)} \right| |\varepsilon_{x_{H1}}^{(1,q)}| \right. \\ & \left. + \left| b^{(1)} \right| |\tilde{z}_\star^{(1)} + \tilde{g}_2^{(1)}\cos(x_1^{(1)})| \right) d\tau + \int_0^t \left| e^{A_L^{(1,q)}(t-\tau)}L^{(1,q)} \right| |\bar{d}_q^{(1)}| d\tau. \end{aligned} \quad (4.186)$$

where $\tilde{z}_\star^{(1)}$ varies with the selected sensor fault diagnosis architecture as shown next:

- following a **distributed** state estimation approach, the constant $\bar{z}_\star^{(1)}$ is selected such that $\left| \sin(x_1^{(2)}) - \sin(y_1^{(2)}) \right| \leq \left| x_1^{(2)} - y_1^{(2)} \right| \leq \bar{d}_1^{(2)} = \bar{z}_\star^{(1)}$ if $\bar{d}_1^{(2)} \leq 2$, while if $\bar{d}_1^{(2)} > 2$, then $\bar{z}_\star^{(1)}$ is selected such that $\left| \sin(x_1^{(2)}) - \sin(y_1^{(2)}) \right| \leq \left| \sin(x_1^{(2)}) \right| + \left| \sin(y_1^{(2)}) \right| \leq 2 = \bar{z}_\star^{(1)}$;
- following a **decentralized** state estimation approach, the constant $\bar{z}_\star^{(1)}$ is selected such that $\left| \sin(x_1^{(2)}) - \sin(x_{1ref}^{(2)}) \right| \leq \left| x_1^{(2)} - x_{1ref}^{(2)} \right| \leq \bar{x}_1^{(2)} = \bar{z}_\star^{(1)}$ if $\bar{x}_1^{(2)} \leq 2$, while if $\bar{x}_1^{(2)} > 2$, then $\bar{z}_\star^{(1)}$ is selected such that $\left| \sin(x_1^{(2)}) - \cos(x_{1ref}^{(2)}) \right| \leq \left| \sin(x_1^{(2)}) \right| + \left| \sin(x_{1ref}^{(2)}) \right| \leq 2 = \bar{z}_\star^{(1)}$

In order for $\left| \varepsilon_{x_H}^{(1,q)} \right|$ to be bounded, we need to determine a bound on $\left| \varepsilon_{x_{H1}}^{(1,q)} \right|$, as follows; given that

$$\varepsilon_{x_{H1}}^{(1,q)} = Q_2 \varepsilon_{x_H}^{(1,q)} \quad (4.187)$$

where Q_2 is defined in (4.178), it yields

$$\begin{aligned} \left| \varepsilon_{x_{H1}}^{(1,q)} \right| &\leq \left| Q_2 e^{A_L^{(1,q)} t} \right| \left| \varepsilon_{x_H}^{(1,q)}(0) \right| + \int_0^t \left| Q_2 e^{A_L^{(1,q)}(t-\tau)} Q_1 \right| \left(\left| g_2^{(1)} \right| \left| \varepsilon_{x_{H1}}^{(1,q)} \right| \right. \\ &\quad \left. + \left| b^{(1)} \right| \bar{z}_\star^{(1)} + \bar{g}_2^{(1)} \left| \cos(x_1^{(1)}) \right| \right) d\tau \\ &\quad + \int_0^t \left| Q_2 e^{A_L^{(1,q)}(t-\tau)} L^{(1,q)} \right| \bar{d}_q^{(1)} d\tau. \end{aligned} \quad (4.188)$$

Let us assume that there is a bound $\bar{x}^{(1,q)}$ such that $\left| \varepsilon_{x_H}^{(1,q)}(0) \right| \leq \bar{x}^{(1,q)}$, and select the values of $\rho_0^{(1,q)}$, $\xi_0^{(1,q)}$, $\rho^{(1,q)}$, $\xi^{(1,q)}$, $\rho_d^{(1,q)}$, $\xi_d^{(1,q)}$ such that

$$\left| Q_2 e^{A_L^{(1,q)} t} \right| \leq \rho_0^{(1,q)} e^{-\xi_0^{(1,q)} t}, \quad (4.189)$$

$$\left| Q_2 e^{A_L^{(1,q)} t} Q_1 \right| \leq \rho^{(1,q)} e^{-\xi^{(1,q)} t} \quad (4.190)$$

$$\left| Q_2 e^{A_L^{(1,q)} t} L^{(1,q)} \right| \leq \rho_d^{(1,q)} e^{-\xi_d^{(1,q)} t}. \quad (4.191)$$

Moreover,

$$\begin{aligned}
\bar{g}_2^{(1)} |\cos(x_1^{(1)})| &= \bar{g}_2^{(1)} |\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1,q)}) + \cos(\hat{x}_{H1}^{(1,q)})| \\
&\leq \bar{g}_2^{(1)} |\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1,q)})| + \bar{g}_2^{(1)} |\cos(\hat{x}_{H1}^{(1,q)})| \\
&\leq \bar{g}_2^{(1)} |\varepsilon_{x_{H1}}^{(1,q)}| + \bar{g}_2^{(1)} |\cos(\hat{x}_{H1}^{(1,q)})|. \tag{4.192}
\end{aligned}$$

Using (4.189)-(4.192), we have

$$\begin{aligned}
|\varepsilon_{x_{H1}}^{(1,q)}| &\leq \rho_0^{(1,q)} e^{-\xi_0^{(1,q)} t} \bar{x}^{(1,q)} + \int_0^t \rho^{(1,q)} e^{-\xi^{(1,q)}(t-\tau)} \left(|g_2^{(1)}| |\varepsilon_{x_{H1}}^{(1,q)}| \right. \\
&\quad \left. + |b^{(1)}| \bar{z}_\star^{(1)} + \bar{g}_2^{(1)} |\cos(\hat{x}_{H1}^{(1,q)})| + \bar{g}_2^{(1)} |\varepsilon_{x_{H1}}^{(1,q)}| \right) d\tau \\
&\quad + \int_0^t \rho_d^{(1,q)} e^{-\xi_d^{(1,q)}(t-\tau)} \bar{d}_q^{(1)} d\tau \tag{4.193}
\end{aligned}$$

leading to

$$|\varepsilon_{x_{H1}}^{(1,q)}| \leq E_H^{(1,q)} + \int_0^t \rho^{(1,q)} e^{-\xi^{(1,q)}(t-\tau)} (|g_2^{(1)}| + \bar{g}_2^{(1)}) |\varepsilon_{x_{H1}}^{(1,q)}| d\tau \tag{4.194}$$

with

$$\begin{aligned}
E_H^{(1,q)} &= \rho_0^{(1,q)} e^{-\xi_0^{(1,q)} t} \bar{x}^{(1,q)} + \int_0^t \rho^{(1,q)} e^{-\xi^{(1,q)}(t-\tau)} \bar{g}_2^{(1)} |\cos(\hat{x}_{H1}^{(1,q)})| d\tau \\
&\quad + \frac{\rho_d^{(1,q)}}{\xi_d^{(1,q)}} \bar{d}_q^{(1)} (1 - e^{-\xi_d^{(1,q)} t}) + \frac{\rho^{(1,q)}}{\xi^{(1,q)}} |b^{(1)}| \bar{z}_\star^{(1)} (1 - e^{-\xi^{(1,q)} t}) \tag{4.195}
\end{aligned}$$

Hence, by applying the Bellman-Gronwall Lemma [Ioannou and Sun [1995]], an upper bound on $|\varepsilon_{x_{H1}}^{(1,q)}|$ such that

$$|\varepsilon_{x_{H1}}^{(1,q)}| \leq Z_{H1}^{(1,q)} \tag{4.196}$$

is determined as

$$\begin{aligned}
Z_{H1}^{(1,q)} &= \rho^{(1,q)} (|g_2^{(1)}| + \bar{g}_2^{(1)}) \int_0^t e^{-\left(\xi^{(1,q)} - \rho^{(1,q)} (|g_2^{(1)}| + \bar{g}_2^{(1)})\right)(t-\tau)} E_H^{(1,q)} d\tau \\
&\quad + E_H^{(1,q)} \tag{4.197}
\end{aligned}$$

which is finite if $\xi^{(1,q)} > \rho^{(1,q)} (|g_2^{(1)}| + \bar{g}_2^{(1)})$. Thus, the observer $\mathcal{O}^{(1,q)}$ is stable, if the gain $L^{(1,q)}$ is designed such that $A^{(1)} - L^{(1,q)} C^{(1,q)}$

is stable and there are positive constants $\rho^{(1,q)}, \xi^{(1,q)}$ such that $|Q_2 e^{(A^{(1)} - L^{(1,q)} C^{(1,q)})t} Q_1| \leq \rho^{(1,q)} e^{-\xi^{(1,q)}t}$ and

$$\xi^{(1,q)} > \rho^{(1,q)} (|g_2^{(2)}| + \bar{g}_2^{(2)}). \quad (4.198)$$

Observer Associated with the Modules $\mathcal{M}^{(2,1)}$ and $\mathcal{M}^{(2,2)}$

Taking into account the model of $\Sigma^{(2)}$ in (4.169), the nonlinear observer $\mathcal{O}^{(2,q)}$ associated with the monitoring module $\mathcal{M}^{(2,q)}$ is structured as

$$\begin{aligned} \mathcal{O}^{(2,q)} : \quad \dot{\hat{x}}^{(2,q)} = & A^{(2)} \hat{x}^{(2,q)} + Q_1 \left(g_1^{(2)} u^{(2)} \right) + Q_1 \left(g_2^{(2)} \sin(\hat{x}_1^{(2,q)}) \right) \\ & + Q_1 \left(b^{(2)} \cos(z_\star^{(2)}) \right) + L^{(2,q)} \left(y_q^{(2)} - C_q^{(2)} \hat{x}^{(2,q)} \right) \end{aligned} \quad (4.199)$$

where Q_1 is given in (4.166), while $z_\star^{(2)}$ varies with the sensor fault diagnosis approach. Particularly, given (2.21) and (4.182), the signal $z_\star^{(2)}$ is determined as

- in a **distributed** configuration of $\mathcal{O}^{(2,q)}$: $z_\star^{(2)} = y_1^{(1)}$
- in a **decentralized** configuration of $\mathcal{O}^{(2,q)}$: $z_\star^{(2)} = x_{1ref}^{(1)}$.

Performing the same stability analysis as in Section 4.5.1 for observer $\mathcal{O}^{(1,q)}$, the nonlinear observer $\mathcal{O}^{(2,q)}$ is stable if the gain $L^{(2,q)}$ is chosen such that $A^{(2)} - L^{(2,q)} C^{(2,q)}$ is stable and there are positive constants $\rho^{(2,q)}, \xi^{(2,q)}$ such that $|Q_2 e^{(A^{(2)} - L^{(2,q)} C^{(2,q)})t} Q_1| \leq \rho^{(2,q)} e^{-\xi^{(2,q)}t}$ and $\xi^{(2,q)} > \rho^{(2,q)} (|g_2^{(2)}| + \bar{g}_2^{(2)})$.

4.5.2 Computation of Adaptive Thresholds

Adaptive Thresholds associated with the Modules $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(1,2)}$

Under healthy conditions, the j -th adaptive threshold $\bar{\varepsilon}_{y_{Hj}}^{(1,q)}$ associated with the monitoring module $\mathcal{M}^{(1,q)}$ is designed to bound the residual under healthy conditions, defined as

$$\varepsilon_{y_{H1}}^{(1,1)} = C_1^{(1)} \varepsilon_{x_H}^{(1,1)} + d_1^{(1)} = \begin{bmatrix} 1 & 0 \end{bmatrix} \varepsilon_{x_H}^{(1,1)} + d_1^{(1)}, \quad (4.200)$$

$$\varepsilon_{y_{H2}}^{(1,2)} = C_2^{(1)} \varepsilon_{x_H}^{(1,2)} + d_2^{(1)} = \begin{bmatrix} 0 & 1 \end{bmatrix} \varepsilon_{x_H}^{(1,2)} + d_2^{(1)} \quad (4.201)$$

Taking into account (4.187), (4.195), (4.197) and that $\left| \varepsilon_{y_{Hj}}^{(1,q)} \right| \leq \left| C_j^{(1)} \varepsilon_{x_H}^{(1,q)} \right| + \bar{d}_j^{(1,q)}$, the adaptive threshold for $j = 1$ is determined as

$$\begin{aligned} \bar{\varepsilon}_{y_1}^{(1,1)} &= \rho^{(1,1)} \left(|g_2^{(1)}| + \bar{g}_2^{(1)} \right) \int_0^t e^{-\left(\xi^{(1,1)} - \rho^{(1,1)} \left(|g_2^{(1)}| + \bar{g}_2^{(1)} \right) \right) (t-\tau)} E^{(1,1)} d\tau \\ &\quad + E^{(1,1)} + \bar{d}_1^{(1)}, \end{aligned} \quad (4.202)$$

$$\begin{aligned} E^{(1,1)} &= \rho_0^{(1,1)} e^{-\xi_0^{(1,1)} t} \bar{x}^{(1,1)} + \int_0^t \rho^{(1,1)} e^{-\xi^{(1,1)} (t-\tau)} \bar{g}_2^{(1)} |\cos(\hat{x}_1^{(1,1)})| d\tau \\ &\quad + \frac{\rho_d^{(1,1)}}{\xi_d^{(1,1)}} \bar{d}_1^{(1)} \left(1 - e^{-\xi_d^{(1,1)} t} \right) + \frac{\rho^{(1,1)}}{\xi^{(1,1)}} |b^{(1)}| \bar{z}_\star^{(1)} \left(1 - e^{-\xi^{(1,1)} t} \right) \end{aligned} \quad (4.203)$$

and is implemented as

$$\bar{\varepsilon}_{y_1}^{(1,1)} = \left(\frac{\rho^{(1,1)} (|g_2^{(1)}| + \bar{g}_2^{(1)})}{s + \xi^{(1,1)} - \rho^{(1,1)} (|g_2^{(1)}| + \bar{g}_2^{(1)})} + 1 \right) \left[E^{(1,1)} \right] + \bar{d}_1^{(1)}, \quad (4.204)$$

$$\begin{aligned} E^{(1,1)} &= \frac{\rho^{(1,1)}}{\xi^{(1,1)}} |b^{(1)}| \bar{z}_\star^{(1)} (1 - e^{-\xi^{(1,1)} t}) + \frac{\rho^{(1,1)}}{s + \xi^{(1,1)}} \left[\bar{g}_2^{(1)} |\cos(\hat{x}_1^{(1,q)})| \right] \\ &\quad + \rho_0^{(1,1)} e^{-\xi_0^{(1,1)} t} \bar{x}^{(1,1)} + \frac{\rho_d^{(1,1)}}{\xi_d^{(1,1)}} \bar{d}_1^{(1)} (1 - e^{-\xi_d^{(1,1)} t}). \end{aligned} \quad (4.205)$$

The adaptive threshold $\bar{\varepsilon}_{y_2}^{(1,2)}$ is computed as follows; taking into account (4.201) and the solution (4.185) for $q = 2$, a bound on $\left| \varepsilon_{y_{H2}}^{(1,2)} \right|$ is defined as:

$$\begin{aligned} \left| \varepsilon_{y_{H2}}^{(1,2)} \right| &\leq \left| C_2^{(1)} e^{A_L^{(1,2)} t} \right| \left| \varepsilon_{x_H}^{(1,2)}(0) \right| + \int_0^t \left| C_2^{(1)} e^{A_L^{(1,2)} (t-\tau)} L^{(1,2)} \right| \bar{d}_2^{(1)} d\tau \\ &\quad + \int_0^t \left| C_2^{(1)} e^{A_L^{(1,2)} (t-\tau)} Q_1 \right| \left(\left| g_2^{(1)} (\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1,2)})) \right| \right. \\ &\quad \left. + |b^{(1)}| \bar{z}_\star^{(1)} + \left| \bar{g}_2^{(1)} \cos(x_1^{(1)}) \right| \right) d\tau + \bar{d}_2^{(1)}. \end{aligned} \quad (4.206)$$

Then, by selecting $\alpha_0^{(1,2)}$, $\zeta_0^{(1,2)}$, $\alpha^{(1,2)}$, $\zeta^{(1,2)}$, $\alpha_d^{(1,2)}$, $\zeta_d^{(1,2)}$ such that

$$\left| C_2^{(1)} e^{A_L^{(1,2)} t} \right| \leq \alpha_0^{(1,2)} e^{-\zeta_0^{(1,2)} t}, \quad (4.207)$$

$$\left| C_2^{(1)} e^{A_L^{(1,2)} t} Q_1 \right| \leq \alpha^{(1,q)} e^{-\zeta^{(1,q)} t}, \quad (4.208)$$

$$\left| C_2^{(1)} e^{A_L^{(1,2)} t} L^{(1,2)} \right| \leq \alpha_d^{(1,2)} e^{-\zeta_d^{(1,2)} t}, \quad (4.209)$$

the adaptive threshold $\bar{\varepsilon}_{y_2}^{(1,2)}$ is determined as

$$\begin{aligned} \bar{\varepsilon}_{y_2}^{(1,2)} = & \alpha_0^{(1,2)} e^{-\zeta_0^{(1,2)} t} \bar{x}^{(1,2)} + \int_0^t \alpha^{(1,2)} e^{-\zeta^{(1,2)}(t-\tau)} \left(\left(|g_2^{(1)}| + \bar{g}_2^{(1)} \right) Z_1^{(1,2)} \right. \\ & \left. + \bar{g}_2^{(1)} |\cos(\hat{x}_1^{(1,2)})| \right) d\tau + \frac{\alpha_d^{(1,2)}}{\zeta_d^{(1,2)}} \bar{d}_2^{(1)} (1 - e^{-\zeta_d^{(1,2)} t}) \\ & + \frac{\alpha^{(1,2)}}{\zeta^{(1,2)}} |b^{(1)}| \bar{z}_\star^{(1)} (1 - e^{-\zeta^{(1,2)} t}) + \bar{d}_2^{(1)} \end{aligned} \quad (4.210)$$

$$\begin{aligned} Z_1^{(1,2)} = & \rho^{(1,2)} \left(|g_2^{(1)}| + \bar{g}_2^{(1)} \right) \int_0^t e^{-\left(\xi^{(1,2)} - \rho^{(1,2)} \left(|g_2^{(1)}| + \bar{g}_2^{(1)} \right) \right) (t-\tau)} E^{(1,2)} d\tau \\ & + E^{(1,2)} \end{aligned} \quad (4.211)$$

$$\begin{aligned} E^{(1,2)} = & \rho_0^{(1,2)} e^{-\xi_0^{(1,2)} t} \bar{x}^{(1,2)} + \int_0^t \rho^{(1,2)} e^{-\xi^{(1,2)}(t-\tau)} \bar{g}_2^{(1)} |\cos(\hat{x}_1^{(1,2)})| d\tau \\ & + \frac{\rho_d^{(1,2)}}{\xi_d^{(1,2)}} \bar{d}_2^{(1)} (1 - e^{-\xi_d^{(1,2)} t}) + \frac{\rho^{(1,2)}}{\xi^{(1,2)}} |b^{(1)}| \bar{z}_\star^{(1)} (1 - e^{-\xi^{(1,2)} t}) \end{aligned} \quad (4.212)$$

The implementation of the adaptive threshold $\bar{\varepsilon}_{y_2}^{(1,2)}$ can be realized using linear filters; i.e.,

$$\begin{aligned} \bar{\varepsilon}_{y_2}^{(1,2)} = & \alpha_0^{(1,2)} e^{-\zeta_0^{(1,2)} t} \bar{x}^{(1,2)} + \frac{\alpha_d^{(1,2)}}{\zeta_d^{(1,2)}} \bar{d}_2^{(1)} (1 - e^{-\zeta_d^{(1,2)} t}) \\ & + \frac{\alpha^{(1,2)}}{\zeta^{(1,2)}} |b^{(1)}| \bar{z}_\star^{(1)} (1 - e^{-\zeta^{(1,2)} t}) + \frac{\alpha^{(1,2)}}{s + \zeta^{(1,2)}} \left[\bar{g}_2^{(1)} |\cos(\hat{x}_1^{(1,2)})| \right. \\ & \left. + \left(|g_2^{(1)}| + \bar{g}_2^{(1)} \right) Z_1^{(1,2)} \right] + \bar{d}_2^{(1)} \end{aligned} \quad (4.213)$$

with

$$Z_1^{(1,2)} = \left(\frac{\rho^{(1,2)}(|g_2^{(1)}| + \bar{g}_2^{(1)})}{s + \xi^{(1,2)} - \rho^{(1,2)}(|g_2^{(1)}| + \bar{g}_2^{(1)})} + 1 \right) [E^{(1,2)}] \quad (4.214)$$

$$\begin{aligned} E^{(1,2)} = & \frac{\rho^{(1,2)}}{\xi^{(1,2)}} |b^{(1)}| \bar{z}_\star^{(1)} (1 - e^{-\xi^{(1,2)}t}) + \frac{\rho^{(1,2)}}{s + \xi^{(1,2)}} [\bar{g}_2^{(1)} |\cos(\hat{x}_1^{(1,2)})|] \\ & + \rho_0^{(1,2)} e^{-\xi_0^{(1,2)}t} \bar{x}^{(1,2)} + \frac{\rho_d^{(1,2)}}{\xi_d^{(1,2)}} \bar{d}_2^{(1)} (1 - e^{-\xi_d^{(1,2)}t}). \end{aligned} \quad (4.215)$$

Adaptive thresholds associated with the modules $\mathcal{M}^{(2,1)}$ and $\mathcal{M}^{(2,2)}$

Following the same procedure presented in Section 4.5.2, the adaptive thresholds $\bar{\varepsilon}_{y_1}^{(2,1)}$ and $\bar{\varepsilon}_{y_2}^{(2,2)}$ associated with the module $\mathcal{M}^{(2,1)}$ can be implemented as:

$$\bar{\varepsilon}_{y_1}^{(2,1)} = \left(\frac{\rho^{(2,1)}(|g_2^{(2)}| + \bar{g}_2^{(2)})}{s + \xi^{(2,1)} - \rho^{(2,1)}(|g_2^{(2)}| + \bar{g}_2^{(2)})} + 1 \right) [E^{(2,1)}] + \bar{d}_1^{(2)}, \quad (4.216)$$

$$\begin{aligned} E^{(2,1)} = & \frac{\rho^{(2,1)}}{\xi^{(2,1)}} |b^{(2)}| \bar{z}_\star^{(2)} (1 - e^{-\xi^{(2,1)}t}) + \frac{\rho^{(2,1)}}{s + \xi^{(2,1)}} [\bar{g}_2^{(2)} |\sin(\hat{x}_1^{(2,q)})|] \\ & + \rho_0^{(2,1)} e^{-\xi_0^{(2,1)}t} \bar{x}^{(2,1)} + \frac{\rho_d^{(2,1)}}{\xi_d^{(2,1)}} \bar{d}_1^{(2)} (1 - e^{-\xi_d^{(2,1)}t}), \end{aligned} \quad (4.217)$$

$$\begin{aligned} \bar{\varepsilon}_{y_2}^{(2,2)} = & \alpha_0^{(2,2)} e^{-\zeta_0^{(2,2)}t} \bar{x}^{(2,2)} + \frac{\alpha_d^{(2,2)}}{\zeta_d^{(2,2)}} \bar{d}_2^{(2)} (1 - e^{-\zeta_d^{(2,2)}t}) \\ & + \frac{\alpha^{(2,2)}}{\zeta^{(2,2)}} |b^{(2)}| \bar{z}_\star^{(2)} (1 - e^{-\zeta^{(2,2)}t}) + \frac{\alpha^{(2,2)}}{s + \zeta^{(2,2)}} [\bar{g}_2^{(2)} |\sin(\hat{x}_1^{(2,2)})|] \\ & + \left(|g_2^{(2)}| + \bar{g}_2^{(2)} \right) Z_1^{(2,2)} + \bar{d}_2^{(2)} \end{aligned} \quad (4.218)$$

$$Z_1^{(2,2)} = \left(\frac{\rho^{(2,2)}(|g_2^{(2)}| + \bar{g}_2^{(2)})}{s + \xi^{(2,2)} - \rho^{(2,2)}(|g_2^{(2)}| + \bar{g}_2^{(2)})} + 1 \right) [E^{(2,2)}] \quad (4.219)$$

$$\begin{aligned} E^{(2,2)} = & \frac{\rho^{(2,2)}}{\xi^{(2,2)}} |b^{(2)}| \bar{z}_\star^{(2)} (1 - e^{-\xi^{(2,2)}t}) + \frac{\rho^{(2,2)}}{s + \xi^{(2,2)}} [\bar{g}_2^{(2)} |\sin(\hat{x}_1^{(2,2)})|] \\ & + \rho_0^{(2,2)} e^{-\xi_0^{(2,2)}t} \bar{x}^{(2,2)} + \frac{\rho_d^{(2,2)}}{\xi_d^{(2,2)}} \bar{d}_2^{(2)} (1 - e^{-\xi_d^{(2,2)}t}). \end{aligned} \quad (4.220)$$

4.5.3 Sensor Fault Detection Decision Logic

The monitoring module $\mathcal{M}^{(I,q)}$ detects the presence of sensor faults, when the analytical redundancy relation $\mathcal{E}^{(I,q)} = \mathcal{E}_q^{(I)}$, where $\mathcal{E}_q^{(I)}$ is defined as

$$\mathcal{E}_q^{(I)} : \left| \varepsilon_{y_q}^{(I,q)} \right| \leq \bar{\varepsilon}_{y_q}^{(I,q)} \quad (4.221)$$

is not satisfied and the decision function (output of the module $\mathcal{M}^{(I,q)}$) is determined as

$$D^{(I,q)}(t) = D_q^{(I,q)}(t) = \begin{cases} 0, & \text{if } t < T_q^{(I,q)} \\ 1, & \text{otherwise} \end{cases} \quad (4.222)$$

$$T_q^{(I,q)} = \min \left\{ t : \left| \varepsilon_{y_q}^{(I,q)}(t) \right| - \bar{\varepsilon}_{y_q}^{(I,q)}(t) > 0 \right\} \quad (4.223)$$

In a *decentralized* architecture, when $D^{(I,q)}(t) = 1$, it is inferred that a sensor fault affecting the sensor $\mathcal{S}_q^{(I)}$ is guaranteed to have occurred. In a *distributed* architecture, when $D^{(1,q)}(t) = 1$, it is inferred that either $\mathcal{S}_q^{(1)}$ or $\mathcal{S}_1^{(2)}$ or both are guaranteed to be faulty, or equivalently a sensor fault in the set $\{\mathcal{S}_q^{(1)}, \mathcal{S}_1^{(2)}\}$ is guaranteed to have occurred. Similarly, when $D^{(2,q)}(t) = 1$, it is inferred that either $\mathcal{S}_q^{(2)}$ or $\mathcal{S}_1^{(1)}$ or both are faulty, or equivalently a sensor fault in the set $\{\mathcal{S}_q^{(2)}, \mathcal{S}_1^{(1)}\}$ is guaranteed to have occurred.

4.5.4 Simulation Results

This section presents the simulation results of the application of the distributed and decentralized sensor fault diagnosis method presented in Sections 4.5.1, 4.5.2 and 4.5.3 to the two nonlinear interconnected subsystems given in (4.165) and (4.169) with the following parameters: $a_1^{(1)} = 9$, $a_2^{(1)} = -2$, $g_1^{(1)} = 1$, $g_2^{(1)} = 1$, $b^{(1)} = 3$, $\tilde{g}_2^{(1)} = 3\%g_2^{(1)} = 0.03$, $\bar{g}_2^{(1)} = 5\%g_2^{(1)} = 0.05$, $a_1^{(2)} = 12$, $a_2^{(2)} = -11$, $g_1^{(2)} = 1$, $g_2^{(2)} = 2$, $b^{(2)} = 12$, $\tilde{g}_2^{(2)} = 3\%g_2^{(2)} = 0.06$ and $\bar{g}_2^{(2)} = 5\%g_2^{(2)} = 0.1$. The simulated reference signals are $x_{1ref}^{(1)}(t) = x_{1ref}^{(2)}(t) = x_1^o \sin(t)$ with $x_1^o = 2\pi$. In the distributed sensor fault diagnosis architecture, the control inputs

$u^{(1)}$ and $u^{(2)}$ are implemented as:

$$u^{(1)} = \frac{1}{g_1^{(1)}} \left[-a_1^{(1)} y_1^{(1)} - a_2^{(1)} y_2^{(1)} - g_2^{(1)} \cos(y_1^{(1)}) - b^{(1)} \sin(y_1^{(2)}) + v^{(1)} \right], \quad (4.224)$$

$$u^{(2)} = \frac{1}{g_1^{(2)}} \left[-a_1^{(2)} y_1^{(2)} - a_2^{(2)} y_2^{(2)} - g_2^{(2)} \sin(y_1^{(2)}) - b^{(2)} \cos(y_1^{(1)}) + v^{(2)} \right], \quad (4.225)$$

$$v^{(I)} = \ddot{x}_{1ref}^{(I)} - \kappa_0 e_1^{(I)} - \kappa_1 e_2^{(I)} \quad (4.226)$$

$$e_1^{(I)} = y_1^{(I)} - x_{1ref}^{(I)} \quad (4.227)$$

$$e_2^{(I)} = y_2^{(I)} - \dot{x}_{1ref}^{(I)} \quad (4.228)$$

In the decentralized sensor fault diagnosis architecture, the control inputs are implemented as

$$u^{(1)} = \frac{1}{g_1^{(1)}} \left[-a_1^{(1)} y_1^{(1)} - a_2^{(1)} y_2^{(1)} - g_2^{(1)} \cos(y_1^{(1)}) - b^{(1)} \sin(x_{1ref}^{(2)}) + v^{(1)} \right], \quad (4.229)$$

$$u^{(2)} = \frac{1}{g_1^{(2)}} \left[-a_1^{(2)} y_1^{(2)} - a_2^{(2)} y_2^{(2)} - g_2^{(2)} \sin(y_1^{(2)}) - b^{(2)} \cos(x_{1ref}^{(1)}) + v^{(2)} \right], \quad (4.230)$$

where $v^{(1)}$ is implemented as in (4.226)-(4.228). In both controller configurations, the controller gains are $\kappa_0^{(I)} = 8$ and $\kappa_1^{(I)} = 9$, $I = 1, 2$. The sensor measurements are corrupted by uniformly random noise satisfying Assumption 4, with $d_q^{(I)} = 5\% x_1^\circ = 0.3142$. Under healthy conditions, the states of both subsystems when the distributed and decentralized control scheme is applied are shown in Fig. 4.4.

The implementation of the distributed and decentralized sensor fault diagnosis method presented in Sections 4.5.1, 4.5.2 and 4.5.3 is realized using the parameters shown in Table 4.1 and the aforementioned system characteristics. It is noted that the adaptive thresholds in the distributed sensor fault diagnosis scheme have been designed using $z_\star^{(1)} = \bar{d}_1^{(2)}$ and $z_\star^{(2)} = \bar{d}_1^{(1)}$, while the adaptive thresholds in the decentralized sensor fault diagnosis scheme have been designed using $z_\star^{(1)} = z_\star^{(2)} = 2$, since the bound $\bar{x}_1^{(I)}$ such that $|x_1^{(I)}(t) - x_{1ref}^{(I)}(t)| \leq \bar{x}_1^{(I)}$

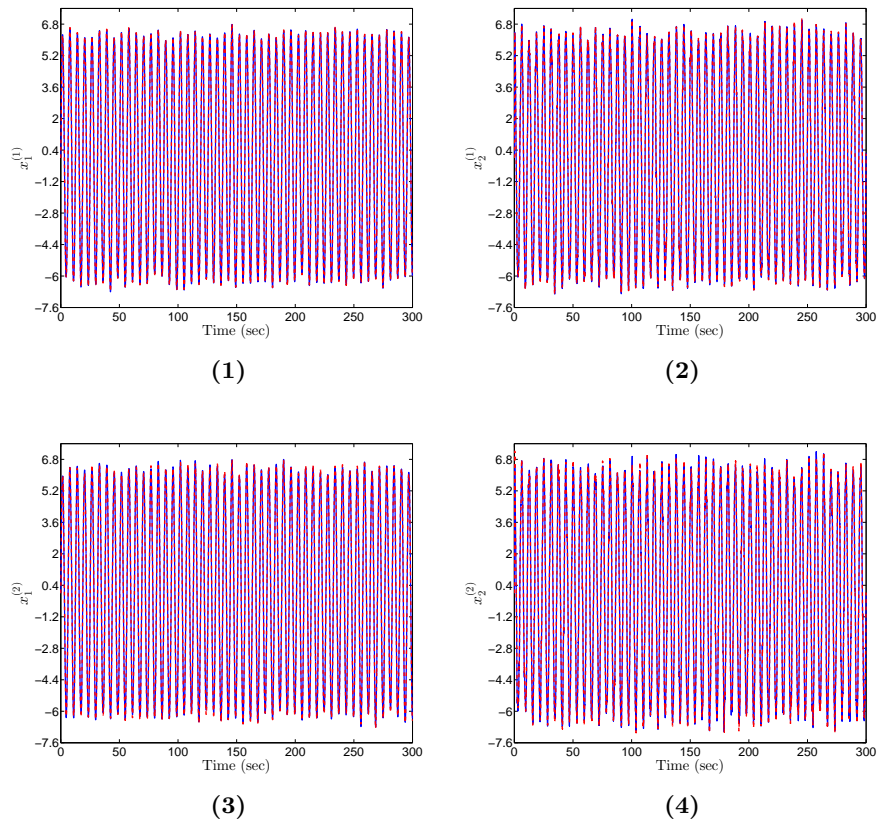


Figure 4.3: States of subsystems $\Sigma^{(1)}$ (subfigures (1),(2)) and $\Sigma^{(2)}$ (subfigures (3),(4)) operating in distributed (blue,solid line) and decentralized (green, dashed line) control scheme.

		$\mathcal{M}^{(I,q)}$			
		(1, 1)	(1, 2)	(2, 1)	(2, 2)
$\mathcal{O}^{(I,q)}$	$\hat{x}^{(I,q)}(0)$	$[0, 0]^\top$	$[0, 0]^\top$	$[0, 0]^\top$	$[0, 0]^\top$
	$\text{eig}(A_L^{(I,q)})$	$\{-5, -6\}$	$\{-10, -11\}$	$\{-14, -15\}$	$\{-14, -15\}$
	$L^{(I,q)}$	$[9, 21]^\top$	$[13.22, 19]^\top$	$[18, 24]^\top$	$[18.5, 18]^\top$
$\bar{\varepsilon}_{y_q}^{(I,q)}$	$\rho_0^{(I,q)}$	1.1196	4.6383	1.6568	5.0804
	$\xi_0^{(I,q)}$	3.75	7	10.5	10.5
	$\rho^{(I,q)}$	0.125	0.8333	0.0667	0.875
	$\xi^{(I,q)}$	2.5	5	7	7
	$\rho_d^{(I,q)}$	10.8055	19.9720	21.0897	21.0155
	$\xi_d^{(I,q)}$	4.75	7.5	13.3	13.3
	$\alpha_0^{(I,q)}$	-	1.8409	-	1.8
	$\zeta_0^{(I,q)}$	-	7.5	-	10.5
	$\alpha^{(I,q)}$	-	1	-	1
	$\zeta^{(I,q)}$	-	5	-	7
	$\alpha_d^{(I,q)}$	-	19.4554	-	18.2777
	$\zeta_d^{(I,q)}$	-	7.5	-	13.3

Table 4.1: Design parameters used in the decentralized and distributed sensor fault diagnosis architecture in the illustrative example.

under worst-case sensor fault conditions is larger than 2 for both $I = 1$ and $I = 2$. Indicatively, by simulating the closed-loop systems where all sensors in both subsystems become abruptly faulty at $t = 150\text{sec}$ with a fault magnitude $\phi_q^{(I)} = 50\%x_1^\circ$ for $q = 1, 2$, the bound on the amplitude of the magnitude of $x_1^{(I)}(t) - x_{1ref}^{(I)}(t)$ is around 10 for both subsystems as, shown in Fig. 4.3.

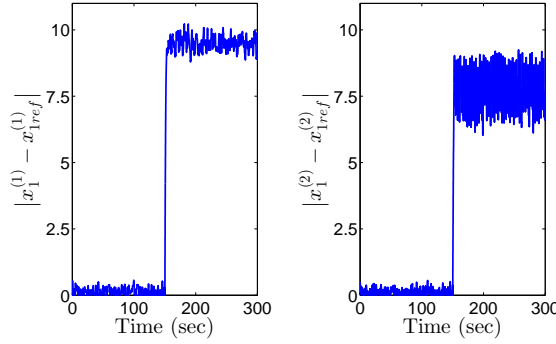


Figure 4.4: Magnitude of the tracking error $x_1^{(I)}(t) - x_{1ref}^{(I)}(t)$ for $I = 1, 2$, when all sensors in both subsystems become abruptly faulty at $t = 150\text{ sec}$ with fault magnitude $\phi_q^{(I)} = 50\%x_1^\circ = \pi$ for $q = 1, 2$

We have simulated several sensor fault scenarios with the single occurrence of abrupt, permanent fault whose evolution rate is $\kappa_{f_q}^{(I)} = 10^4$, time of fault occurrence is $T_{f_q}^{(I)} = 150\text{ sec}$ and fault function is an offset (see Fig. 2.2), that is, $\phi_q^{(I)}(t) = 25\%x_1^\circ = \pi/2$, for all $I, q = 1, 2$. The simulated scenarios are shown below:

Sensor fault scenario 1: Single occurrence of sensor fault $f_1^{(1)}$

Sensor fault scenario 2: Single occurrence of sensor fault $f_1^{(2)}$

Sensor fault scenario 3: Single occurrence of sensor fault $f_2^{(1)}$

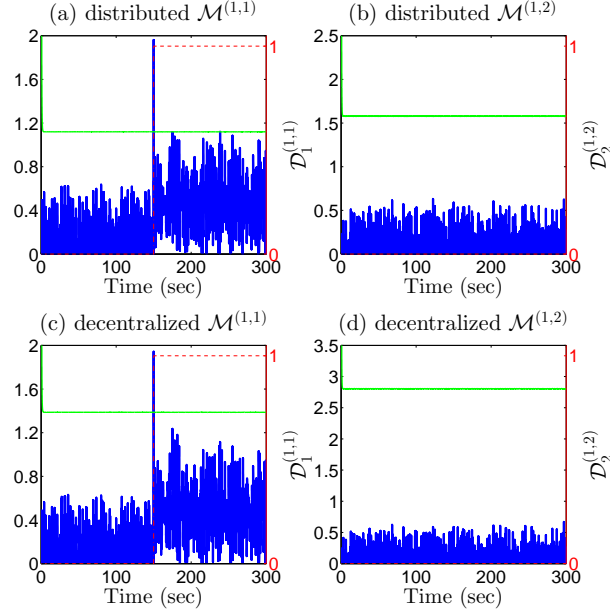
Sensor fault scenario 4: Single occurrence of sensor fault $f_2^{(2)}$

Figures 4.5 and 4.6 illustrate the sensor fault detection decision-making process of the agents $\mathcal{M}^{(1)}$ (subfigures 4.5-1 and 4.6-1) and $\mathcal{M}^{(2)}$ (subfigures 4.5-2 and 4.6-2) for the simulated scenarios 1 and 2, respectively.

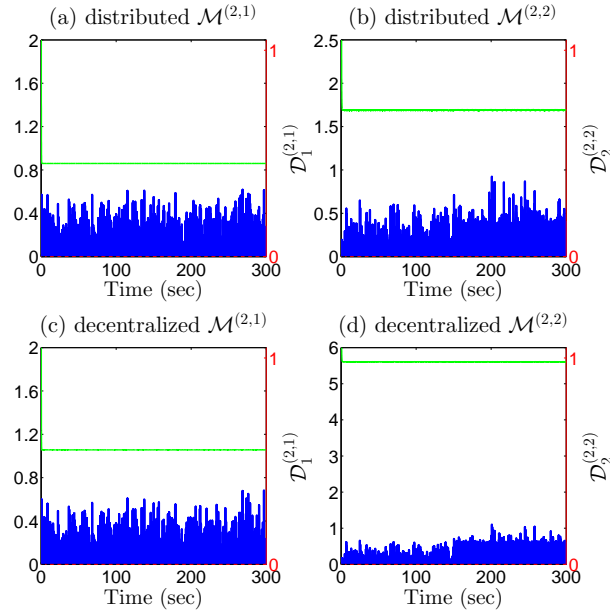
Figure 4.7 presents the sensor fault detection decision-making process of the agent $\mathcal{M}^{(1)}$ for the simulated scenario 3 (Fig. 4.7-1), and of the agent $\mathcal{M}^{(2)}$ for the simulated scenario 4 (Fig. 4.7-2). In every subfigure, we present the temporal evolution of the magnitude of the residual $|\varepsilon_{y_q}^{(I,q)}(t)|$ (blue line), the adaptive threshold $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (green line) and the decision function $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$. Based on Fig. 4.5-1, the monitoring module $\mathcal{M}^{(1,1)}$ detects sensor fault $f_1^{(1)}$ just after its occurrence in both architectures (distributed and decentralized). Note that the adaptive threshold $\bar{\varepsilon}_{y_2}^{(1,2)}$ is much lower in the distributed scheme than in the decentralized scheme. This difference stems from the difference of the values of $z_\star^{(I)}$ and $\bar{z}_\star^{(I)}$ (see Section 4.5.1), as well as the implementation and the design parameters for the observer and the adaptive threshold of the modules $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(1,2)}$ (see Table 4.1).

Based on Fig. 4.5-2, the occurrence of sensor fault $f_1^{(1)}$ was not detectable from the modules $\mathcal{M}^{(2,1)}$ (Fig. 4.5-2(a)) and $\mathcal{M}^{(2,2)}$ (Fig. 4.5-2(b)) of the second monitoring agent, although the measurements of sensor $\mathcal{S}^{(1)}\{1\}$ are used in the distributed state estimation performed by the modules $\mathcal{M}^{(2,1)}$ and $\mathcal{M}^{(2,2)}$. As in the case of modules $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(1,2)}$, the adaptive thresholds of $\mathcal{M}^{(2,1)}$ and $\mathcal{M}^{(2,2)}$ computed following the distributed approach are lower than the adaptive thresholds that are computed following a decentralized approach. This fact is more obvious in the case of adaptive threshold $\bar{\varepsilon}_{y_2}^{(2,2)}$, which is much lower in the distributed scheme than in the decentralized scheme. Moreover, comparing the adaptive thresholds $\bar{\varepsilon}_{y_2}^{(1,2)}$ and $\bar{\varepsilon}_{y_2}^{(2,2)}$ computed in a decentralized way (see Figs. 4.5-1(b) and 4.5-2(b)), it can be observed that $\bar{\varepsilon}_{y_2}^{(2,2)}$ is much larger than $\bar{\varepsilon}_{y_2}^{(1,2)}$. This is not the case when the adaptive thresholds $\bar{\varepsilon}_{y_2}^{(1,2)}$ and $\bar{\varepsilon}_{y_2}^{(2,2)}$ are computed in a distributed way. This is due to the fact that $\frac{\alpha^{(2,2)}}{\zeta^{(2,2)}} \left| b^{(2)} \right| \bar{z}_\star^{(2)}$ in (4.218) (correspondingly for (4.220)) is much larger when $z_\star^{(2)} = 2$ following a decentralized computation than when $z_\star^{(2)} = 0.3142$ following a distributed computation, while $\frac{\alpha^{(2,2)}}{\zeta^{(2,2)}} \left| b^{(2)} \right| \bar{z}_\star^{(2)}$ is much larger than $\frac{\alpha^{(1,2)}}{\zeta^{(1,2)}} \left| b^{(1)} \right| \bar{z}_\star^{(1)}$ in (4.213) (correspondingly for (4.215)) where $b^{(1)} = 3$ while $b^{(1)} = 12$.

According to Fig. 4.6-2, in the second simulated scenario the module

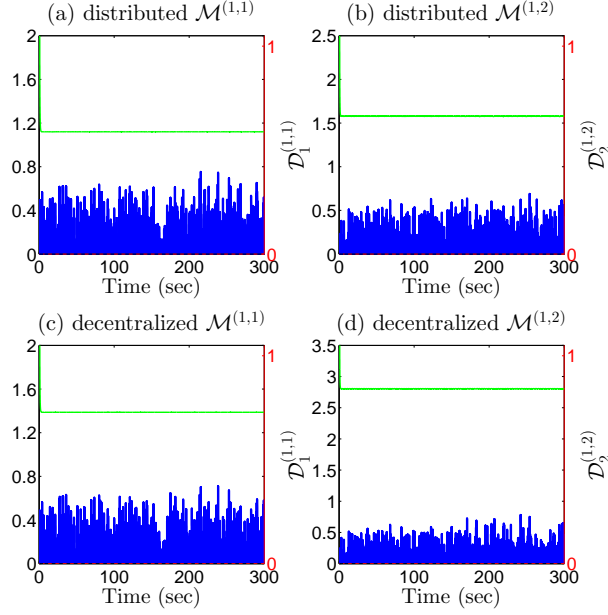


(1) Agent $\mathcal{M}^{(1)}$ deployed in a distributed and decentralized scheme.

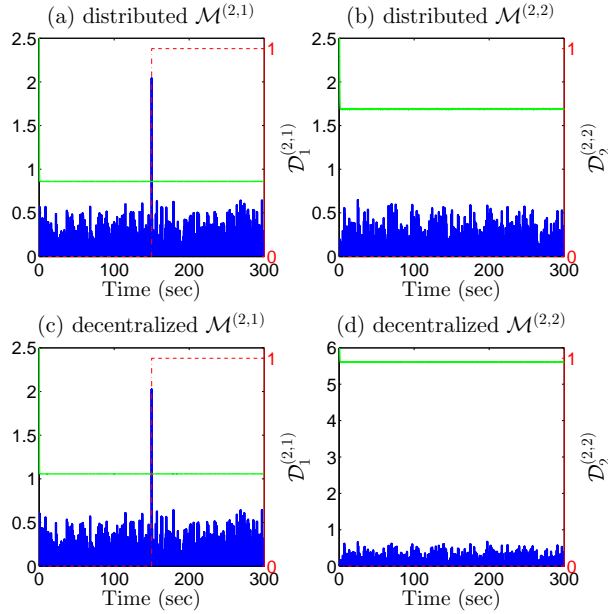


(2) Agent $\mathcal{M}^{(2)}$ deployed in a distributed and decentralized scheme.

Figure 4.5: Sensor fault detection for the simulated scenario 1. Subfigures 1(a)-2(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (green line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.



(1) Agent $\mathcal{M}^{(1)}$ deployed in a distributed and decentralized scheme.



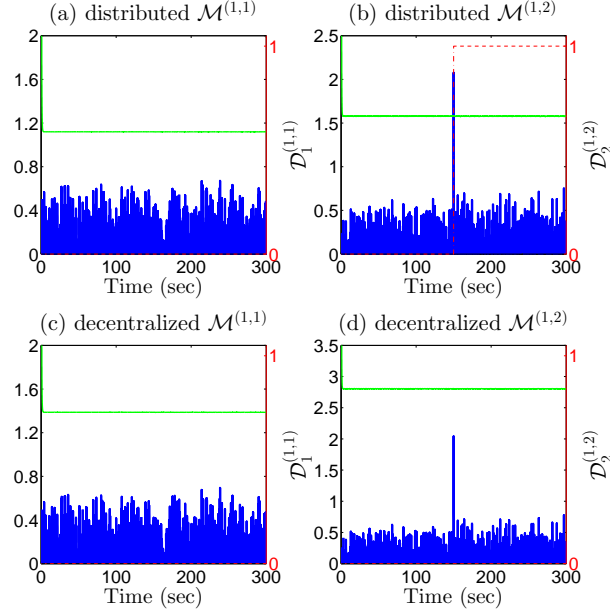
(2) Agent $\mathcal{M}^{(2)}$ deployed in a distributed and decentralized scheme.

Figure 4.6: Sensor fault detection for the simulated scenario 2. Subfigures 1(a)-2(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (green line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.

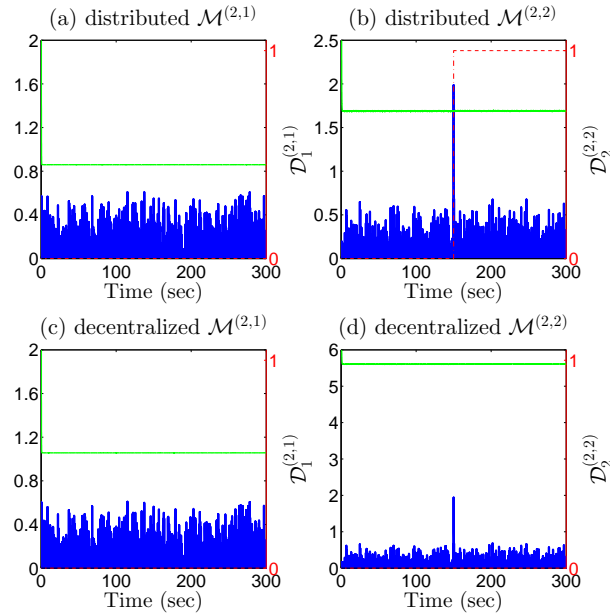
$\mathcal{M}^{(2,1)}$, which has been designed to be sensitive to sensor fault $f_1^{(2)}$ and $f_1^{(1)}$ successfully detected the occurrence of $f_1^{(2)}$ in both architectures, while the module $\mathcal{M}^{(2,2)}$ is insensitive to $f_1^{(2)}$. On the contrary, the effects of $f_1^{(2)}$ on the the residuals and adaptive thresholds of $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(1,2)}$ were very weak to be detectable. Based on Fig. 4.7-1, in the third simulated scenario, the module $\mathcal{M}^{(1,2)}$ succeed to detect the occurrence of sensor fault $f_2^{(1)}$ only within a distributed sensor fault diagnosis scheme. In the decentralized scheme, the adaptive threshold $\bar{\varepsilon}_{y_2}^{(1,2)}$ is much larger than the one in the distributed scheme, and it is difficult for the magnitude of the residual to exceed the threshold. Similar results are obtained in the forth simulated scenario, where the occurrence of the sensor fault $f_2^{(2)}$ was detected by the module $\mathcal{M}^{(2,2)}$ designed in a distributed way, but the sensor fault occurrence was missed in the decentralized architecture.

4.6 Summary and Discussion

in this chapter, we presented the sensor fault detection process of the distributed or decentralized agents that monitor the healthy status of the sensors in each of the interconnected physical system in a network of CPS, as well as of the centralized agent that monitors all sensors of the network of CPS. In all diagnostic architectures (distributed, decentralized, centralized) the sensor fault detection of the monitoring agent is executed through several monitoring modules, which result after the decomposition of the monitoring of the sensors. The design of the monitoring module $\mathcal{M}^{(I,q)}$ in a distributed or decentralized configuration is summarized in Table 4.2. The basic steps for detecting sensor faults are the generation of residuals, the computation of adaptive thresholds and the sensor fault detection decision logic. The residual generation is realized based on a nonlinear (linear) observer, whose structure varies with the diagnostic architecture. Based on the conditions for guaranteeing the stability of the state estimation error dynamics (Theorems 4.1 and 4.3), we may infer that the use of a monitoring agent dedicated to each of the interconnected subsystems (non-centralized approach) may facilitate the design of a nonlinear observer compared to the de-



(1) Agent $\mathcal{M}^{(1)}$ deployed in a distributed and decentralized scheme.



(2) Agent $\mathcal{M}^{(2)}$ deployed in a distributed and decentralized scheme.

Figure 4.7: Sensor fault detection for the simulated scenario 3 (subfigure 1) and 4 (subfigure 1). Subfigures 1(a)-2(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (green line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.

sign of the single agent that monitors the network of CPS, handling it as a monolithic system.

Comparing the bounds on the state estimation error for the non-centralized architectures, we can infer that the transmission of sensor information between the agents (distributed scheme) reduces the state estimation error. The conditions for a stable observer used by the monitoring module can be less strict, if the subsystem has linear dynamics or the nonlinear dynamics are functions of a linear combination of states that are measured by the local sensor set monitored by the module. In general, the formulated observer stability conditions ensures that the adaptive thresholds are finite bounds for the residuals, implying that the associated set of analytical redundancy relations is satisfied under healthy conditions (robustness), and is sensitive to subsets of local and propagated faults (structured fault sensitivity).

Following a distributed or decentralized approach, the sensor fault diagnosis method can be scalable. Particularly, if a new CPS is connected with the network of cyber-physical systems (e.g. an extra mobile robot is incorporated in a swarm of robots, an additional vehicle is included in the automated highway system) a new monitoring agent can be plugged-in to the non-centralized diagnostic scheme [Boem et al. [2015b]]. At the same time, only the monitoring agents of the CPS that will be interconnected with the new CPS have to be modified by changing the interconnection function $h^{(I)}$ and bound on the modeling uncertainty $\bar{\eta}^{(I)}$ along with their arguments, i.e. the measured or unmeasurable states of the new CPS, as well as the Lipschitz constant λ_{h_I} . Thus, the observer $\mathcal{O}^{(I,q)}$ and adaptive thresholds $\bar{\varepsilon}_{y_j}^{(I,q)}$ will be modified for all monitoring modules of the agents that monitor the CPS being in the neighborhood of the new CPS. The flexibility of a non-centralized architecture can be exploited also in the case that a CPS should be unplugged in order to mitigate the effects of a severe fault. In a centralized approach, the inclusion of a new CPS will lead to the modification of the the nonlinear dynamics γ and bound on the modeling uncertainty $\bar{\eta}$, and consequently the modification of the observer $\mathcal{O}^{(q)}$ and adaptive thresholds $\bar{\varepsilon}_{y_j}^{(q)}$ for all monitoring modules.

Table 4.2: Design of the monitoring module $\mathcal{M}^{(I,q)}$ in a distributed and decentralized configuration for $I \in \{1, \dots, M\}$ and $q \in \{1, \dots, N_I\}$ ($\Sigma^{(I)}$: subsystem, $\mathcal{S}^{(I)}$: sensor set, $\mathcal{S}^{(I)}\{j\}$: j -th sensor, $\mathcal{S}^{(I,q)}$: local sensor set, $\mathcal{S}_z^{(I)}$: set of neighboring sensors).

$\Sigma^{(I)}$	$\dot{x}^{(I)} = A^{(I)}x^{(I)} + \gamma^{(I)}(x^{(I)}, u^{(I)}) + h^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}) + \eta^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t)$		
$\mathcal{S}^{(I)}$	$y^{(I)} = C^{(I)}x^{(I)} + d^{(I)} + f^{(I)}$		
$\mathcal{S}^{(I,q)}$	$y^{(I,q)} = C^{(I,q)}x^{(I)} + d^{(I,q)} + f^{(I,q)}$		
$\mathcal{S}_z^{(I)}$	$y_z^{(I)} = C_z^{(I)}z^{(I)} + d_z^{(I)} + f_z^{(I)}$		
Inputs of $\mathcal{M}^{(I,q)}$	$u^{(I)}$	Distributed	$y_z^{(I)}, \zeta_r^{(I)}$
	$y^{(I,q)}$	Decentralized	$z_r^{(I)}$
Outputs of $\mathcal{M}^{(I,q)}$			
Design Parameters of $\mathcal{M}^{(I,q)}$	$L^{(I,q)}, \bar{x}^{(I,q)}, \bar{d}^{(I,q)}, \bar{d}_j^{(I)}, \rho^{(I,q)}, \xi^{(I,q)}$	Distributed	$\bar{\zeta}_r^{(I)}, \bar{d}_z^{(I)}$
	$\rho_d^{(I,q)}, \xi_d^{(I,q)}, \alpha_j^{(I,q)}, \zeta_j^{(I,q)}, \alpha_{d_j}^{(I,q)}, \zeta_{d_j}^{(I,q)}$	Decentralized	$\bar{z}_r^{(I)}$
$\mathcal{O}^{(I,q)}$	$\dot{\hat{x}}^{(I,q)} = A^{(I)}\hat{x}^{(I,q)} + \gamma^{(I)}(\hat{x}^{(I,q)}, u^{(I)})$ $+ h^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, w_\star^{(I)})$ $+ L^{(I,q)}(y^{(I,q)} - C^{(I,q)}\hat{x}^{(I,q)})$	Distributed	$w_\star^{(I)} = \begin{bmatrix} \zeta_r^{(I)} \\ y_z^{(I)} \end{bmatrix}$
	$\hat{x}^{(I,q)}(0) = 0$	Decentralized	$w_\star^{(I)} = z_r^{(I)}$
Residual	$\varepsilon_{y_j}^{(I,q)} = y_j^{(I)} - C_j^{(I)}\hat{x}_j^{(I,q)}, j \in \mathcal{J}^{(I,q)} = \{j \in \{1, \dots, m\} : \mathcal{S}^{(I)}\{j\} \in \mathcal{S}^{(I,q)}\}$		

Adaptive Threshold	$\begin{aligned}\bar{\varepsilon}_{y_j}^{(I,q)} &= Y_j^{(I,q)}(t) + \frac{\alpha_j^{(I,q)}}{s + \zeta_j^{(I,q)}} [\Lambda_I Z^{(I,q)}] \\ &\quad + \frac{\alpha_j^{(I,q)}}{s + \zeta_j^{(I,q)}} \left[\bar{\eta}^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, w_\star^{(I)}, t) \right] \\ Y_j^{(I,q)} &= \alpha_j^{(I,q)} \bar{x}^{(I,q)} e^{-\zeta_j^{(I,q)} t} + \bar{d}_j^{(I)} \\ &\quad + \frac{\alpha_{d_j}^{(I,q)} \bar{d}^{(I,q)}}{\zeta_{d_j}^{(I,q)}} (1 - e^{-\zeta_{d_j}^{(I,q)} t}) \\ &\quad + \frac{\alpha_j^{(I,q)} (\lambda_{h_I} + \lambda_{\eta_I}) \bar{w}_\star^{(I)}}{\zeta_j^{(I,q)}} (1 - e^{-\zeta_j^{(I,q)} t})\end{aligned}$	Distributed	$\begin{aligned}w_\star^{(I)} &= \begin{bmatrix} \zeta_r^{(I)} \\ y_z^{(I)} \end{bmatrix} \\ \bar{w}_\star^{(I)} &= \bar{d}_z^{(I)} + \bar{\zeta}^{(I)}\end{aligned}$
	$\begin{aligned}Z^{(I,q)} &= \left(1 + \frac{\rho^{(I,q)} \Lambda_I}{s + \xi^{(I,q)} - \rho^{(I,q)} \Lambda_I} \right) E^{(I,q)} \\ E^{(I,q)} &= \rho^{(I,q)} \bar{x}^{(I,q)} e^{-\xi^{(I,q)} t} \\ &\quad + \frac{\rho_{d-}^{(I,q)} \bar{d}^{(I,q)}}{\xi_{d-}^{(I,q)}} (1 - e^{-\xi_{d-}^{(I,q)} t}) \\ &\quad + \frac{\rho^{(I,q)}}{s + \xi^{(I,q)}} \bar{\eta}^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, w_\star^{(I)}, t) \\ &\quad + \frac{\rho^{(I,q)} (\lambda_{h_I} + \lambda_{\eta_I}) \bar{w}_\star^{(I)}}{\xi^{(I,q)}} (1 - e^{-\xi^{(I,q)} t})\end{aligned}$	Decentralized	$\begin{aligned}w_\star^{(I)} &= z_r^{(I)} \\ \bar{w}_\star^{(I)} &= \bar{z}_r^{(I)}\end{aligned}$
Decision	$\begin{aligned}D^{(I,q)}(t) &= \begin{cases} 0, & \text{if } D_j^{(I,q)}(t) = 0, \forall j \in \mathcal{J}^{(I,q)} \\ 1, & \text{otherwise} \end{cases} \\ D_j^{(I,q)}(t) &= \begin{cases} 0, & \text{if } t < \min \left\{ t : \left \varepsilon_{y_j}^{(I,q)}(t) \right - \varepsilon_{y_j}^{(I,q)}(t) > 0 \right\} \\ 1, & \text{otherwise} \end{cases}\end{aligned}$		

5

Sensor Fault Isolation

The process of sensor fault isolation is executed by every monitoring agent $\mathcal{M}^{(I)}$, though the the aggregation module $\mathcal{A}^{(I)}$ (in the centralized case the aggregation module is denoted by \mathcal{A}) as shown in Fig. 3.1. After the first time that a monitoring module $\mathcal{M}^{(I,q)}$ detects sensor faults, the aggregation module $\mathcal{A}^{(I)}$ is activated for isolating the combination of sensor faults that has affected the sensor set $\mathcal{S}^{(I)}$. In the sequel, the multiple sensor fault isolation decision logic is initially presented for a decentralized architecture (local sensor fault isolation), and then for a distributed architecture where a global agent is used to collect the decisions of the monitoring agents in order to isolate the combination of sensor faults propagated in the network of cyber-physical systems (CPS). The multiple sensor fault isolation decision logic for a centralized architecture is omitted, since it is similar to the multiple sensor fault isolation decision logic applied in a decentralized architecture.

In all cases, the decisions obtained by the N_I modules of the monitoring agent $\mathcal{M}^{(I)}$, $I \in \{1, \dots, M\}$, constitute the observed pattern of sensor faults denoted by $D^{(I)}(t)$; i.e.,

$$D^{(I)}(t) = \left[D^{(I,1)}(t), \dots, D^{(I,N_I)}(t) \right]^\top \quad (5.1)$$

where $D^{(I,q)}$, $q \in \{1, \dots, N_I\}$ is defined through (4.94)-(4.96). As long as $D^{(I)}(t) = \mathbf{0}_{N_I}$ ($\mathbf{0}_{N_I}$ is a zero vector of length N_I), the module $\mathcal{A}^{(I)}$ is not activated and the agent $\mathcal{M}^{(I)}$ infers that there is no sensor fault in $\mathcal{S}^{(I)}$.

5.1 Decentralized Multiple Sensor Fault Isolation

In a decentralized architecture, the output of $\mathcal{M}^{(I)}$ is to infer the possibly faulty sensors in $\mathcal{S}^{(I)}$ without exchanging any information with neighboring agents. The diagnostic reasoning of $\mathcal{M}^{(I)}$ is based on the diagnosis set that contains all the possible combinations of sensor faults that may have occurred in the sensor set $\mathcal{S}^{(I)}$.

5.1.1 Sensor Fault Diagnosis Set

The observed pattern $D^{(I)}(t)$ defined in (5.1) is compared to a number of theoretical patterns, which are the columns of the multiple sensor fault signature matrix, denoted by $F^{(I)}$, $I \in \{1, \dots, M\}$. The matrix $F^{(I)}$ is binary, consisting of N_I rows and N_{cI} columns, where $N_{cI} = 2^{m_I} - 1$ (m_I is the number of sensors in the sensor set $\mathcal{S}^{(I)}$). The q -th row of the matrix $F^{(I)}$ corresponds to the q -th set of ARRs $\mathcal{E}^{(I,q)}$, $q \in \{1, \dots, N_I\}$, while every column corresponds to a different combination of sensor faults in $\mathcal{S}^{(I)}$. The theoretical pattern of sensor fault combination $F_i^{(I)}$ is defined as:

$$F_i^{(I)} = \left[F_{1i}^{(I)}, \dots, F_{N_I i}^{(I)} \right]^\top, \quad (5.2)$$

where $F_{qi}^{(I)} = 1$, if at least one sensor fault that belongs to the combination $\mathcal{F}_{c_i}^{(I)}$, $i \in \{1, \dots, N_{cI}\}$, can provoke the violation of (or else is involved in) $\mathcal{E}^{(I,q)}$, $q \in \{1, \dots, N_I\}$ and $F_{qi}^{(I)} = 0$ otherwise. The design of $F^{(I)}$ exploits the structure sensor fault sensitivity property of $\mathcal{E}^{(I,q)}$, described in Lemma 4.5. The observed sensor fault pattern $D^{(I)}(t)$ is consistent with the i -th theoretical sensor fault pattern $F_i^{(I)}$, if $D^{(I,q)}(t) = F_{qi}^{(I)}$ for all $q \in \{1, \dots, N_I\}$ (consistency test).

When the consistency test is satisfied, we determine the sensor fault diagnosis set $\mathcal{D}_s^{(I)}(t)$ that contains the combination(s) of sensor faults,

whose associated theoretical fault pattern(s) is (are) consistent to the observed pattern [Koscielny et al. [2012]]; i.e.

$$\mathcal{D}_s^{(I)}(t) = \left\{ \mathcal{F}_{c_i}^{(I)} : i \in \mathcal{I}_D^{(I)}(t) \right\}, \quad (5.3)$$

where

$$\mathcal{I}_D^{(I)}(t) = \left\{ i : F_i^{(I)} = D^{(I)}(t), i \in \{1, \dots, N_{CI}\} \right\} \quad (5.4)$$

The diagnosis set $\mathcal{D}_s^{(I)}(t)$ may contain a single or more than one combination of sensor faults. The pattern $D^{(I)}(t)$ may change over time, while it is possible that, at some time instant, the consistency test is not satisfied. This may happen if there are two or more identical theoretical patterns of sensor faults, and some possible observed patterns do not match the theoretical patterns. In this case, the sensor fault diagnosis set $\mathcal{D}_s^{(I)}(t)$ contains the sensor fault combinations involved in the violated ARRs [Cordier et al. [2004]]; i.e.

$$\mathcal{D}_s^{(I)}(t) = \bigcap_{q \in \mathcal{Q}(t)} \text{Supp}(\mathcal{E}^{(I,q)}), \quad (5.5)$$

where $\text{Supp}(\mathcal{E}^{(I,q)})$ is the support of $\mathcal{E}^{(I,q)}$, which is the set of sensor fault combinations $\mathcal{F}_{c_i}^{(I)}$ for which $F_{qi}^{(I)} = 1$ and

$$\mathcal{Q}^{(I)}(t) = \left\{ q : D^{(I,q)}(t) = 1, q \in \{1, \dots, N_I\} \right\}. \quad (5.6)$$

In general, $\left\{ \mathcal{F}_{c_i}^{(I)} : i \in \mathcal{I}_D^{(I)}(t) \right\} \subseteq \bigcap_{q \in \mathcal{Q}(t)} \text{Supp}(\mathcal{E}^{(I,q)})$, i.e. the consistency test provides a diagnosis set with smaller (or equal) cardinality compared to the diagnosis set defined in (5.5). When the diagnosis set $\mathcal{D}_s^{(I)}(t)$ is defined based on (5.3), we may say that the observed behavior of $\mathcal{S}^{(I)}$ is consistent with its expected faulty behavior, represented by the theoretical patterns.

In network of CPS, the number of all possible multiple sensor fault combinations may be too large, leading to a large number of theoretical sensor fault patterns to which the observed patterns should be compared, increasing the computational complexity. For this reason, instead of designing a single multiple sensor fault signature matrix $F^{(I)}$, we may create $\ell^* > 1$ matrices of smaller dimension in the agent

$\mathcal{M}^{(I)}$. The design of these matrices is realized taking into account that there may be more than one cluster of overlapping local sensor sets, while there is a cluster of disjoint local sensor sets. For example, let $\mathcal{S}^{(I)}$ be decomposed in $\mathcal{S}^{(I,1)}, \dots, \mathcal{S}^{(I,8)}$ with $\mathcal{S}^{(I,1)} \cap \mathcal{S}^{(I,2)} \neq \emptyset$, $\mathcal{S}^{(I,2)} \cap \mathcal{S}^{(I,3)} \neq \emptyset$, and $\mathcal{S}^{(I,4)} \cap \mathcal{S}^{(I,5)} \neq \emptyset$, while there are three disjoint local sensor sets $\mathcal{S}^{(I,6)}, \mathcal{S}^{(I,7)}, \mathcal{S}^{(I,8)}$. In this case, we may consider three clusters of local sensor sets; the first cluster consists of $\mathcal{S}^{(I,1)}, \mathcal{S}^{(I,2)}, \mathcal{S}^{(I,3)}$; the second cluster includes $\mathcal{S}^{(I,4)}, \mathcal{S}^{(I,5)}$; the third cluster is comprised of $\mathcal{S}^{(I,6)}, \mathcal{S}^{(I,7)}, \mathcal{S}^{(I,8)}$. For each of the three clusters, we may design a sensor fault signature matrix $F^{(I,\ell)}$, $\ell = 1, 2, 3$, of lower dimension compared to a single sensor fault signature matrix $F^{(I)}$.

5.1.2 Diagnostic Reasoning

The multiple sensor fault isolation decision is obtained applying diagnostic reasoning to the resultant sensor fault diagnosis set. Particularly, based on $\mathcal{D}_s^{(I)}(t)$ and taking into account permanent faults (see Section 2.2), we may infer that

- at least one of the sensor fault combinations $\mathcal{F}_{c_i}^{(I)}$ that belongs to $\mathcal{D}_s^{(I)}(t)$ has occurred,
- the sensor fault $f_j^{(I)}$ such that $f_j^{(I)} \in \bigcap_{i \in \mathcal{I}_D^{(I)}(t)} \mathcal{F}_{c_i}^{(I)}$ with $\bigcap_{i \in \mathcal{I}_D^{(I)}(t)} \mathcal{F}_{c_i}^{(I)} \neq \emptyset$, is guaranteed to have occurred,
- the occurrence of the sensor fault combination $\mathcal{F}_{c_i}^{(I)}$ such that $\mathcal{F}_{c_i}^{(I)} \not\subset \bigcap_{q \in \mathcal{Q}(t)} \text{Supp}(\mathcal{E}^{(I,q)})$, is excluded.

When $f_j^{(I)}$ is guaranteed to have occurred, the associated sensor $\mathcal{S}^{(I)}\{j\}$ is isolated as faulty. If $f_j^{(I)}$ belongs to some of the diagnosed sensor fault combinations, but not to all of them, the sensor $\mathcal{S}^{(I)}\{j\}$ is characterized as possibly faulty.

Based on this diagnostic reasoning and assuming permanent faults, we do not only conclude to some sensor fault combinations but also we

can exclude the occurrence of some other fault combinations, enhancing the fault isolation procedure. Thus, the sensor fault diagnosis set may be updated overtime. After each update of $\mathcal{D}_s^{(I)}(t)$, we may reduce the number of columns of the matrix $F^{(I)}$, by omitting the columns that correspond to the sensor fault combinations, which have been excluded. Moreover, at every time instant, the sensor fault diagnosis set $\mathcal{D}_s^{(I)}(t)$ separates the set $\mathcal{E}^{(I)} = \{\mathcal{E}^{(I,1)}, \dots, \mathcal{E}^{(I,N_I)}\}$ into two subsets:

- the first subset, denoted by $\mathcal{E}_{\mathcal{D}_s^{(I)}}^{(I)}(t)$, contains the violated sets of ARRs, i.e. $\mathcal{E}_{\mathcal{D}_s^{(I)}}^{(I)}(t) = \{\mathcal{E}^{(I,q)} : q \in \mathcal{Q}(t)\}$
- the second set contains the sets of ARRs that have not been violated, denoted by $\neg\mathcal{E}_{\mathcal{D}_s^{(I)}}^{(I)}(t)$ (\neg is the negation symbol), defined as $\neg\mathcal{E}_{\mathcal{D}_s^{(I)}}^{(I)}(t) = \mathcal{E}^{(I)} \setminus \mathcal{E}_{\mathcal{D}_s^{(I)}}^{(I)}(t)$

When the diagnosis set $\mathcal{D}_s^{(I)}(t)$ arises from the consistency test and there is a set of ARRs $\mathcal{E}^{(I,q^*)}$ that belongs to $\neg\mathcal{E}_{\mathcal{D}_s^{(I)}}^{(I)}(t)$, it is implied that $D^{(I,q^*)}(t) = F_{q^*i}^{(I)} = 0$ for all $i \in \mathcal{I}_D^{(I)}(t)$. The semantics of $F_{q^*i}^{(I)} = 0$ is that none of the sensor faults, which belong to the diagnosed sensor fault combinations $\mathcal{F}_{c_i}^{(I)}$, can provoke the violation of $\mathcal{E}^{(I,q^*)}$ and the associated local sensor set $\mathcal{S}^{(I,q^*)}$ does not include any possibly faulty sensor. Moreover, since $D^{(I,q^*)}(t) = 0$, the module $\mathcal{M}^{(I,q^*)}$ infers the non-faulty condition of the local sensor set $\mathcal{S}^{(I,q^*)}$ (within an exoneration framework (see Section 4.3)).

Based on the previous analysis, in the case that $\mathcal{D}_s^{(I)}(t)$ includes two or more sensor fault combinations, the module $\mathcal{M}^{(I,q^*)}$ can be used to isolate a single sensor fault combination among the combinations included in $\mathcal{D}_s^{(I)}(t)$, as follows. Let $\mathcal{S}^{(I)}\{j\}$ be diagnosed as possibly faulty sensor that belongs to the sensor group $\mathcal{S}^{(I,k)}$ monitored by the module $\mathcal{M}^{(I,k)}$. The measurements of $\mathcal{S}^{(I)}\{j\}$ may be also used by the module $\mathcal{M}^{(I,q^*)}$ that monitors the sensor group $\mathcal{S}^{(I,q^*)}$ ($\mathcal{S}^{(I,q^*)} \cap \mathcal{S}^{(I,k)} = \emptyset$). Given that $\mathcal{S}^{(I,q^*)}$ is healthy, the module $\mathcal{M}^{(I,q^*)}$ provides a healthy estimation model $\hat{x}_H^{(I,q)}(t)$ defined in (4.4). Hence, the module $\mathcal{M}^{(I,q^*)}$ may obtain a decision for the presence of fault in

$\mathcal{S}^{(I)}\{j\}$ for $j \in \{1, \dots, m_I\}$, $j \notin \mathcal{J}^{(I,q^*)}$, based on the following ARR:

$$\mathcal{E}_j^{(I,q^*)} : \left| y_j^{(I)}(t) - C_j^{(I)} \hat{x}_H^{(I,q^*)}(t) \right| - \bar{\varepsilon}_{y_{Hj}}^{(I,q^*)}(t) \leq 0, \quad (5.7)$$

where $y_j^{(I)}$ is defined in (2.3), $C_j^{(I)} \mathbb{R}^{1 \times n_I}$ and $\bar{\varepsilon}_{y_{Hj}}^{(I,q^*)}$ is defined through (4.84)-(4.87) for $j \in \{1, \dots, m_I\}$, $j \notin \mathcal{J}^{(I,q^*)}$, taking into account $\hat{x}^{(I,q^*)} = \hat{x}_H^{(I,q^*)}$. The ARR defined in (5.7) is sensitive only to sensor fault $f_j^{(I)}$, $j \in \{1, \dots, m_I\}$, $j \notin \mathcal{J}^{(I,q)}$.

It is noted that following a centralized approach for diagnosing sensor faults in a network of CPS, a small group of neighboring monitoring modules can be used for isolating faulty sensors in a local sensor set that is isolated as faulty, as previously described. Therefore, the proposed sensor FDI scheme is expected to work even in the presence of limited communication link. The minimum communication requirement that guarantees that the centralized approach works "properly" is the minimum set of neighboring monitoring modules that can communicate in order to achieve isolation, i.e. a small group of neighboring monitoring modules that can be used for isolating the faulty sensors within a local sensor set that is localized as faulty.

The following simple example is used to illustrate the proposed multiple sensor fault isolation decision logic.

Example 5.1. Consider the sensor set $\mathcal{S}^{(I)}$ comprised of three sensors ($m_I = 3$) (see Fig. 5.1). Let $\mathcal{S}^{(I)}$ be decomposed into three local sensor sets ($N_I = 3$) such that $\mathcal{S}^{(I,1)} = \{\mathcal{S}^{(I)}\{1\}\}$ (black box), $\mathcal{S}^{(I,2)} = \{\mathcal{S}^{(I)}\{2\}, \mathcal{S}^{(I)}\{3\}\}$ (blue box) and $\mathcal{S}^{(I,3)} = \{\mathcal{S}^{(I)}\{3\}\}$ (green box). Each of the three groups is monitored by a module $\mathcal{M}^{(I,q)}$, $q = 1, 2, 3$.

Given that all possible sensor fault combinations are $\mathcal{F}_{c_1}^{(I)} = \{f_1^{(I)}\}$, $\mathcal{F}_{c_2}^{(I)} = \{f_2^{(I)}\}$, $\mathcal{F}_{c_3}^{(I)} = \{f_3^{(I)}\}$, $\mathcal{F}_{c_4}^{(I)} = \{f_1^{(I)}, f_2^{(I)}\}$, $\mathcal{F}_{c_5}^{(I)} = \{f_1^{(I)}, f_3^{(I)}\}$, $\mathcal{F}_{c_6}^{(I)} = \{f_2^{(I)}, f_3^{(I)}\}$ and $\mathcal{F}_{c_7}^{(I)} = \{f_1^{(I)}, f_2^{(I)}, f_3^{(I)}\}$, the multiple sensor fault signature matrix is formulated as shown in Table 5.1. For example, if the observed sensor fault pattern $D^{(I)}(t)$ equals to $[1, 1, 0]^\top$, then the sensor fault diagnosis set is $\mathcal{D}_s^{(I)}(t) = \{\mathcal{F}_{c_4}^{(I)}\}$. Since the diagnosis set includes a single combination of sensor faults, we may infer that

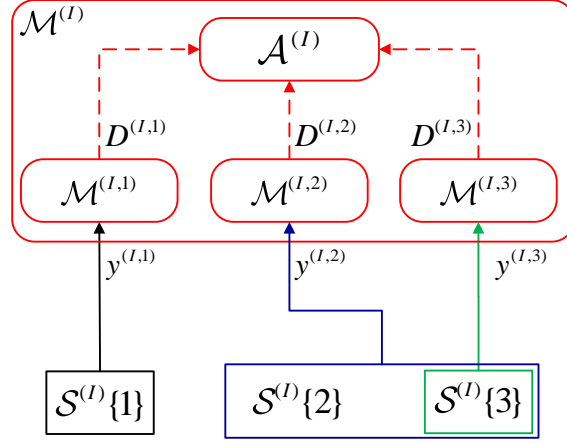


Figure 5.1: Example 5.1: Decomposition of the sensor set $\mathcal{S}^{(I)} = \{\mathcal{S}^{(I)}\{1\}, \mathcal{S}^{(I)}\{2\}, \mathcal{S}^{(I)}\{3\}\}$ into three local sensor sets: $\mathcal{S}^{(I,1)} = \{\mathcal{S}^{(I)}\{1\}\}$ (black box), $\mathcal{S}^{(I,2)} = \{\mathcal{S}^{(I)}\{2\}, \mathcal{S}^{(I)}\{3\}\}$ (blue box) and $\mathcal{S}^{(I,3)} = \{\mathcal{S}^{(I)}\{3\}\}$ (green box).

	$\mathcal{F}_{c_1}^{(I)}$	$\mathcal{F}_{c_2}^{(I)}$	$\mathcal{F}_{c_3}^{(I)}$	$\mathcal{F}_{c_4}^{(I)}$	$\mathcal{F}_{c_5}^{(I)}$	$\mathcal{F}_{c_6}^{(I)}$	$\mathcal{F}_{c_7}^{(I)}$
$\mathcal{E}^{(I,1)}$	1	0	0	1	1	0	1
$\mathcal{E}^{(I,2)}$	0	1	1	1	1	1	1
$\mathcal{E}^{(I,3)}$	0	0	1	0	1	1	1

Table 5.1: Multiple sensor fault signature matrix $F^{(I)}$ of example 5.1.

both $f_1^{(I)}$ and $f_2^{(I)}$ have occurred and both sensors $\mathcal{S}^{(I)}\{1\}$ and $\mathcal{S}^{(I)}\{2\}$ are faulty, while the occurrence of the sensor fault combinations $\{f_1^{(I)}\}$, $\{f_2^{(I)}\}$, $\{f_3^{(I)}\}$ and $\{f_2^{(I)}, f_3^{(I)}\}$ are excluded. On the other hand, if $D^{(I)}(t) = [0, 1, 1]^\top$, the diagnosis set $\mathcal{D}_s^{(I)}(t)$ contains more than one sensor fault combination, i.e. $\mathcal{D}_s^{(I)}(t) = \{\mathcal{F}_{c_3}^{(I)}, \mathcal{F}_{c_6}^{(I)}\}$. Based on the diagnostic reasoning decision logic, we may infer that: (i) the sensor fault $f_3^{(I)}$ has occurred or both $f_3^{(I)}$ and $f_2^{(I)}$ have occurred, (ii) the sensor fault $f_3^{(I)}$ is guaranteed to have occurred, and (iii) the occurrence of the sensor fault combinations $\{f_1^{(I)}\}$, $\{f_2^{(I)}\}$ and $\{f_1^{(I)}, f_2^{(I)}\}$ is excluded. Therefore, sensor $\mathcal{S}^{(I)}\{3\}$ is diagnosed as faulty, while $\mathcal{S}^{(I)}\{2\}$ as possibly faulty. Since the set $\mathcal{E}^{(I,1)}$ is not violated, we can use the module $\mathcal{M}^{(I,1)}$, as described in Section 5.1.2, in order to infer the status of sensor fault $f_2^{(I)}$. For this reason, the measurements of $\mathcal{S}^{(I)}\{2\}$ can be used by the module $\mathcal{M}^{(I,1)}$ in order to obtain a decision on the occurrence of a fault in $\mathcal{S}^{(I)}\{2\}$ based on the following set of ARRs:

$$\mathcal{E}_2^{(I,1)} : \quad |y_2^{(I)}(t) - C_2^{(I)} \hat{x}_H^{(I,1)}(t)| - \bar{\varepsilon}_{y_{2H}}^{(I,1)}(t) \leq 0. \quad (5.8)$$

If at some time instant, $D^{(I)}(t) = [1, 0, 1]^\top$, the consistency test cannot provide any result. In this case, the sensor fault diagnosis set is determined through (5.5), i.e. $\mathcal{D}_s^{(I)} = \text{Supp}(\mathcal{E}^{(I,1)}) \cap \text{Supp}(\mathcal{E}^{(I,3)}) = \{\mathcal{F}_{c_1}^{(I)}, \mathcal{F}_{c_4}^{(I)}, \mathcal{F}_{c_5}^{(I)}, \mathcal{F}_{c_7}^{(I)}\} \cap \{\mathcal{F}_{c_3}^{(I)}, \mathcal{F}_{c_5}^{(I)}, \mathcal{F}_{c_6}^{(I)}, \mathcal{F}_{c_7}^{(I)}\} = \{\mathcal{F}_{c_5}^{(I)}, \mathcal{F}_{c_7}^{(I)}\}$. Hence, we may infer that: (i) the sensor faults $f_1^{(I)}$ and $f_3^{(I)}$ have occurred or all sensor faults have occurred, (ii) $f_1^{(I)}$ and $f_3^{(I)}$ are guaranteed to have occurred, and (iii) the occurrence of the sensor fault combinations $\{f_1^{(I)}\}$, $\{f_2^{(I)}\}$, $\{f_3^{(I)}\}$, $\{f_1^{(I)}, f_2^{(I)}\}$ and $\{f_2^{(I)}, f_3^{(I)}\}$ is excluded.

5.2 Distributed Multiple Sensor Fault Isolation

In a distributed architecture, the multiple sensor fault isolation is realized in two levels; locally, the monitoring agent $\mathcal{M}^{(I)}$ isolates sensor faults based on the decisions of the monitoring modules $\mathcal{M}^{(I,q)}$, and globally, the decisions of the monitoring agents are combined for iso-

lating multiple propagated sensor faults.

5.2.1 Local Sensor Fault Isolation Decision Logic

The monitoring agent $\mathcal{M}^{(I)}$ uses a binary fault signature matrix $F^{(I)}$, which is similar to the fault signature matrix used in a decentralized architecture (see Section 5.1.1), but in a distributed architecture, $F^{(I)}$ consists of N_I rows and $N_{CI} + 2$ columns, where $N_{CI} = 2^{m_I} - 1$ (m_I is the number of sensors in the sensor set $\mathcal{S}^{(I)}$); the q -th row corresponds to the q -th set of ARRs $\mathcal{E}^{(I,q)}$, $q \in \{1, \dots, N_I\}$; the i -th column, for all $i \in \{1, \dots, N_{CI}\}$ corresponds to the i -th combination of sensor faults that may affect the sensor set $\mathcal{S}^{(I)}$, denoted by $\mathcal{F}_{c_i}^{(I)}$, while the $N_{CI} + 1$ column corresponds to the fault vector $f_z^{(I)}$, that is sensor faults from neighboring CPS, and the $N_{CI} + 2$ column corresponds to the union of all combinations of $f_z^{(I)}$ and $\mathcal{F}_{c_i}^{(I)}$ for all $i \in \{1, \dots, N_{CI}\}$. For example, if $m_I = 2$, $F^{(I)}$ has five columns, corresponding to the following sensor fault combination: $\mathcal{F}_{c_1}^{(I)} = \{f_1^{(I)}\}$, $\mathcal{F}_{c_2}^{(I)} = \{f_2^{(I)}\}$, $\mathcal{F}_{c_3}^{(I)} = \{f_1^{(I)}, f_2^{(I)}\}$, $\mathcal{F}_{c_4}^{(I)} = \{f_z^{(I)}\}$ and $\mathcal{F}_{c_5}^{(I)} = \bigcup_{i \in \{1,2,3\}} \{f_z^{(I)}, \mathcal{F}_{c_i}^{(I)}\}$. It is noted that the theoretical pattern $F_{N_{CI}+1}^{(I)}$ is a column vector with 1 in each element, since the sensor fault vector $f_z^{(I)}$ is involved in each $\mathcal{E}^{(I,q)}$. Similarly, the theoretical pattern $F_{N_{CI}+2}^{(I)}$ is a vector with 1 in each element, representing the qualitative sensitivity of each $\mathcal{E}^{(I,q)}$ to the combination of $f_z^{(I)}$ with every sensor fault combination $\mathcal{F}_{c_i}^{(I)}$, $i \in \{1, \dots, N_{CI}\}$. The design of $F^{(I)}$ exploits the structure sensor fault sensitivity property of $\mathcal{E}^{(I,q)}$, described in Lemma 4.4.

The decisions obtained by the N_I modules of the agent $\mathcal{M}^{(I)}$ constitute the observed pattern of sensor faults, $D^{(I)}(t)$, affecting $\mathcal{S}^{(I)}$ and $\mathcal{S}_z^{(I)}$, which is defined in (5.1). Following the same procedure presented in Section 5.1.1 for determining the sensor fault diagnosis, the observed pattern, $D^{(I)}(t)$ is compared to each of the $N_{CI} + 2$ theoretical patterns $F_i^{(I)}$, $i \in \{1, \dots, M\}$, defined in (5.2) in order to determine the sensor fault diagnosis set $\mathcal{D}_s^{(I)}(t)$ that includes the sensor fault combinations that have possibly occurred and is computed either through (5.3 or (5.5) as described in Section 5.1.1.

The outputs of the agent $\mathcal{M}^{(I)}$, $I \in \{1, \dots, M\}$ are the diagnosis set $\mathcal{D}_s^{(I)}$ and the decision on the presence of sensor faults in $\mathcal{S}_z^{(I)}$, represented by the function $D_z^{(I)}(t)$:

$$D_z^{(I)}(t) = \begin{cases} 0, & \text{if } f_z^{(I)} \notin \mathcal{D}_s^{(I)}(t) \text{ and } f_p^{(I)} \notin \mathcal{D}_s^{(I)}(t) \\ 1, & \text{if } f_z^{(I)} \in \mathcal{D}_s^{(I)}(t) \text{ or } f_p^{(I)} \in \mathcal{D}_s^{(I)}(t) \end{cases} \quad (5.9)$$

where $f_p^{(I)} \in \mathbb{R}^{m_I^*}$, $m_I^* \leq m_I$ collectively amounts for the sensor faults that are propagated from the agent $\mathcal{M}^{(I)}$ to its neighboring agents due to the exchange of sensor information. The diagnostic reasoning behind the decision of the agent $\mathcal{M}^{(I)}$ made for multiple sensor fault isolation relies on the resultant diagnosis set and is similar to the diagnostic reasoning described in Section 5.1.2. In a distributed architecture if $f_z^{(I)}$ belongs to $\mathcal{D}_s^{(I)}(t)$, then at least one sensor fault in $\mathcal{S}_z^{(I)}$ might have been propagated to $\mathcal{M}^{(I)}$ from a neighboring agent. On the other hand, if $f_p^{(I)}$ belongs to $\mathcal{D}_s^{(I)}(t)$, the agent $\mathcal{M}^{(I)}$ has propagated sensor faults to neighboring agents. The faults $f_p^{(I)}$ may provoke the violation of ARRs of neighboring agents. For this reason, the global decision logic is applied in order to isolate propagated sensor faults.

5.2.2 Global Sensor Fault Isolation Decision Logic

The primary goal of the global decision logic \mathcal{G} is to isolate sensor faults that have been propagated from neighboring agents. If $D_z^{(I)}(t) = 0$ for all $I \in \{1, \dots, M\}$, then no information (decisions) is processed by the global decision logic; otherwise the global decision logic is based on the theoretical patterns of propagated sensor faults, which describe the involvement of the sensor faults $f_z^{(I)}$ in the set of ARRs, $\mathcal{E}^{(I)}$, for all $I \in \{1, \dots, M\}$, defined as

$$\mathcal{E}^{(I)} = \bigcup_{q \in \{1, \dots, N_I\}} \mathcal{E}^{(I, q)}. \quad (5.10)$$

The theoretical patterns are the columns of the sensor fault signature matrix F^z , which has M rows and $N_c = 2^p - 1$ columns, where $p \leq \sum_{I=1}^M p_I$ (p_I is the length of $f_z^{(I)}$). The I -th row corresponds to the set of ARRs $\mathcal{E}^{(I)}$ and the k -th column, $k \in \{1, \dots, N_c\}$ to the k -th

combination of sensor faults, denoted by $\mathcal{F}_{c_k}^z$ that affect $\bigcup_{I=1}^N \mathcal{S}_z^{(I)}$. The k -th theoretical pattern of transmitted faulty sensor information is the k -th column of F^z defined as:

$$F_k^z = [F_{1k}^z, \dots, F_{Nk}^z]^\top \quad (5.11)$$

where: a) $F_{Ik}^z = 1$, $I \in \{1, \dots, M\}$, $k \in \{1, \dots, Nc\}$, if at least one sensor fault included in $\mathcal{F}_{c_k}^z$ is involved in $\mathcal{E}^{(I)}$ and affects $\mathcal{S}^{(I)}$, b) $F_{Ik}^z = 0$ if none of the sensor faults included in $\mathcal{F}_{c_k}^z$ is involved in $\mathcal{E}^{(I)}$, and c) $F_{Ik}^z = *$, if at least one sensor fault included in $\mathcal{F}_{c_k}^z$ is involved in $\mathcal{E}^{(I)}$ and affects $\mathcal{S}_z^{(I)}$, while none of the sensor faults included in $\mathcal{F}_{c_k}^z$ affects $\mathcal{S}^{(I)}$.

The meaning of $F_{Ik}^z = 1$ is that a sensor fault of $\mathcal{F}_{c_k}^z$ that affects $\mathcal{S}^{(I)}$ necessarily provokes the violation of $\mathcal{E}^{(I)}$. In other words, $\mathcal{E}^{(I)}$ is very sensitive to sensor faults $f^{(I)}$. The meaning of $F_{Ik}^z = *$ is that a sensor fault of $\mathcal{F}_{c_k}^z$ that belongs to $f_z^{(I)}$ and is involved in $\mathcal{E}^{(I)}$, i.e. a sensor fault that is propagated from a neighboring agent, can explain why $\mathcal{E}^{(I)}$ is violated at some time instant, but $\mathcal{E}^{(I)}$ may happen to be satisfied while this sensor fault in $\mathcal{F}_{c_k}^z$ has occurred. In this way we discriminate local from propagated sensor faults. This may happen if $\mathcal{E}^{(I)}$ is not very sensitive to $f_z^{(I)}$ (see Section 6).

The combinatorial process of the decisions of the N monitoring agents is performed by initially determining the observed pattern of propagated sensor faults as:

$$D_z(t) = [D_z^{(1)}(t), \dots, D_z^{(M)}(t)]^\top. \quad (5.12)$$

The observed pattern $D_z(t)$ is compared to each of the columns of F_k^z . If $D_z(t)$ is consistent to F_k^z , i.e. $D_z(t) = F_k^{(z)}$ the diagnosis set of propagated sensor faults is defined as:

$$\mathcal{D}_s^z(t) = \left\{ \mathcal{F}_{c_k}^z : k \in \mathcal{I}_z(t) \right\}, \quad (5.13)$$

where $\mathcal{I}_z(t)$ is an index set defined as $\mathcal{I}_z(t) = \left\{ k : F_{Ik}^z = D_z^{(I)}(t), k \in \{1, \dots, Nc\} \right\}$ for all $I \in \{1, \dots, M\}$. If $F_{Ik}^z = *$ then $D_z^{(I)}(t) = F_{Ik}^z$ if either $D_z^{(I)}(t) = 0$ or $D_z^{(I)}(t) = 1$. If

$D_z(t)$ is not consistent to F_k^z , then $\mathcal{D}_s^z(t)$ contains the sensor fault combinations that belong to the support of the violated $\mathcal{E}^{(I)}$, i.e. the sensor fault combinations $\mathcal{F}_{c_k}^z$ for which $F_{I_k}^z = 1$ or $F_{I_k}^z = *$. The set $\mathcal{D}_s^z(t)$ is used to update the non-empty diagnosis set $\mathcal{D}_s^{(I)}(t)$ of $\mathcal{M}^{(I)}$ by excluding $f_z^{(I)}$ and its combinations, if $f_z^{(I)} \notin \mathcal{D}_s^z(t)$.

The following simple example is used for illustrating the proposed multiple sensor fault isolation decision logic.

Example 5.2. Consider three monitoring agents $\mathcal{M}^{(1)}$, $\mathcal{M}^{(2)}$ and $\mathcal{M}^{(3)}$ monitoring the sensor sets $\mathcal{S}^{(1)}$, $\mathcal{S}^{(2)}$ and $\mathcal{S}^{(3)}$, respectively, while (i) $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ use the measurements of $\mathcal{S}_z^{(1)} = \mathcal{S}_z^{(2)} = \{\mathcal{S}^{(3)}\{4\}\}$ transmitted from $\mathcal{M}^{(3)}$, and (ii) $\mathcal{M}^{(3)}$ uses the measurements of $\mathcal{S}_z^{(3)} = \{\mathcal{S}^{(1)}\{5\}, \mathcal{S}^{(2)}\{1\}\}$ transmitted from $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, as shown in Fig. 5.2.

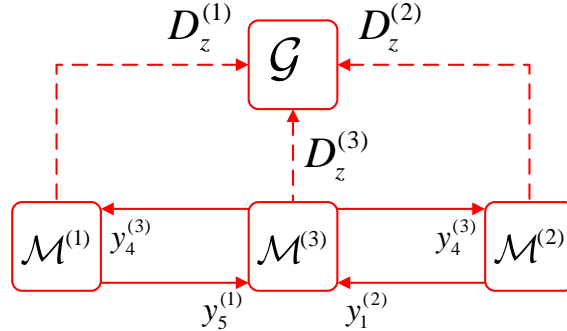


Figure 5.2: Example 5.2: Exchange of sensor information between agents $\mathcal{M}^{(1)}$, $\mathcal{M}^{(2)}$ and $\mathcal{M}^{(3)}$.

The sensor fault signature matrix F^z is designed as shown in Table 5.2, where $\mathcal{F}_{c_1}^z = \{f_5^{(1)}\}$, $\mathcal{F}_{c_2}^z = \{f_1^{(2)}\}$, $\mathcal{F}_{c_3}^z = \{f_4^{(3)}\}$, $\mathcal{F}_{c_4}^z = \{f_5^{(1)}, f_1^{(2)}\}$, $\mathcal{F}_{c_5}^z = \{f_5^{(1)}, f_4^{(3)}\}$, $\mathcal{F}_{c_6}^z = \{f_1^{(2)}, f_4^{(3)}\}$ and $\mathcal{F}_{c_7}^z = \{f_5^{(1)}, f_1^{(2)}, f_4^{(3)}\}$. If the observed pattern is $D_z(t) = [0, 1, 0]^\top$, the diagnosis set is $\mathcal{D}_s^z(t) = \{\mathcal{F}_{c_2}^z\}$, then the agent $\mathcal{M}^{(2)}$ infers that local sensor faults have occurred. If the observed pattern is $D_z(t) = [0, 1, 1]^\top$, the diagnosis set is $\mathcal{D}_s^z(t) = \{\mathcal{F}_{c_2}^z, \mathcal{F}_{c_3}^z, \mathcal{F}_{c_7}^z\}$, then (i) the agent $\mathcal{M}^{(2)}$ infers that local faults have

occurred or sensor faults have propagated from the agent $\mathcal{M}^{(3)}$, and (ii) the agent $\mathcal{M}^{(3)}$ infers that local faults have occurred or sensor faults have propagated from the agent $\mathcal{M}^{(2)}$ and not from $\mathcal{M}^{(1)}$.

	$\mathcal{F}_{c_1}^z$	$\mathcal{F}_{c_2}^z$	$\mathcal{F}_{c_3}^z$	$\mathcal{F}_{c_4}^z$	$\mathcal{F}_{c_5}^z$	$\mathcal{F}_{c_6}^z$	$\mathcal{F}_{c_7}^z$
$\mathcal{E}^{(1)}$	1	0	*	1	1	*	1
$\mathcal{E}^{(2)}$	0	1	*	1	*	1	1
$\mathcal{E}^{(3)}$	*	*	1	*	1	1	1

Table 5.2: Example sensor fault signature matrix F_z .

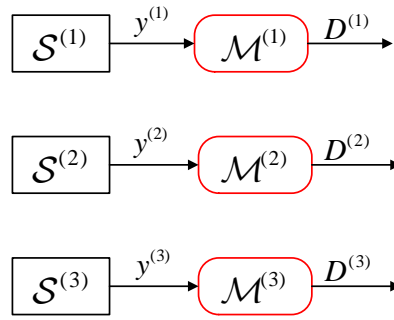
5.3 Other Established Fault Isolation schemes

in this section, we provide an overview of some established observer-based schemes for fault isolation, and some approaches for performing diagnostic reasoning.

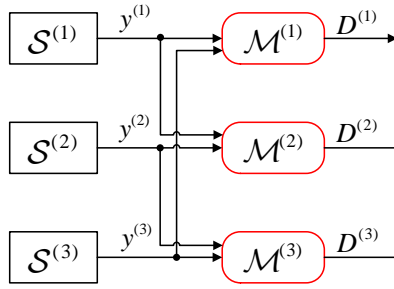
5.3.1 Observer-based Fault Isolation Schemes

A bank of observer-based residual generators is commonly proposed in the literature for the isolation of faults and especially of multiple faults. The isolation of multiple faults may be performed by using a single residual generator, but the design of the residual generator can be in some cases unrealizable [Ding [2008]]. There are two basic isolation schemes that rely on a bank of observers; the dedicated observer-based scheme (DOS), and the generalized observer-based scheme (GOS) [Chen and Patton [1999]]. An illustrative example of the GOS and DOS for sensor fault isolation is shown in Fig. 5.3. For the sake of generality, we describe three sets of sensors instead of three sensors.

For the decision making process of every observer-based module $\mathcal{M}^{(j)}$ that monitors one sensor set (Fig. 5.3-1), or more than one (Fig. 5.3-2), a set of structured residuals $\varepsilon_y^{(j)}$ is generated, and their magnitude is compared to the set of corresponding thresholds $\bar{\varepsilon}_y^{(j)}$ (in a norm based framework, the norm of the structured residual vector is compared to a threshold). As observed in Fig. 5.3-1, every observer-based



(1) Dedicated Observer-based Scheme.



(2) Generalized Observer-based Scheme.

Figure 5.3: Illustrative example of the basic observer-based fault isolation schemes ($\mathcal{S}^{(j)}$: set of sensors, $y^{(j)}$: measurements of $\mathcal{S}^{(j)}$, $\mathcal{M}^{(j)}$: observed-based monitoring module, $D^{(j)}$: decision of $\mathcal{M}^{(j)}$, $j \in \{1, 2, 3\}$).

module $\mathcal{M}^{(j)}$ is designed to monitor one set of sensors $\mathcal{S}^{(j)}$. The decision logic of the monitoring modules deployed in a DOS is the following:

- Under healthy conditions (no faults) of $\mathcal{S}^{(j)}$, every threshold in $\bar{\varepsilon}_y^{(j)}$ should bound the corresponding residual in $\varepsilon_y^{(j)}$.
- If at least one residual exceeds its threshold, then $D^{(j)} = 1$
- If $D^{(j)} = 1$, then the module $\mathcal{M}^{(j)}$ detects the presence of sensor faults in $\mathcal{S}^{(j)}$.

In a GOS scheme (see Fig. 5.3-2), every observed-based module $\mathcal{M}^{(j)}$ is designed to monitor all sets of sensors except for one. The decision logic implemented in the monitoring module $\mathcal{M}^{(j)}$ is the following:

- Under healthy conditions (no faults) of the sets $\mathcal{S}^{(j)}$ monitored by $\mathcal{M}^{(j)}$ (e.g. $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(3)}$ that are monitored by $\mathcal{M}^{(1)}$), every threshold in $\bar{\varepsilon}_y^{(j)}$ should bound the corresponding residual in $\varepsilon_y^{(j)}$.
- If at least one residual exceeds its threshold, then $D^{(j)} = 1$
- If $D^{(j)} = 1$, then the module $\mathcal{M}^{(j)}$ detects the presence of sensor faults in its monitored sensor sets $\mathcal{S}^{(j)}$ (e.g. the module $\mathcal{M}^{(1)}$ detects the presence of faults either in $\mathcal{S}^{(1)}$ or in $\mathcal{S}^{(3)}$ or in both).

Let us assume the special case that only one set of sensors is subject to faults, e.g. $\mathcal{S}^{(1)}$. By applying the DOS, the module $\mathcal{M}^{(1)}$ may detect and isolate the fault(s) in $\mathcal{S}^{(1)}$. In this case there is no need to combine the decisions of the three modules (no need for combinatorial decision logic) in order to isolate faults in $\mathcal{S}^{(1)}$. On the contrary, by applying the GOS, (i) the module $\mathcal{M}^{(1)}$ may detect the presence of faults either in $\mathcal{S}^{(1)}$ or in $\mathcal{S}^{(3)}$, (ii) the module $\mathcal{M}^{(2)}$ may detect the presence of faults either in $\mathcal{S}^{(1)}$ or in $\mathcal{S}^{(2)}$, and (iii) the module $\mathcal{M}^{(3)}$ does not detect any fault in $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(3)}$. In order to isolate the faulty set, the decisions of all modules should be combine for isolating $\mathcal{S}^{(1)}$ by applying the GOS. The difference in the combinatorial decision logic is described in more detail in the Section 5.3.2.

Now let us assume that more than one set of sensors are faulty, e.g. $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$. By applying the DOS, the modules $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ may

detect and isolate the faults in $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$, respectively. By applying the GOS scheme, all modules may detect the faulty sensor sets, making the isolation unfeasible. Thus, the GOS is not suitable for the isolation of multiple faulty sensor sets. On the contrary DOS can be used for the isolation of multiple faulty sets [Adjallah et al. [1994], Rajaraman et al. [2004, 2006]]. Assuming a single faulty sensor set, both schemes are equivalent with respect to the fault isolation process. However, the difficulty in designing the modules $\mathcal{M}^{(j)}$ due to observability issues may be higher in a DOS than in a GOS scheme, since the sensors used for the design of the observer of $\mathcal{M}^{(j)}$ in a DOS is less than in a GOS.

It is worth to discuss that in the literature there are some research works on the application of ‘hybrid’ GOS-like scheme, aiming at overcoming the weakness of the GOS for isolation of multiple faulty sets. For instance, in [Samy et al. [2011]], the authors propose a methodology following a GOS approach in order to detect and isolate consecutive sensor faults (see Section 2.2). Instead of using a bank of observers, they use a bank of neural network (NN) models, where each NN model is driven by all sensors except for one, and the estimation of a sensor output used for generating the residuals is realized by the NN model. When a single sensor fault occurs, it is isolated as faulty by applying the GOS isolation decision logic. Then, assuming that there will be no fault for a time interval, the NN model that is not driven by the isolated faulty sensor provides its (healthy) estimation, which replaces the faulty sensor output in the remainder NN models. In order words, the isolated faulty sensor is switched off, and its estimated output is used in the GOS-like scheme.

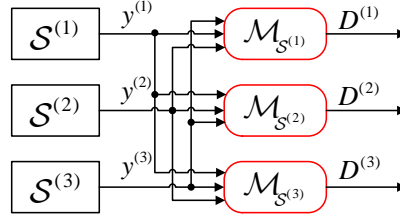
In order to overcome any observability issues that may be raised in the GOS and DOS due to the fact that every module uses less sensor sets than the total sets of sensors, several researchers have proposed an observer-based (or observer-like) scheme, in which all monitoring modules are driven by all sensor sets, but they are designed to be sensitive to some sensor sets. Particularly, the monitoring module, denoted by $\mathcal{M}_{\mathcal{S}^{(i,j)}}$ is *only* sensitive to $\mathcal{S}^{(i,j)} = \{\mathcal{S}^{(i)}, \mathcal{S}^{(j)}\}$ with $\mathcal{S}^{(i,i)} = \{\mathcal{S}^{(i)}\}$. Figure 5.4 presents an example of a GOS-like scheme applied for the sensor set $\mathcal{S} = \{\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \mathcal{S}^{(3)}\}$. The way for ensuring this type of selective

fault sensitivity of the monitoring modules varies.

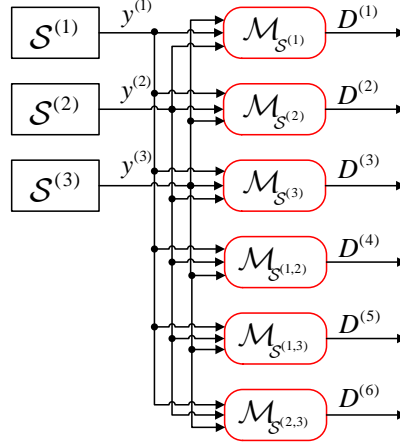
In [Zhang et al. [2008], Zhang [2011], Boem et al. [2013c]], the authors developed a GOS-like methodology, where every module uses an adaptive nonlinear estimation scheme that provides not only the estimation of the system but also the estimation of the sensor fault, and is used for residual generation. Every module is sensitive to a set of sensors in the sense that the adaptive nonlinear estimation scheme is designed under the working hypothesis that only this sensor set is faulty. If this hypothesis is valid then this module will estimate the faults in this sensor set. The adaptive thresholds are then designed under the aforementioned working hypothesis, and using an upper bound on the parameter estimation error, i.e. the error between the faults and their estimations. Then a faulty set is isolated if all modules detect faults, except for the module that is sensitive to this specific set. For example based on Fig. 5.4-1, when $\mathcal{M}_{\mathcal{S}^{(1)}}$ and $\mathcal{M}_{\mathcal{S}^{(2)}}$ detect faults (since at least one residual may exceed the corresponding threshold) and $\mathcal{M}_{\mathcal{S}^{(3)}}$ does not detect any fault because all the residuals remain below the corresponding thresholds, then it is inferred that the sensor set $\mathcal{S}^{(3)}$ is faulty.

In [Chen and Saif [2007]], the authors designed a GOS-based methodology, where every module $\mathcal{M}^{(j)}$ uses a nonlinear sliding mode observer, which is sensitive to faults that occur in the monitored set $\mathcal{S}^{(j)}$, but is driven by the remainder sensors $\mathcal{S} \setminus \mathcal{S}^{(j)}$. Similarly to [Zhang et al. [2008], Zhang [2011], Boem et al. [2013c]], the thresholds in $\mathcal{M}_{\mathcal{S}^{(j)}}$ are designed to bound the residuals under the assumption that $\mathcal{S}^{(j)}$ is the only faulty set. Then a faulty set is isolated if all modules detect faults, except for the module that is sensitive to this specific set.

In [Ding [2008]], a bank of linear residual generators is proposed, where every module is driven by the outputs of all the sensor sets but it is sensitive to faults occurring in a specific number of sensor sets, and it is decoupled from the effects of faults in the remainder sensor sets. In order to decouple the fault effects, the author in [Ding [2008]] proposes the use of an unknown input observer by every monitoring module that is capable of decoupling the fault effects of particular sensors by treating them as unknown disturbances. Given m sensors, for



(1) For isolation of a single faulty sensor set.



(2) For isolation of multiple faulty sensor sets.

Figure 5.4: Illustrative example of an observer-based scheme with selective sensitivity of modules. In this scheme, the monitoring module $\mathcal{M}_{\mathcal{S}^{(i,j)}}$ is sensitive to $\mathcal{S}^{(i,j)} = \{\mathcal{S}^{(i)}, \mathcal{S}^{(j)}\}$ with $\mathcal{S}^{(i,i)} = \{\mathcal{S}^{(i)}\}$, and decoupled from the effects of faults occur in $\{\mathcal{S} \setminus \mathcal{S}^{(i,j)}\}$.

isolating $k^* = 2^m - 2$ possible combinations of sensor faults, k^* UIO-based monitoring modules need to be designed such that each observer is totally decoupled from $k - 1$ sensor fault combinations. In this way, every module may detect and isolate a single faulty set, since it is sensitive to sensor faults in this set only. For example, based on Fig. 5.4-2, six monitoring modules are designed for isolating one or more faulty sensor sets.

Taking into account the centralized architecture in Section 3.1, the fault isolation scheme proposed in this tutorial can be characterized as a dedicated observer-like scheme, but with some additional treatment in order to handle the observer stability (see Algorithm 1). Particularly, every monitoring module is designed to be sensitive to sensor faults in one sensor set, and may detect and isolate this sensor set as faulty, as proposed in [Ding [2008]]. The difference between the two schemes is in the number of modules needed for isolation of multiple faulty sets. This can be observed taking into account Fig. 5.3-1 and Fig. 5.4-2.

It is noted that the grouping of the sensors, i.e. the sensor sets can help the isolation process. In this way, the isolation process is realized in more than one stage following a top-down design. Particularly, we start with seeking faulty subsets of sensors that belong in a large set of sensors, and then we continue with searching for the faulty sensors within the faulty subsets. In some cases, the isolation of a faulty set of sensors, instead isolating the faulty sensors in the set, can be enough in order to activate some remedial actions necessary for compensating fast the fault effects. If the goal is to isolate the faulty sensors, then the overlapping of the faulty sensor sets, i.e. some sensors may belong to more than one sensor sets, can contribute to achieve this goal. However, in this case it is necessary to combine the decisions of the monitoring modules and apply a combinatorial diagnostic reasoning to obtain the final decision. There are different ways to design the combinatorial diagnostic reasoning. In [Chen and Saif [2007]], the authors developed a combinatorial decision logic based on the integer solution of algebraic equations that involve combinations. In [Ding [2008]] and in this tutorial, the combinatorial decision logic relies on a fault signature matrix, as analytically described in Section 5.3.2. The diagnostic reasoning may

vary with respect to the underlying assumptions and the goals of the diagnosis.

5.3.2 Combinatorial Diagnostic Reasoning

One of the popular approaches to perform combinatorial diagnostic reasoning is using a fault signature matrix, as discussed in Section 5.1.1. Following this approach, the observed pattern is compared with the theoretical patterns, i.e. the elements of the fault signature matrix. There are two ways to carry out this comparison: (i) the column-based reasoning, and (ii) the row-based reasoning [Cordier et al. [2004]]. While the first technique is mostly applied by the FDI community, the second technique is applied by the DX community. It is worth to note that the common tools used by the FDI community can be found in the field of automatic control, such as stability theory, analytical mathematical analysis, system engineering. On the other hand DX community uses tools from artificial intelligence, such as logic, search, knowledge representation and reasoning. The last ten years, there is a significant effort to combine the FDI- and DX-based tools in order to enhance the fault detection and isolation process [Travé-Massuyès [2014]].

In order to exemplify the two different approaches to diagnostic reasoning, let us consider the fault signature matrix shown in Table 5.3, where the term ‘ARR- j ’ stands for the j -th ARR, $j = 1, 2, 3$ and we assume 5 faults, e.g. sensor faults. Table 5.3 shows the five single faults, and one combination of multiple sensor faults. Now, let us consider that

Table 5.3: Illustrative example of a fault signature matrix ($\{\dots\}$ represents of the remainder combinations of sensor faults)

	f_1	f_2	f_3	f_4	f_5	$\{f_1, f_3\}$	$\{\dots\}$
ARR-1	1	0	1	1	0	1	1
ARR-2	0	1	0	1	1	0	1
ARR-3	1	1	1	0	1	1	1

the observed pattern is

$$D(t) = \begin{bmatrix} 1, 0, 1, \end{bmatrix}^T \quad (5.14)$$

When the fault isolation decision is obtained

- following the column-based approach, the observed pattern is compared to every column of the fault signature matrix, and the resultant diagnosis set is

$$\mathcal{D}_s(t) = \{\{f_1\}, \{f_3\}, \{f_1, f_3\}\} \quad (5.15)$$

- following the row-based approach, the diagnosis set is obtained by using the support of the violated analytical redundancy relations, which in this case, are ARR-1 and ARR-3; i.e. $Supp(ARR-1) = \{\{f_1\}, \{f_3\}, \{f_4\}, \{f_1, f_3\}, \{\dots\},\}$ and $Supp(ARR-3) = \{\{f_1\}, \{f_2\}, \{f_3\}, \{f_5\}, \{f_1, f_3\}, \{\dots\},\}$. Then the minimal diagnosis set is determined as:

$$\mathcal{D}_s^{\min}(t) = \{\{f_1\}, \{f_3\}, \{f_2, f_4\}, \{f_4, f_5\}\} \quad (5.16)$$

while the (possible) diagnosis set \mathcal{D}_s consists of all the combinations of the elements of \mathcal{D}_s^{\min} , i.e. the minimal diagnoses.

The difference in the diagnosis sets obtained following a column- and row-based approach stems from the fact that the column-based approach is formulated in an exoneration framework (see Section 4.3) and uses both the knowledge of an ‘1’ and ‘0’ in the diagnosis process. Particularly, ‘1’ in the fault signature matrix implies that the ARR- j is necessarily violated when the faults or fault combinations that belong to its support occur. So when the q -th element of the observed pattern D^q is ‘1’, then at least one of the fault combinations in the support of the violated ARR- q is possible to have occurred. At the same time when the q -th element of the observed pattern D^q is ‘0’, the combinations of faults in the support of the non-violated ARR- q are not considered as possible to have occurred. On the other hand, the row-based diagnostic reasoning approach ignores the information that D^q is ‘0’.

Comparing the column-based and row-based diagnosis set, we can deduce that the cardinality of the column-based diagnosis set is less than the cardinality of the row-based diagnosis set. The smaller the cardinality the smaller the ambiguity in the diagnosis, since less combinations of faults are possible to have occurred. On the other hand

taking into account the exoneration framework (see Section 4.3), the column-based diagnosis reasoning is not robust with respect to delays in the violation of the ARR_s, or even the non-violation of some ARR_s. The fault signature matrix depicts in an implicit way (structurally) the sensitivity of the ARR_s to every fault combination. This means that when a single fault or fault combination can provoke the violation of more than one ARR, according to the fault signature matrix (positions of ‘1’) *all* ARR_s should be violated *at the same time*. In practice, this is not always the case, and some ARR_s can be more sensitive to the occurrence of a fault combination at a specific time instant than other ARR_s.

The delayed violation of some ARR_s may lead to incorrect fault isolation for a certain time interval since the diagnosis set is designed to change over time (in both column- and row-based reasoning approaches). Let us consider the fault signature matrix shown in Table 5.3 and that the time instant t_1 the observed pattern $D(t_1) = [1, 0, 1]^T$. In this case, the column-based diagnosis set $\mathcal{D}_s(t_1)$ will be equal to the right-hand side of (5.15). At the time instant t_2 with $t_2 > t_1$, the observed pattern $D(t_1)$ may become $D(t_2) = [1, 1, 1]^T$. Then, the resultant diagnosis set is equal to the remainder fault combinations that do not appear in Table 5.3. Among them is the fault combination e.g. $\{f_2, f_4\}$, which is not included in the diagnosis set $\mathcal{D}_s(t_1)$. This means that the fault combination $\{f_2, f_4\}$ may have actually occurred, but there was a delay in the violation of ARR-2, while ARR-1 and ARR-3 were violated at the same time. It could be even worse if ARR-2 would never be violated, so the outcome of the diagnosis would be the set $\mathcal{D}_s(t_1)$. A wrong fault isolation may provoke the activation of wrong counteracting actions for compensating the fault effects.

Following a row-based diagnostic reasoning approach in this example, the diagnosis set at the time instant t_1 would contain the fault combination $\{f_2, f_4\}$ and many more, including the column-based diagnosis set. In this way, we would not obtain a wrong fault isolation decision, because the fault isolation decision is very conservative. The conservativeness of the row-based diagnostic reasoning may delay the activation of the appropriate remedial actions, because it leads to a

large number of fault combinations that have possibly occurred.

The delayed violation or even the non-violation of some ARR_s that may impact the effectiveness of the column-based diagnosis reasoning is a design issue, which should be taken into account during the design of both the fault detection and isolation procedure. At the first stage (fault detection), it is necessary to perform a systematic fault sensitivity analysis for every ARR, or equivalently, fault detectability analysis for every monitoring module in order to ensure that all ARR_s are sensitive to a fault combination in the same way and at the same time. At this stage, we should ensure that every ARR is sensitive to all fault combinations, and exclude the case of the non-violation of ARR_s. At the second stage (fault isolation), we may wait for a time interval (dwell-time) before obtaining the decision based on the resultant diagnosis set. The dwell-time can be defined taking into account that the residual generation is a dynamic process, and that both the residuals and the adaptive thresholds are evolving over time. In other words, the design of the nonlinear observer and the adaptive thresholds plays an important role in the determination of the dwell-time. In any case a larger value for the dwell time can be selected, but it may delay the activation of remedial actions.

An interesting approach for enhancing the fault isolation process has been proposed in [Meseguer et al. [2010]]. In this research work, the authors proposed an interval observer-based fault diagnosis technique, where the fault isolation is realized by using four fault signature matrices that are related to: (i) the fault sensitivity, (ii) binary fault sensitivity (iii) the time interval within which a fault signal (which represents a degree of satisfaction of an analytical redundancy relation of residuals and adaptive thresholds) is expected to be violated, and (iv) the order according to which the fault signals are activated. In the first case, the element $\{i, j\}$ of the fault signature matrix contains a real value that describes how easily the j -th fault will cause the i -th residual to exceed its corresponding threshold, i.e. how easily the i -th fault signal is activated. In the second case, the binary fault signature matrix results from the ‘fault sensitivity’ signature matrix, by replacing the non-zero values of this matrix with ones. In the third case,

the element $\{i, j\}$ of the fault signature matrix includes a time interval within which the i -th fault signal is expected to be activated after the occurrence of the j -th fault. In the latter case, the j -th column of the fault signature matrix contains the theoretical order of activation of the fault signals after the occurrence of the j -th fault. The computation of these matrices requires the knowledge of time of fault occurrence, and an upper and lower bound of every fault. The first requirement is very difficult to be satisfied; therefore the computations are realized by using the first time instant that an observed fault signal is activated.

The combinatorial diagnostic reasoning proposed in this tutorial relies on an interplay between the column- and row-based approaches. Specifically, in the centralized and decentralized architecture, and in the local monitoring agents of the distributed architecture, when the observed sensor fault pattern matches at least one of the theoretical sensor fault patterns, then the column-based reasoning is applied, otherwise the row-based diagnostic reasoning is used. The combinatorial diagnostic reasoning implemented in the global monitoring agent is based on a fault signature matrix that contains ‘1’ and ‘*’, aiming at discriminating the sensitivity of an ARR to local and propagated faults respectively. The use of ‘1’ in the $\{i, j\}$ element of the fault signature matrix means that the i -th ARR is necessarily violated when the j -th local fault combination occurs, while ‘*’ means that the i -th ARR might be violated when the j -th fault combination is propagated from another system. The problem of delayed violation or non-violation of some ARRs is tackled by performing a fault detectability and isolability analysis described in detail in Chapter 6. The rationale for performing this analysis is that the monitoring modules *should be designed to be sensitive* to any fault combination that affects their associated sensor sets under worst-case assumptions of modeling uncertainties, i.e. the design parameters of the nonlinear observer and the adaptive thresholds are selected such that the sensitivity of the ARRs to specific sets of faults is guaranteed.

5.4 Illustrative Example

The objective of this section is to illustrate the multiple sensor fault isolation decision logic for the two interconnected systems presented in Example 2.2, taking into account the distributed sensor fault detection process presented in Section 4.5. Aiming at isolating multiple sensor faults that may affect the two interconnected subsystems, the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ use the aggregation modules $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$, which collect and process combinatorially the decisions of the modules $\mathcal{M}^{(1,1)}$, $\mathcal{M}^{(1,2)}$, and $\mathcal{M}^{(2,1)}$, $\mathcal{M}^{(2,2)}$, respectively, as shown in Fig. 5.5.

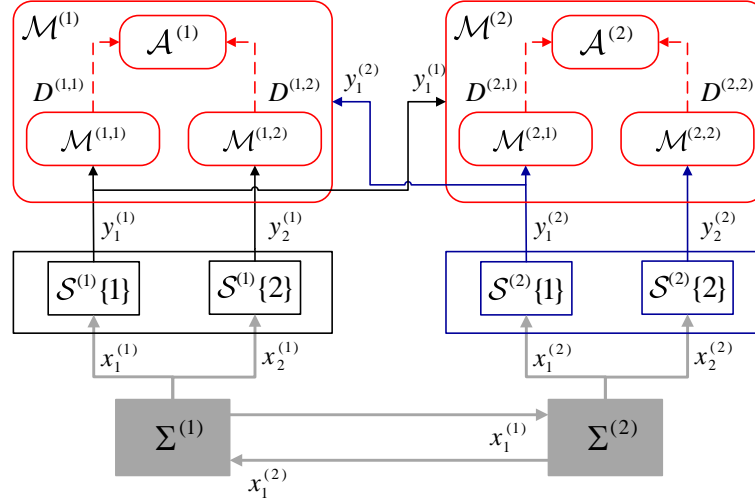


Figure 5.5: Illustrative example: First-level distributed sensor fault diagnosis scheme for two interconnected systems.

Following a decentralized sensor fault diagnosis architecture for the two interconnected nonlinear systems, the aggregation module $\mathcal{A}^{(I)}$ can isolate locally sensor faults affecting the subsystem $\Sigma^{(I)}$ for all $I = 1, 2$. On the other hand, following a distributed sensor fault diagnosis architecture, the aggregation module $\mathcal{A}^{(I)}$ should convey a decision on the propagation of sensor faults to the global aggregation module \mathcal{G} , as shown in Fig. 5.6.

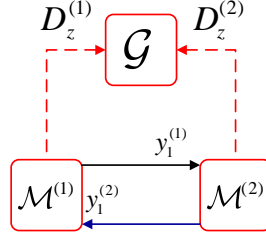


Figure 5.6: Illustrative example: Second-level distributed sensor fault diagnosis scheme for two interconnected systems.

5.4.1 Local Sensor Fault Isolation Decision Logic

The observed pattern of sensor faults used by the aggregation module $\mathcal{A}^{(I)}$ is defined as

$$D^{(I)}(t) = \left[D^{(I,1)}(t), D^{(I,2)}(t) \right]^\top, \quad I = 1, 2. \quad (5.17)$$

The observed pattern is compared at every time instant to the theoretical patterns of sensor faults that correspond to the columns of the matrix $F^{(I)}$.

Decentralized Configuration of Sensor Fault Signature Matrix

In a decentralized sensor fault diagnosis scheme, the sensor fault signature matrix is determined in Table 5.4. Comparing the observed pattern $D^{(I)}$ to the columns of the sensor fault signature matrix $F^{(I)}$ results in the following diagnosis set:

- If $D^{(I)}(t) = [1, 0]^\top$, then $\mathcal{D}_s^{(I)}(t) = \{f_1^{(I)}\}$
- If $D^{(I)}(t) = [0, 1]^\top$, then $\mathcal{D}_s^{(I)}(t) = \{f_2^{(I)}\}$
- If $D^{(I)}(t) = [1, 1]^\top$, then $\mathcal{D}_s^{(I)}(t) = \left\{ \{f_1^{(I)}, f_2^{(I)}\} \right\}$

In any case, the diagnosis set contains one sensor fault combination, implying the isolation of multiple local sensor faults.

	$\{f_1^{(I)}\}$	$\{f_2^{(I)}\}$	$\{f_1^{(I)}, f_2^{(I)}\}$
$\mathcal{E}^{(I,1)}$	1	0	1
$\mathcal{E}^{(I,2)}$	0	1	1

Table 5.4: Decentralized configuration of the fault signature matrix $F^{(I)}$, $I = 1, 2$.**Distributed Configuration of Sensor Fault Signature Matrix**

In a distributed sensor fault diagnosis scheme, the sensor fault signature matrix $F^{(I)}$ is formulated, taking into account the matrix in Table 5.4 and the effects of the transmitted sensor faults; i.e. the matrices $F^{(1)}$ and $F^{(2)}$ are shown in Tables 5.5 and 5.6.

	$\{f_1^{(1)}\}$	$\{f_2^{(1)}\}$	$\{f_1^{(1)}, f_2^{(1)}\}$	$\{f_1^{(2)}\}$	$\bigcup_{i \in \{1,2,3\}} \{f_1^{(2)}, \mathcal{F}_{c_i}^{(1)}\}$
$\mathcal{E}^{(1,1)}$	1	0	1	1	1
$\mathcal{E}^{(1,2)}$	0	1	1	1	1

Table 5.5: Distributed configuration of sensor fault signature matrix $F^{(1)}$ ($\mathcal{F}_{c_1}^{(1)} = \{f_1^{(1)}\}$, $\mathcal{F}_{c_2}^{(1)} = \{f_2^{(1)}\}$ and $\mathcal{F}_{c_3}^{(1)} = \{f_1^{(1)}, f_2^{(1)}\}$).

	$\{f_1^{(2)}\}$	$\{f_2^{(2)}\}$	$\{f_1^{(2)}, f_2^{(2)}\}$	$\{f_1^{(1)}\}$	$\bigcup_{i \in \{1,2,3\}} \{f_1^{(1)}, \mathcal{F}_{c_i}^{(2)}\}$
$\mathcal{E}^{(2,1)}$	1	0	1	1	1
$\mathcal{E}^{(2,2)}$	0	1	1	1	1

Table 5.6: Distributed configuration of sensor fault signature matrix $F^{(2)}$ ($\mathcal{F}_{c_1}^{(2)} = \{f_1^{(2)}\}$, $\mathcal{F}_{c_2}^{(2)} = \{f_2^{(2)}\}$ and $\mathcal{F}_{c_3}^{(2)} = \{f_1^{(2)}, f_2^{(2)}\}$).

Comparing the observed pattern $D^{(I)}$ to the columns of the sensor fault signature matrix $F^{(I)}$ shown in Tables 5.5 and 5.6 results in the following diagnosis set:

- If $D^{(I)}(t) = [1, 0]^\top$, then $\mathcal{D}_s^{(I)}(t) = \{f_1^{(I)}\}$
- If $D^{(I)}(t) = [0, 1]^\top$, then $\mathcal{D}_s^{(I)}(t) = \{f_2^{(I)}\}$
- If $D^{(I)}(t) = [1, 1]^\top$, then $\mathcal{D}_s^{(I)}(t) = \{f_1^{(I)}, f_2^{(I)}\}$

$$\left\{ \left\{ f_1^{(I)}, f_2^{(I)} \right\}, \left\{ f_1^{(Q)} \right\}, \bigcup_{i \in \{1,2,3\}} \left\{ f_1^{(Q)}, \mathcal{F}_{c_i}^{(I)} \right\} \right\} \quad \text{for } I \neq Q, \\ I, Q = 1, 2$$

When $D^{(I)}(t) = [1, 1]^\top$, the diagnosis set contains more than one sensor fault combination. The diagnosed sensor fault combinations are considered to have possibly occurred. If $D^{(1)}(t) = [1, 1]^\top$, it is inferred that there are multiple local sensor faults affecting the subsystem $\Sigma^{(1)}$ or sensor fault $f_1^{(2)}$ has occurred (possibly affecting the decision of $\mathcal{M}^{(1)}$) or both local and propagated faults have occurred, that is $\{f_1^{(1)}, f_1^{(2)}\}$ or $\{f_2^{(1)}, f_1^{(2)}\}$ or $\{\{f_1^{(2)}, f_2^{(2)}\}, f_1^{(2)}\}$ has possibly occurred.

The decision on the propagation of sensor faults of the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ are represented by the following decision functions

$$D_z^{(1)}(t) = \begin{cases} 0, & \text{if } f_1^{(2)} \notin \mathcal{D}_s^{(1)}(t) \text{ and } f_1^{(1)} \notin \mathcal{D}_s^{(1)}(t) \\ 1, & \text{if } f_1^{(2)} \in \mathcal{D}_s^{(1)}(t) \text{ or } f_1^{(1)} \in \mathcal{D}_s^{(1)}(t) \end{cases} \quad (5.18)$$

$$D_z^{(2)}(t) = \begin{cases} 0, & \text{if } f_1^{(2)} \notin \mathcal{D}_s^{(1)}(t) \text{ and } f_1^{(2)} \notin \mathcal{D}_s^{(2)}(t) \\ 1, & \text{if } f_1^{(2)} \in \mathcal{D}_s^{(1)}(t) \text{ or } f_1^{(2)} \in \mathcal{D}_s^{(2)}(t) \end{cases} \quad (5.19)$$

5.4.2 Global Sensor Fault Isolation Decision Logic

The aggregation module compares the observed pattern of propagated sensor faults, determined as

$$D_z(t) = [D_z^{(1)}(t), D_z^{(2)}(t)]^\top, \quad (5.20)$$

to the columns of the sensor fault signature matrix F^z , shown in Table 5.7, where $\mathcal{E}^{(I)} = \mathcal{E}^{(I,1)} \cup \mathcal{E}^{(I,2)}$ for $I = 1, 2$.

	$\{f_1^{(1)}\}$	$\{f_1^{(2)}\}$	$\{f_1^{(1)}, f_1^{(2)}\}$
$\mathcal{E}^{(1)}$	1	*	1
$\mathcal{E}^{(2)}$	*	1	1

Table 5.7: Global sensor fault signature matrix F^z , $I = 1, 2$.

The diagnosis process is realized as follows:

- If $D_z(t) = [1, 0]^\top$, then $\mathcal{D}_s^z(t) = \{f_1^{(1)}\}$
- If $D_z(t) = [0, 1]^\top$, then $\mathcal{D}_s^z(t) = \{f_1^{(2)}\}$
- If $D_z(t) = [1, 1]^\top$, then $\mathcal{D}_s^z(t) = \left\{ \{f_1^{(1)}\}, \{f_1^{(2)}\}, \{f_1^{(1)}, f_1^{(2)}\} \right\}$

If $\mathcal{D}_s^z(t) = \{f_1^{(1)}\}$ and the diagnosis set of the agent $\mathcal{M}^{(1)}$ is $\mathcal{D}_s^{(1)}(t) = \{f_1^{(1)}\}$ (leading to $D_z^{(1)} = 1$) then it is inferred that $f_1^{(1)}$ has occurred. If $\mathcal{D}_s^z(t) = \{f_1^{(1)}\}$ and $\mathcal{D}_s^{(1)}(t) = \left\{ \{f_1^{(1)}, f_2^{(1)}\}, \{f_1^{(2)}\}, \bigcup_{i \in \{1,2,3\}} \{f_1^{(2)}, \mathcal{F}_{c_i}^{(1)}\} \right\}$, the agent $\mathcal{M}^{(1)}$ infers that both local sensor faults $f_1^{(1)}$ and $f_2^{(1)}$ have occurred and there was no propagation of sensors faults. The diagnostic reasoning in the case that $\mathcal{D}_s^z(t) = \{f_1^{(2)}\}$ is similar to the diagnostic reasoning when $\mathcal{D}_s^z(t) = \{f_1^{(1)}\}$. In the case that $\mathcal{D}_s^z(t) = \left\{ \{f_1^{(1)}\}, \{f_1^{(2)}\}, \{f_1^{(1)}, f_1^{(2)}\} \right\}$, it is inferred that either sensor fault $f_1^{(1)}$ (correspondingly for $f_1^{(2)}$) has occurred in $\Sigma^{(1)}$ with its effects on the decision of the agent $\mathcal{M}^{(2)}$ being strong enough to cause the violation of both $\mathcal{E}^{(2,1)}$ and $\mathcal{E}^{(2,2)}$ or both subsystems $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are affected by local sensor faults.

5.4.3 Simulation Results

This section presents the local and global decision logic of the distributed and decentralized sensor fault diagnosis method presented in Sections 4.5.1, 4.5.2 and 4.5.3 to the two nonlinear interconnected subsystems given in (4.165) and (4.169). Note that the implementation of the two agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ has been realized using the system characteristics and design parameters described in Section (4.5.4). According to the simulation results of the four sensor fault scenarios (see Section (4.5.4)) shown in Figs. 4.5, 4.6 and 4.7, the single occurrence of sensor faults $f_1^{(1)}$ and $f_1^{(2)}$ have been successfully detected by the modules $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(2,1)}$ respectively in both sensor fault diagnosis schemes. On the other hand, the single occurrence of sensor faults $f_2^{(1)}$

and $f_2^{(2)}$ was respectively detected by the modules $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,2)}$ only in the distributed architecture.

In the first simulated scenario, the decisions of the two modules $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(1,2)}$ constitute the observed pattern of sensor faults that equals to $D^{(1)}(t) = [1 \ 0]^\top$ for $t \geq 150.0023$ sec in the distributed scheme and for $t \geq 150.0032$ sec in the decentralized scheme. Then, $D^{(1)}$ is compared to the columns of the sensor fault signature matrix $F^{(1)}$ shown in Table 5.4 in the case of the decentralized architecture, and Table 5.5 in the case of the distributed architecture. In both cases, the diagnosis set equals to $\mathcal{D}_s^{(1)}(t) = \{f_1^{(1)}\}$. Similar diagnosis decision making process is applied by the agent $\mathcal{M}^{(2)}$ for isolating the sensor fault $f_1^{(2)}$, resulting in the diagnosis set $\mathcal{D}_s^{(2)}(t) = \{f_1^{(2)}\}$ for $t \geq 150.0023$ sec in the distributed scheme and for $t \geq 150.0032$ sec in the decentralized scheme. In the third simulated scenario, the observed pattern is $D^{(1)}(t) = [0 \ 1]^\top$ for $t \geq 150.0023$ sec in the distributed scheme, while in the decentralized scheme $D^{(1)}(t) = [0 \ 0]^\top$. The comparison of $D^{(1)}$ to the columns of the sensor fault signature matrix $F^{(1)}$ shown in Table 5.4 leads to the diagnosis set $\mathcal{D}_s^{(1)}(t) = \{f_2^{(1)}\}$. The diagnosis results of the forth simulated scenario are similar to the results of the third simulated scenario, which are based on the same decision logic. The resultant diagnosis set is $\mathcal{D}_s^{(2)}(t) = \{f_2^{(2)}\}$ for $t \geq 150.0023$ sec.

We have simulated four scenarios of multiple sensor faults, which are abrupt and permanent faults with evolution rate $\kappa_{f_q}^{(I)} = 10^4$ and fault function $\phi_q^{(I)}(t) = 25\%x_1^\circ = \pi/2$, for all $I, q = 1, 2$. The new simulated scenarios are shown below:

Sensor fault scenario 5: Occurrence of sensor faults $f_1^{(2)}$ and $f_2^{(2)}$ at $T_{f_1}^{(1)} = 150$ sec.

Sensor fault scenario 6: Occurrence of sensor faults $f_2^{(1)}$ and $f_2^{(2)}$ at $T_{f_1}^{(1)} = 150$ sec.

Sensor fault scenario 7: Occurrence of sensor faults $f_1^{(1)}$, $f_2^{(1)}$ at $T_{f_1}^{(1)} = 150$ sec, and $f_1^{(2)}$ at $T_{f_1}^{(1)} = 170$ sec.

Sensor fault scenario 8: Occurrence of sensor faults $f_1^{(1)}$, $f_2^{(1)}$ and $f_1^{(2)}$ at $T_{f_1}^{(1)} = 150$ sec.

Figure 5.7-1 presents the distributed decision making process of

the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ for the simulated scenario 5. Based on these results, the observed pattern of $\mathcal{M}^{(2)}$ is $D^{(2)}(t) = [1 \ 1]^\top$ for $t \geq 150.0023$ sec. Comparing $D^{(2)}$ to the columns of the sensor fault signature matrix shown in Table 5.6, we obtain the diagnosis set $\mathcal{D}_s^{(2)}(t) = \left\{ \{f_1^{(2)}, f_2^{(2)}\}, \{f_1^{(1)}\}, \bigcup_{i \in \{1,2,3\}} \{f_1^{(1)}, \mathcal{F}_{c_i}^{(2)}\} \right\}$. Thus, in order for the agent $\mathcal{M}^{(2)}$ to distinguish between local and propagated sensor faults, it transmits its decision $D_z^{(2)} = 1$ to the global diagnosis module \mathcal{G} . Then the module \mathcal{G} collects also the decision of $\mathcal{M}^{(1)}$, which is $D_z^{(1)} = 0$ (since $D^{(1)}(t) = [0 \ 0]^\top$), formulating the observed pattern of propagated sensor faults $D_z(t) = [0 \ 1]^\top$. By comparing $D_z(t)$ to the columns of the fault signature matrix F^z , we obtain the diagnosis set $\mathcal{D}_z^s(t) = \{f_1^{(2)}\}$. Since $f_1^{(1)} \notin \mathcal{D}_z^s$, the diagnosis set of $\mathcal{M}^{(2)}$ is updated by excluding the combinations of sensor faults containing $f_1^{(1)}$, implying that $\mathcal{D}_s^{(2)}(t) = \left\{ \{f_1^{(2)}, f_2^{(2)}\} \right\}$. The decision process of the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ for the 6th sensor fault scenario is shown in Fig. 5.7-2. For $t \geq 150.0023$ sec both the observed patterns of the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ are $D^{(1)}(t) = D^{(2)}(t) = [0 \ 1]^\top$, which is compared to the fault signature matrix shown in Table 5.6, leading to the diagnosis set $\mathcal{D}_s^{(I)}(t) = \{f_2^{(I)}\}$ for $I = 1$ and $I = 2$. Since the measurements of the sensor $\mathcal{S}^{(1)}\{2\}$ are not transmitted to the agent $\mathcal{M}^{(2)}$ and vice versa, there is no reason for activating the global decision logic.

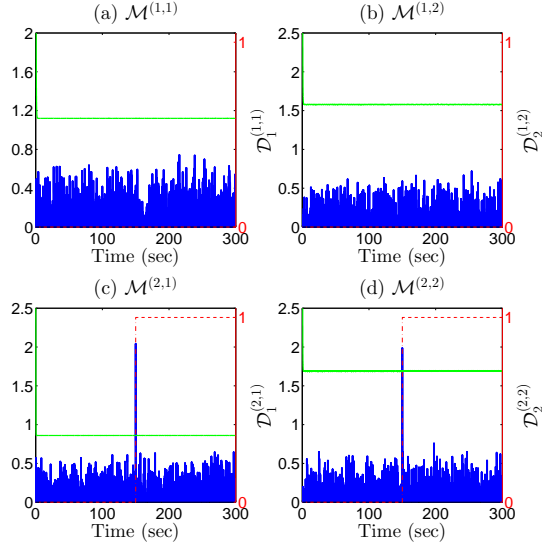
Figure 5.8-1 illustrates the simulation results of the 7th fault scenario. The diagnosis set of agent $\mathcal{M}^{(1)}$ after $t \geq 150.0023$ is $\mathcal{D}_s^{(1)}(t) = \left\{ \{f_1^{(1)}, f_2^{(1)}\}, \{f_1^{(2)}\}, \bigcup_{i \in \{1,2,3\}} \{f_1^{(2)}, \mathcal{F}_{c_i}^{(1)}\} \right\}$ and the decision about the propagation of sensor faults is $D_z^{(1)}(t) = 1$. Thus, the global aggregation module is activated and the observed pattern of propagated sensor faults is $D_z(t) = [1, 0]^\top$ for $t \in [150.0023, 170.0062)$ sec, implying that the diagnosis set of propagated sensor faults is $\mathcal{D}_z^s(t) = \{f_1^{(1)}\}$. By combining $\mathcal{D}_s^{(1)}(t)$, $\mathcal{D}_z^s(t)$, we exclude the occurrence of $f_1^{(1)}$ and we conclude that both sensor faults $f_1^{(1)}$ and $f_2^{(1)}$ have occurred. After $t = 170.0062$, the observed pattern of the agent $\mathcal{M}^{(2)}$ becomes $D^{(2)}(t) = [1 \ 0]^\top$ leading to the diagnosis set $\mathcal{D}_s^{(2)}(t) = \{f_1^{(2)}\}$ and the isolation of the local fault

by the agent $\mathcal{M}^{(2)}$ without activating the global aggregation module.

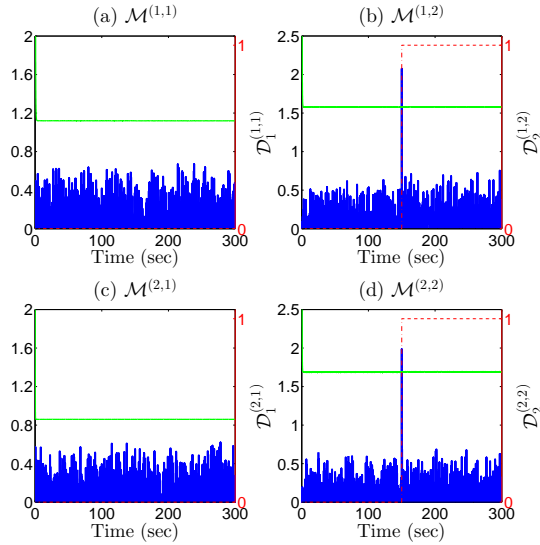
The results of the 8th simulated sensor fault scenario are depicted in Fig. 5.8-2. At the time instant $t = 150.0023$, both monitoring agents detect faults and their corresponding aggregation modules collect the decisions of the monitoring modules, which constitute the following observed patterns of sensor faults: $D^{(1)}(t) = [1 \ 1]^\top$ and $D^{(2)}(t) = [1 \ 0]^\top$. By comparing the observed patterns to the columns of $F^{(I)}$ shown in Table 5.6, we obtain the diagnosis sets $\mathcal{D}_s^{(1)}(t) = \left\{ \{f_1^{(1)}, f_2^{(1)}\}, \{f_1^{(2)}\}, \bigcup_{i \in \{1,2,3\}} \{f_1^{(2)}, \mathcal{F}_{c_i}^{(1)}\} \right\}$ and $\mathcal{D}_s^{(2)}(t) = \left\{ \{f_1^{(2)}\} \right\}$. The global aggregation module collects the decisions of $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, which form the observed pattern $D_z(t) = [1, 1]^\top$ leading to $\mathcal{D}_s^z(t) = \left\{ \{f_1^{(1)}\}, \{f_1^{(2)}\}, \{f_1^{(1)}, f_1^{(2)}\} \right\}$. In this scenario the global isolation decision is that the fault combinations that are possible to have occurred are $\{f_1^{(1)}, f_2^{(1)}\}, \{f_1^{(2)}\}, \bigcup_{i \in \{1,2,3\}} \{f_1^{(2)}, \mathcal{F}_{c_i}^{(1)}\}$.

5.5 Summary and Discussion

In this chapter, we presented the decision-making process for isolating multiple sensor faults in a network of cyber-physical systems. Irrespective of the diagnostic architecture, the multiple sensor fault isolation is realized through aggregation modules that collect the decisions of the monitoring modules locally, and the decisions of the monitoring agents globally. Then, the aggregation modules process the decisions based on combinatorial logic and diagnostic reasoning that are formulated based on structured robustness and sensor fault sensitivity. In a decentralized architecture (correspondingly for a centralized architecture), the goal of the local aggregation module is to isolate the combination of sensor faults that have occurred in the sensor set used for monitoring and controlling each of the interconnected subsystems (the monolithic system in a centralized approach). In a distributed architecture, the local aggregation module may infer the possible sensor fault propagation due to exchange of sensor information, while the global aggregation module aims to isolate multiple propagated sensor faults.

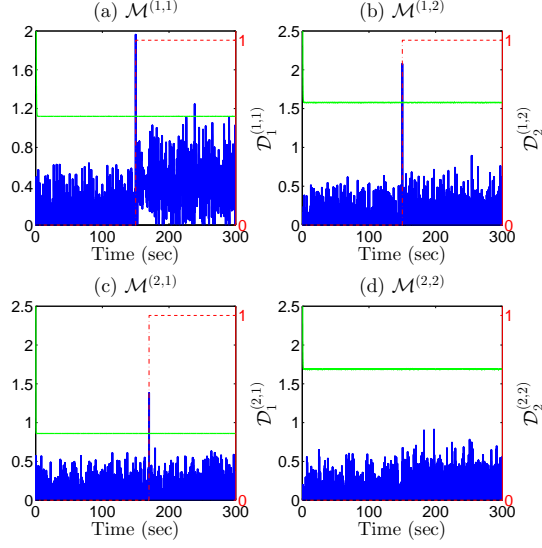


(1) Simulated scenario 5

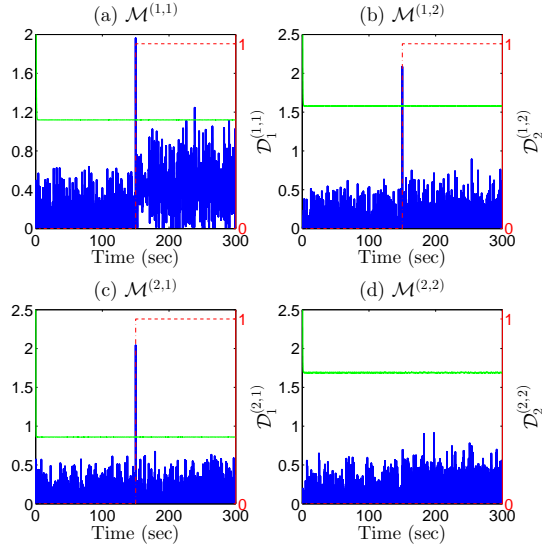


(2) Simulated scenario 6

Figure 5.7: Distributed decision making process of the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ for the simulated scenarios 5 and 6. Subfigures (a)-(d) show the temporal evolution of $|\varepsilon_{y_q}^{(I,q)}(t)|$ (blue line), $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (green line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.



(1) Simulated scenario 7



(2) Simulated scenario 8

Figure 5.8: Distributed decision making process of the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ for the simulated scenarios 7 and 8. Subfigures (a)-(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (green line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.

6

Performance Analysis

In this chapter, we analyze the performance of the three diagnostic schemes with respect to sensor fault detectability and the multiple sensor fault isolability. There are several techniques to perform fault detectability and isolability analysis, which vary with the model that describes the system. In Chapter 4 of [Ding [2008]], fault detectability is characterized as a structural property that should be analyzed independent of the system inputs and uncertainties. In this respect, the author provides algebraic conditions for characterizing the sets of detectable faults, taking into account the input and output matrix of the system as well as the fault distribution matrices. In [Düştügör et al. [2006]], the detectable faults are analyzed in a structural framework by using graph-based tools (e.g. causal matching). The analysis is realized in three levels, according to the prior knowledge about the faults, i.e. type of fault, the way that a fault affects the system, and the model of the fault. Isolability conditions are obtained by using a fault signature matrix and isolability matrix. In [Frisk et al. [2012]], the authors derived diagnosability (detectability and isolability) properties taking into account the structural model of the system and three different types of causalities, i.e. derivative, integral, and mixed causal-

ities. In Chapter 4 of [Ding [2008]], and in [Düştégör et al. [2006]] the detectability and isolability conditions are deduced taking into account the structure of the model and its properties, without using any diagnosis system. On the contrary in [Travé-Massuyès et al. [2006]], diagnosability properties are investigated with respect to component-supported analytical redundancy relations. In [Chen and Patton [1999]], the fault detectability conditions in the frequency domain are obtained based on a residual generator designed following a parity space approach, and are distinguished to ‘weak’ and ‘strong’ detectability conditions.

When both uncertainties and inputs are taken into consideration along with a model, the fault detectability and isolability conditions are studied under different assumptions for the uncertainties and inputs. In Chapter 12 of [Ding [2008]], the set of detectable faults is implicitly characterized (with no guarantees) in relation to the designed residual generator and the corresponding threshold in a norm based framework. Then the author proposes a methodology in order to maximize the fault detectability given a false alarm rate. In [Staroswiecki and Comtet-Varga [2001]], detectable and non-detectable faults are defined in an implicit way assuming the stochasticity of the residuals, without providing a guaranteed condition for the detectability. In [Blanke et al. [2016]], weak and strong fault detectability and isolability conditions in the frequency domain are derived by following a parity space approach in which the uncertainties are decoupled. Particularly the fault detectability conditions combine fault sensitivity and disturbance decoupling conditions. The fault isolability conditions are deduced similarly to the detectability conditions after designing several residuals that each of them is sensitive to a subset of faults, and decoupling the effects of both disturbances and faults that does not belong to the specific subset of faults.

In [Polycarpou and Trunov [2000]], guaranteed detectable incipient faults are characterized with respect to worst case assumptions of nonlinear modeling uncertainty (i.e. the modeling uncertainty is assumed to reach its bound), and design parameters of the nonlinear approximation-based diagnostic system. In this work, the authors derived conditions that describe: (i) the detectable effects of faults in re-

lation to the design parameters, the fault evolution rate and the bound on the modeling uncertainty, and (ii) the detectable faults in relation to the bound of the modeling uncertainty. In [Reppa et al. [2015a]], time-varying certain conditions are obtained for characterizing weakly detectable faults, while strongly detectable faults are described by time-invariant conditions. In both cases, the conditions are deduced under worst case assumption of modeling uncertainties, and in relation to the design parameters of the nonlinear observer-based diagnostic system.

In this tutorial, we perform fault detectability and isolability analysis, taking into account nonlinear models of the system, the nonlinear diagnostic system (centralized, decentralized and distributed) and worst case assumptions for the modeling uncertainties. Particularly, we analyze the performance of the proposed multiple sensor fault detection and isolation methodology with respect to the sensor fault detectability of the modules $\mathcal{M}^{(I,q)}$, $I \in \{1, \dots, M\}$, $q \in \{1, \dots, N_I\}$, and multiple sensor fault isolability of the monitoring agent $\mathcal{M}^{(I)}$. For a distributed diagnosis scheme, we investigate the fault detectability in relation to the local and propagated sensor fault effects on the decisions of $\mathcal{M}^{(I,q)}$.

6.1 Sensor Fault Detectability

As a result of the sensor fault detection decision logic presented in Section 4.3, the module $\mathcal{M}^{(I,q)}$ does not raise false alarms as long as the the sensor faults to which $\mathcal{M}^{(I,q)}$ is structurally sensitive do not occur, or equivalently, the presence of sensor faults to which $\mathcal{M}^{(I,q)}$ is structurally sensitive according to Section 4.3 is guaranteed, when $\mathcal{M}^{(I,q)}$ detects them. In addition, certain conditions can be derived, under which the detection of sensor faults in $\mathcal{S}^{(I,q)}$ is guaranteed under worst case assumptions of bounded modeling uncertainties and measurement noise, i.e. assuming that the modeling uncertainties and measurement noise have reached their limits. In practice, the modeling uncertainties and measurement noise may not reach the limits shown in Assumptions 1-4. Thus, the sensor fault detectability conditions are valid for any modeling uncertainty and measurement noise satisfying Assumptions 1-4, as shown next.

Definition 6.1. The effects of sensor faults, to which the module $\mathcal{M}^{(I,q)}$ is structurally sensitive, on the j -th residual and adaptive threshold, denoted by $\varepsilon_{y_{Fj}}^{(I,q)}$ and $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}$, $j \in \mathcal{J}^{(I,q)}$, respectively, for $t \geq T_0^{(I,q)}$, where $T_0^{(I,q)}$ is the first time instant of sensor fault occurrence to which $\mathcal{M}^{(I,q)}$ is structurally sensitive, can be determined as

$$\varepsilon_{y_{Fj}}^{(I,q)} = \varepsilon_{y_j}^{(I,q)} - \varepsilon_{y_{Hj}}^{(I,q)}, \quad (6.1)$$

$$\bar{\varepsilon}_{y_{Fj}}^{(I,q)} = \bar{\varepsilon}_{y_j}^{(I,q)} - \bar{\varepsilon}_{y_{Hj}}^{(I,q)}, \quad (6.2)$$

where $\varepsilon_{y_{Hj}}^{(I,q)}$, $\bar{\varepsilon}_{y_{Hj}}^{(I,q)}$ are the j -th residual and adaptive threshold under healthy conditions, as defined in Sections 4.1 and 4.2, respectively.

It is noted that the sensor fault effects $\varepsilon_{y_{Fj}}^{(I,q)}$ and $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}$ vary with respect to the sensor fault diagnosis architecture (see Sections 4.1 and 4.2). The conditions that characterize the minimum effects of sensor faults to which the module $\mathcal{M}^{(I,q)}$ is structurally sensitive are given in the following Lemma.

Lemma 6.1. The occurrence of sensor faults to which the module $\mathcal{M}^{(I,q)}$ is structurally sensitive according to Lemmas 4.4 and 4.5 is guaranteed to be detected under worst-case conditions, if there exists a time instant $t \geq T_0^{(I,q)}$ such that, for at least one $j \in \mathcal{J}^{(q)}$,

$$\left| \varepsilon_{y_{Fj}}^{(I,q)}(t) \right| - \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) > 2\bar{\varepsilon}_{y_{Hj}}^{(I,q)}(t), \quad (6.3)$$

where $\varepsilon_{y_{Fj}}^{(I,q)}$ and $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}$ are defined through (6.1) and (6.2).

Proof. Taking into account (6.1), for any time instant $t \geq T_0^{(I,q)}$, we have

$$\left| \varepsilon_{y_j}^{(I,q)}(t) \right| = \left| \varepsilon_{y_{Fj}}^{(I,q)}(t) + \varepsilon_{y_{Hj}}^{(I,q)}(t) \right| \geq \left| \varepsilon_{y_{Fj}}^{(I,q)}(t) \right| - \left| \varepsilon_{y_{Hj}}^{(I,q)}(t) \right|. \quad (6.4)$$

Given (4.76), we obtain

$$\left| \varepsilon_{y_j}^{(I,q)}(t) \right| \geq \left| \varepsilon_{y_{Fj}}^{(I,q)}(t) \right| - \bar{\varepsilon}_{y_{Hj}}^{(I,q)}(t). \quad (6.5)$$

By combining (6.3) and (6.5), it yields

$$\left| \varepsilon_{y_j}^{(I,q)}(t) \right| > 2\bar{\varepsilon}_{y_{Hj}}^{(I,q)}(t) + \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) - \bar{\varepsilon}_{y_{Hj}}^{(I,q)}(t) = \bar{\varepsilon}_{y_{Hj}}^{(I,q)}(t) + \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t). \quad (6.6)$$

Given (6.2), inequality (6.6) becomes $\left| \varepsilon_{y_j}^{(I,q)}(t) \right| > \bar{\varepsilon}_{y_j}^{(I,q)}(t)$, which implies the detection of sensor faults in $\mathcal{S}^{(I,q)}$ at the time instant t . \square

The main goal of the designer of a sensor fault diagnosis method, irrespective of the adopted architecture (see Section 3), is to formulate adaptive thresholds that are low as much as possible such that no false alarms are raised under healthy conditions on one hand, while on the other hand, there should be ideally zero number of missed fault detections. Based on condition (6.3), we may infer that the lower the adaptive threshold under healthy conditions the more detectable the sensor faults effects on the residuals and the adaptive thresholds.

The design of the adaptive thresholds depends on the adopted architecture (see Section 4.2). It is worth to discuss how the different designs of adaptive thresholds can affect the sensor fault detectability of the module $\mathcal{M}^{(I,q)}$. For this reason, the adaptive thresholds under healthy conditions for a distributed, decentralized and centralized architecture are presented again below.

Distributed adaptive thresholds under healthy conditions:

$$\begin{aligned} \bar{\varepsilon}_{y_{Hj}}^{(I,q)}(t) = & Y_j^{(I,q)}(t) + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I Z_H^{(I,q)}(\tau) \right. \\ & \left. + \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) \right) d\tau, \end{aligned} \quad (6.7)$$

$$\begin{aligned} Y_j^{(I,q)}(t) = & \alpha_j^{(I,q)} \bar{x}^{(I,q)} e^{-\zeta_j^{(I,q)} t} + \frac{\alpha_{d_j}^{(I,q)} \bar{d}^{(I,q)}}{\zeta_{d_j}^{(I,q)}} \left(1 - e^{-\zeta_{d_j}^{(I,q)} t} \right) + \bar{d}_j^{(I)} \\ & + \frac{\alpha_j^{(I,q)} (\lambda_{h_I} + \lambda_{\eta_I}) (\bar{\zeta}^{(I)} + \bar{d}_z^{(I)})}{\zeta_j^{(I,q)}} \left(1 - e^{-\zeta_j^{(I,q)} t} \right), \end{aligned} \quad (6.8)$$

$$Z_H^{(I,q)}(t) = E_H^{(I,q)}(t) + \rho^{(I,q)} \Lambda_I \int_0^t E_H^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau \quad (6.9)$$

$$\begin{aligned} E_H^{(I,q)}(t) = & \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\ & + \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) d\tau \\ & + \frac{(\lambda_{h_I} + \lambda_{\eta_I}) (\bar{d}_z^{(I)} + \bar{\zeta}^{(I)})}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right). \end{aligned} \quad (6.10)$$

Decentralized adaptive thresholds under healthy conditions:

$$\begin{aligned} \bar{\varepsilon}_{y_{Hj}}^{(I,q)}(t) = & Y_j^{(I,q)}(t) + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I Z_H^{(I,q)}(\tau) \right. \\ & \left. + \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) \right) d\tau \end{aligned} \quad (6.11)$$

$$\begin{aligned} Y_j^{(I,q)}(t) = & \alpha_j^{(I,q)} \bar{x}^{(I)} e^{-\zeta_j^{(I,q)} t} + \frac{\alpha_{d_j}^{(I,q)} \bar{d}^{(I,q)}}{\zeta_{d_j}^{(I,q)}} \left(1 - e^{-\zeta_{d_j}^{(I,q)} t} \right) \\ & + \frac{\alpha_j^{(I,q)} (\lambda_{h_I} + \lambda_{\eta_I}) \bar{z}^{(I)}}{\zeta_j^{(I,q)}} \left(1 - e^{-\zeta_j^{(I,q)} t} \right) + \bar{d}_j^{(I)}. \end{aligned} \quad (6.12)$$

where $Z_H^{(I,q)}$ is defined in (6.9) with

$$\begin{aligned} E_H^{(I,q)}(t) = & \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\ & + \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) d\tau \\ & + \frac{(\lambda_{h_I} + \lambda_{\eta_I}) \bar{z}^{(I)}}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right). \end{aligned} \quad (6.13)$$

Centralized adaptive thresholds under healthy conditions:

$$\begin{aligned} \bar{\varepsilon}_{y_{Hj}}^{(q)}(t) = & \int_0^t \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-\tau)} \left(\Lambda Z_H^{(q)}(\tau) + \bar{\eta}(\hat{x}_H^{(q)}(\tau), u^{(I)}(\tau), \tau) \right) d\tau \\ & + Y_j^{(q)}(t) \end{aligned} \quad (6.14)$$

where

$$Y_j^{(q)}(t) = \alpha_j^{(q)} \bar{x}^{(q)} e^{-\zeta_j^{(q)} t} + \frac{\alpha_{d_j}^{(q)} \bar{d}^{(q)}}{\zeta_{d_j}^{(q)}} \left(1 - e^{-\zeta_{d_j}^{(q)} t} \right) + \bar{d}_j \quad (6.15)$$

$$Z_H^{(q)}(t) = E_H^{(q)}(t) + \rho^{(q)} \Lambda \int_0^t E_H^{(q)}(\tau) e^{-\nu^{(q)}(t-\tau)} d\tau, \quad (6.16)$$

$$\begin{aligned} E_H^{(q)}(t) = & \Phi^{(q)}(t) \bar{x}^{(q)} + \frac{\rho_d^{(q)} \bar{d}^{(q)}}{\xi_d^{(q)}} \left(1 - e^{-\xi_d^{(q)} t} \right) \\ & + \int_0^t \Phi^{(q)}(t-\tau) \bar{\eta}(\hat{x}_H^{(q)}(\tau), u(\tau), \tau) d\tau \end{aligned} \quad (6.17)$$

Comparing the adaptive thresholds following a distributed and decentralized approach, we may infer that the distributed adaptive threshold can be less than a decentralized adaptive threshold, since the bound $\bar{z}^{(I)}$ in (6.12) and (6.13) can be more conservative than the bound $\bar{\zeta}^{(I)} + \bar{d}_z^{(I)}$ in (6.8) and (6.10). This is also valid for the bounds on the modeling uncertainty $\bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_{z_H}^{(I)}, t)$ and $\bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, z_r^{(I)}(t), t)$, mainly because $\hat{x}_H^{(I,q)}$ computed in a distributed scheme is a better estimation of $\hat{x}^{(I)}$ compared to $\hat{x}_H^{(I,q)}$ computed in a decentralized scheme (the bound on the estimation error in a distributed architecture defined through (6.9) and (6.10) is lower than the bound on the estimation error in a decentralized architecture defined through (6.9) and (6.13)). The reference signals $z_r^{(I)}$ may be comparable to the signals $\zeta^{(I)}$ and $y_z^{(I)}$, depending on the performance of the controller $\mathcal{K}^{(I)}$ (see Fig. 2.1). If the interconnection dynamics $h^{(I)}$ and the modeling uncertainty $\eta^{(I)}$ are functions of measurable interconnection states only (i.e. $\zeta^{(I)} = 0, \bar{\zeta}^{(I)} = 0$), the distributive adaptive threshold can be much lower than a decentralized adaptive threshold.

The centralized adaptive threshold can be more conservative than the non-centralized adaptive thresholds, mainly due to the scale of the monolithic system that affects the determination of the Lipschitz constants λ_γ , λ_h and λ_η , as well as the bound on the modeling uncertainty $\bar{\eta}(\hat{x}_H^{(q)}, u, t)$. In addition, the number of sensors in the set $\mathcal{S}^{(q)}$ may be more than in the set $\mathcal{S}^{(I,q)}$, implying a larger bound $\bar{d}^{(q)}$ on the measurement noise $d^{(q)}$ than the bound $\bar{d}^{(I,q)}$ on the measurement noise $d^{(I,q)}$.

The adaptive nature of the thresholds proposed for all architectures can contribute to the reduction of the conservativeness in the decision making process of the monitoring modules $\mathcal{M}^{(I,q)}$ compared to fixed thresholds, i.e. a threshold $\bar{\sigma}^{(I,q)}$ that bounds the residual under healthy conditions such that $|\varepsilon_{y_{Hj}}^{(I,q)}| \leq \bar{\sigma}^{(I,q)}$, prohibiting the occurrence of false alarms. A fixed threshold can be determined by taking into account Assumption 1 and 3 and using a fixed bound on the modeling uncertainty, denoted by $\bar{\eta}_o^{(I)}$, such that

$$|\eta^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t)| \leq \bar{\eta}^{(I)}(x^{(I)}, u^{(I)}, z^{(I)}, t) \leq \bar{\eta}_o^{(I)} \quad (6.18)$$

Using $\bar{\eta}_o^{(I)}$, the threshold in a distributed approach (correspondingly for the decentralized and centralized approach), can be described by

$$\bar{\sigma}^{(I,q)} = \sup_{t \geq t_0} \left\{ Y_j^{(I,q)} + \frac{\alpha_j^{(I,q)} (\Lambda_I Z_H^{(I,q)} + \bar{\eta}_o^{(I)})}{\zeta_j^{(I,q)}} \left(1 - e^{-\zeta_j^{(I,q)} t} \right) \right\} \quad (6.19)$$

$$\begin{aligned} Y_j^{(I,q)} = \sup_{t \geq t_0} \left\{ \alpha_j^{(I,q)} \bar{x}^{(I,q)} e^{-\zeta_j^{(I,q)} t} + \frac{\alpha_{d_j}^{(I,q)} \bar{d}^{(I,q)}}{\zeta_{d_j}^{(I,q)}} \left(1 - e^{-\zeta_{d_j}^{(I,q)} t} \right) + \bar{d}_j^{(I)} \right. \\ \left. + \frac{\alpha_j^{(I,q)} (\lambda_{h_I} + \lambda_{\eta_I}) (\bar{\zeta}^{(I)} + \bar{d}_z^{(I)})}{\zeta_j^{(I,q)}} \left(1 - e^{-\zeta_j^{(I,q)} t} \right) \right\}, \quad (6.20) \end{aligned}$$

$$Z_H^{(I,q)} = \sup_{t \geq t_0} \left\{ E_H^{(I,q)} + \frac{\rho^{(I,q)} \Lambda_I E_H^{(I,q)}}{\nu^{(I,q)}} \left(1 - e^{-\nu^{(I,q)} t} \right) \right\}, \quad (6.21)$$

$$\begin{aligned} E_H^{(I,q)} = \sup_{t \geq t_0} \left\{ \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \right. \\ \left. + \frac{(\lambda_{h_I} + \lambda_{\eta_I}) (\bar{d}_z^{(I)} + \bar{\zeta}^{(I)}) + \bar{\eta}_o^{(I)}}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right) \right\}. \quad (6.22) \end{aligned}$$

Based on the computation of the fixed threshold such that structured robustness is guaranteed, it yields $\bar{\varepsilon}_{y_{Hj}}^{(I,q)} < \bar{\sigma}^{(I,q)}$. By using a fixed threshold, the sensor fault detectability conditions are represented by the following inequality

$$\left| \varepsilon_{y_{Fj}}^{(I,q)}(t) \right| - \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) > 2\bar{\sigma}^{(I,q)}(t) > 2\bar{\varepsilon}_{y_{Hj}}^{(I,q)}. \quad (6.23)$$

Thus, using a fixed threshold can make the sensor fault detection difficult, leading to delayed detections, or even missed detections.

6.1.1 Distributed Analysis

Given that $T_0^{(I,q)}$ is the first time instant of that a fault occurs in $\mathcal{S}^{(I,q)} \cup \mathcal{S}_z^{(I)}$, the effects of sensor faults $f^{(I,q)}$ and $f_z^{(I)}$ on the j -th residual and adaptive threshold are described in Proposition 6.1. The proof

of the Proposition 6.1 is based on the Definition definitionEffects. Particularly, the effects $\varepsilon_{y_{Fj}}^{(I,q)}(t)$ for $t \geq T_0^{(I,q)}$ defined in (6.1) can be analytically described, taking into account that $\varepsilon_{y_j}^{(I,q)} = y_{Hj}^{(I)} + f_j^{(I)} - C_j^{(I)} \hat{x}^{(I,q)}$ and $\varepsilon_{y_{Hj}}^{(I,q)} = y_{Hj}^{(I)} - C_j^{(I)} \hat{x}_H^{(I,q)}$, $j \in \mathcal{J}^{(I,q)}$, where $y_{Hj}^{(I)}$ is the j -th component of $y_H^{(I)}$ defined in (4.6); i.e.,

$$\varepsilon_{y_{Fj}}^{(I,q)} = -C_j^{(I)} \left(\hat{x}^{(I,q)} - \hat{x}_H^{(I,q)} \right) + f_j^{(I)}, \quad (6.24)$$

where $\hat{x}^{(I,q)}$ is the estimation model of $\mathcal{M}^{(I,q)}$ for $t \geq T_0^{(I,q)}$ that satisfies

$$\begin{aligned} \dot{\hat{x}}^{(I,q)} &= A_L^{(I,q)} \hat{x}^{(I,q)} + L^{(I,q)} \left(y_H^{(I,q)} + f^{(I,q)} \right) + \gamma^{(I)}(\hat{x}^{(I,q)}, u^{(I)}) \\ &\quad + h^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, \zeta_r^{(I)}(t), y_{z_H}^{(I)} + f_z^{(I)}), \end{aligned} \quad (6.25)$$

where $y_{z_H}^{(I)}$ is defined in (4.7). Similarly the effects of sensor faults on the adaptive threshold $\varepsilon_{y_{Fj}}^{(I,q)}$ defined in (6.2) can be described analytically, taking into account that after the occurrence of sensor faults in $\mathcal{S}^{(I,q)}$ the adaptive threshold $\bar{\varepsilon}_{y_j}^{(I,q)}$ is defined in (4.79), while $\text{bar}\varepsilon_{y_{Hj}}^{(I,q)}$ is defined in (4.77).

Proposition 6.1. In a distributed architecture, the effects of sensor faults $f^{(I,q)}$ and $f_z^{(I)}$ on the j -th residual and adaptive threshold, i.e. $\varepsilon_{y_{Fj}}^{(I,q)}$ and $\bar{\varepsilon}_{y_j}^{(I,q)}$, $j \in \mathcal{J}^{(I,q)}$, respectively, are defined as

$$\begin{aligned} \varepsilon_{y_{Fj}}^{(I,q)} &= C_j^{(I)} \int_{T_0^{(I,q)}}^t e^{A_L^{(I,q)}(t-\tau)} \left(\gamma^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau)) \right. \\ &\quad - \gamma^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau)) + h^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau)) \\ &\quad - h^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau) + f_z^{(I)}(\tau)) \\ &\quad \left. - L^{(I,q)} f^{(I,q)}(\tau) \right) d\tau + f_j^{(I)} \end{aligned} \quad (6.26)$$

and

$$\begin{aligned} \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) &= \int_{T_0^{(I,q)}}^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I \delta Z^{(I,q)}(t) \right. \\ &\quad + \left(\bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau) + f_z^{(I)}(\tau), \tau) \right. \\ &\quad \left. \left. - \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) \right) \right) d\tau \end{aligned} \quad (6.27)$$

$$\delta Z^{(I,q)}(t) = \delta E^{(I,q)}(t) + \rho^{(I,q)} \Lambda_I \int_{T_0^{(I,q)}}^t \delta E^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau \quad (6.28)$$

$$\begin{aligned} \delta E^{(I,q)}(t) = \int_0^t \Phi^{(I,q)}(t-\tau) & \left(\bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau) + f_z^{(I)}(\tau)) \right. \\ & \left. - \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau)) \right) d\tau, \end{aligned} \quad (6.29)$$

for $t \geq T_0^{(I,q)}$ and for at least one $j \in \mathcal{J}^{(I,q)}$, where $\Phi^{(I,q)}$ is defined in (4.13).

Proof. Based on (4.4) and (6.25),

$$\begin{aligned} \hat{x}^{(I,q)}(t) - \hat{x}_H^{(I,q)}(t) = \int_{T_0^{(I,q)}}^t e^{A_L^{(I,q)}(t-\tau)} & \left(\gamma^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau)) \right. \\ & - \gamma^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau)) \\ & + h^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau) + f_z^{(I)}(\tau)) \\ & - h^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau)) \\ & \left. + L^{(I,q)} f^{(I,q)}(\tau) \right) \end{aligned} \quad (6.30)$$

Hence, the effects of sensor faults $f^{(I,q)}$ and $f_z^{(I)}$ on the j -th residual satisfies (6.26).

By subtracting (4.77) from (4.79), the effects $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t)$ can be expressed as

$$\begin{aligned} \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) = Y_j^{(I,q)}(t) + \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} & \left(\Lambda_I Z^{(I,q)}(\tau) \right. \\ & + \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_z^{(I)}(\tau), \tau) \Big) d\tau \\ - Y_j^{(I,q)}(t) - \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} & \left(\Lambda_I Z_H^{(I,q)}(\tau) \right. \\ & \left. + \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) \right) d\tau \end{aligned} \quad (6.31)$$

Let us define the integral $\mathcal{I}_1(0, t)$ with limits 0, t as

$$\begin{aligned} \mathcal{I}_1(0, t) = \int_0^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} & \left(\Lambda_I Z^{(I,q)}(\tau) \right. \\ & \left. + \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_z^{(I)}(\tau), \tau) \right) d\tau \end{aligned} \quad (6.32)$$

Then,

$$\mathcal{I}_1(0, t) = \mathcal{I}_1(0, T_0^{(I,q)}) + \mathcal{I}_1(T_0^{(I,q)}, t) \quad (6.33)$$

where

$$\begin{aligned} \mathcal{I}_1(0, T_0^{(I,q)}) = & \int_0^{T_0^{(I,q)}} \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I Z_H^{(I,q)}(\tau) \right. \\ & \left. + \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) \right) d\tau \end{aligned} \quad (6.34)$$

since, before the first time of fault occurrence $T_0^{(I,q)}$, $Z^{(I,q)} = Z_H^{(I,q)}$, $\hat{x}^{(I,q)} = \hat{x}_H^{(I,q)}$ and $y_z^{(I)} = y_{z_H}^{(I)}$. Combining (6.31), (6.34) results in

$$\begin{aligned} \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) = & \int_{T_0^{(I,q)}}^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I \left(Z^{(I,q)}(\tau) - Z_H^{(I,q)}(\tau) \right) \right. \\ & + \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_z^{(I)}(\tau), \tau) \\ & \left. - \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) \right) d\tau \end{aligned} \quad (6.35)$$

Let us define

$$\delta Z^{(I,q)} = Z^{(I,q)} - Z_H^{(I,q)} \quad (6.36)$$

$$\delta \bar{\eta}_F^{(I,q)} = \bar{\eta}^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_z^{(I)}, t) - \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, \zeta_r^{(I)}, y_{z_H}^{(I)}, t) \quad (6.37)$$

then (6.35) can be re-written as

$$\bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) = \int_{T_0^{(I,q)}}^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I \delta Z^{(I,q)}(\tau) + \delta \bar{\eta}_F^{(I,q)}(\tau) \right) d\tau \quad (6.38)$$

where

$$\begin{aligned} \delta Z^{(I,q)}(t) &= \delta E^{(I,q)}(t) + \rho^{(I,q)} \Lambda_I \int_0^t \delta E^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau \\ \delta E^{(I,q)}(t) &= E^{(I,q)}(t) - E_H^{(I,q)}(t) \end{aligned} \quad (6.39)$$

where $E^{(I,q)}$ and $E_H^{(I,q)}(t)$ are defined in (4.81) and (6.10). Thus, we

have

$$\begin{aligned}
\delta E^{(I,q)}(t) &= \Phi^{(I,q)}(t) \bar{x}^{(I,q)} + \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\
&\quad + \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_z^{(I)}(\tau), \tau) d\tau \\
&\quad + \frac{(\lambda_{h_I} + \lambda_{\eta_I})(\bar{d}_z^{(I)} + \bar{\zeta}^{(I)})}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right) \\
&\quad - \Phi^{(I,q)}(t) \bar{x}^{(I,q)} - \frac{\rho_d^{(I,q)} \bar{d}^{(I,q)}}{\xi_d^{(I,q)}} \left(1 - e^{-\xi_d^{(I,q)} t} \right) \\
&\quad - \int_0^t \Phi^{(I,q)}(t-\tau) \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), \zeta_r^{(I)}(\tau), y_{z_H}^{(I)}(\tau), \tau) d\tau \\
&\quad - \frac{(\lambda_{h_I} + \lambda_{\eta_I})(\bar{d}_z^{(I)} + \bar{\zeta}^{(I)})}{\xi^{(I,q)}} \left(\rho^{(I,q)} - \Phi^{(I,q)}(t) \right). \tag{6.41}
\end{aligned}$$

implying that

$$\delta E^{(I,q)}(t) = \int_0^t \Phi^{(I,q)}(t-\tau) \delta \bar{\eta}_F^{(I,q)}(\tau) d\tau \tag{6.42}$$

Since, before the first time of fault occurrence $T_0^{(I,q)}$, $\hat{x}^{(I,q)} = \hat{x}_H^{(I,q)}$ and $f_z^{(I)} = 0$, we obtain

$$\delta E^{(I,q)}(t) = \int_{T_0^{(I,q)}}^t \Phi^{(I,q)}(t-\tau) \delta \bar{\eta}_F^{(I,q)}(\tau) d\tau. \tag{6.43}$$

Taking into account that $\delta E^{(I,q)}(t) = 0$ before the first time of fault occurrence $T_0^{(I,q)}$, (6.39) can be re-written as

$$\begin{aligned}
\delta Z^{(I,q)}(t) &= \delta E^{(I,q)}(t) + \rho^{(I,q)} \Lambda_I \int_0^{T_0^{(I,q)}} \delta E^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau \\
&\quad + \rho^{(I,q)} \Lambda_I \int_{T_0^{(I,q)}}^t \delta E^{(I,q)}(\tau) e^{-\nu^{(I,q)}(t-\tau)} d\tau
\end{aligned}$$

leading to (6.28). By combining (6.37) and (6.38) we can describe the effects of sensor faults $f^{(I,q)}$ and $f_z^{(I)}$ on the j -th adaptive threshold as in (6.27)-(6.29). \square

The condition (6.3) along with (6.7)-(6.10) and (6.26)-(6.29) and can be taken into account in the design of the modules $\mathcal{M}^{(I,q)}$ for calibrating the design parameters, since it provides a relationship between sensor faults $f^{(I,q)}$ and $f_z^{(I)}$ and the selected design parameters used for the implementation of the nonlinear observer $\mathcal{N}^{(I,q)}$ (e.g. $L^{(I,q)}$) and adaptive thresholds (e.g. $\rho^{(I,q)}, \xi^{(I,q)}, \alpha_j^{(I,q)}, \zeta_j^{(I,q)}, j \in \mathcal{J}^{(I,q)}$). Based on Proposition 6.1 and assuming that the sensor faults are uniformly bounded (see Assumption 1 in Section 2.1.1), we can analyze the different ways that local and propagated sensor faults may affect every subsystem.

Bounded distributed sensor fault effects: Let us define the following functions:

$$G_0^{(I,q)}(t) = \rho^{(I,q)} e^{-\xi^{(I,q)} t} \quad (6.44)$$

$$G_{0_j}^{(I,q)}(t) = \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)} t} \quad (6.45)$$

$$G_1^{(I,q)}(t) = \rho^{(I,q)} (\lambda_{\gamma_I} + \lambda_{h_I}) e^{-\mu^{(I,q)} t} \quad (6.46)$$

$$G_2^{(I,q)}(t) = \rho^{(I,q)} \Lambda_I e^{-\nu^{(I,q)} t} \quad (6.47)$$

$$G_d^{(I,q)}(t) = \rho_d^{(I,q)} e^{-\xi_d^{(I,q)} t} \quad (6.48)$$

$$G_{d_j}^{(I,q)}(t) = \alpha_{d_j}^{(I,q)} e^{-\zeta_{d_j}^{(I,q)} t} \quad (6.49)$$

with

$$\mu^{(I,q)} = \xi^{(I,q)} - \rho^{(I,q)} (\lambda_{\gamma_I} + \lambda_{h_I}), \quad (6.50)$$

where $\rho^{(I,q)}, \xi^{(I,q)}, \rho_d^{(I,q)}, \xi_d^{(I,q)}, \alpha_j^{(I,q)}, \zeta_j^{(I,q)}, \alpha_{d_j}^{(I,q)}, \zeta_{d_j}^{(I,q)}$ and $\nu^{(I,q)}$ are parameters used for the design of the adaptive thresholds, with $\rho^{(I,q)}, \xi^{(I,q)}$ satisfying (4.11) and conditions in Theorem 4.1, $\rho_d^{(I,q)}, \xi_d^{(I,q)}, \alpha_j^{(I,q)}, \zeta_j^{(I,q)}, \alpha_{d_j}^{(I,q)}, \zeta_{d_j}^{(I,q)}$ satisfying (4.12), (4.72) and (4.73), and $\nu^{(I,q)}$ defined in (4.14).

The distributed sensor fault effects are analyzed in the Lemma 6.2. The proof of the Lemma is taking into account that $|f_z^{(I)}| < \infty$ and $|f^{(I,q)}| < \infty$. Then, in order to compute a bound on $|\varepsilon_{y_{Fj}}^{(I,q)}(t)|$, we have to compute a bound on $|\hat{x}^{(I,q)} - \hat{x}_H^{(I,q)}|$, where $\hat{x}^{(I,q)}$ is described

by (6.25) and $\hat{x}_H^{(I,q)}$ is described by (4.4). The bound on $|\bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t)|$ is computed taking into account (6.27)-(6.29) by bounding each term of $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t)$. The mathematical manipulations of the proof are based on Assumptions 1-4, and applying the Bellman-Gronwall Lemma [Ioannou and Sun [1995]] and linear filtering techniques as presented in (4.82) and (4.83).

Lemma 6.2. Based on Assumptions 1-4 and Proposition 6.1, the effects of the local sensor faults $f^{(I,q)}$ and the effects of the propagated sensor faults $f_z^{(I)}$ on the j -th residual $\varepsilon_{y_{Fj}}^{(I,q)}(t)$ and adaptive threshold $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t)$, $j \in \mathcal{J}^{(I,q)}$ are uniformly bounded and satisfy

$$|\varepsilon_{y_{Fj}}^{(I,q)}(t)| \leq \bar{y}_{Fj}^{(I,q)}(t) \quad (6.51)$$

$$|\bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t)| \leq \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) \quad (6.52)$$

where

$$\begin{aligned} \bar{y}_{Fj}^{(I,q)}(t) = & G_{0j}^{(I,q)}(s) \left((\lambda_{\gamma_I} + \lambda_{h_I}) \left(G_1^{(I,q)}(s) + 1 \right) \right) \\ & \times \left(G_0^{(I,q)}(s) \left[\lambda_{h_I} |f_z^{(I)}(t)| \right] + G_d^{(I,q)}(s) \left[|f^{(I,q)}(t)| \right] \right) \\ & + G_{0j}^{(I,q)}(s) \left[\lambda_{h_I} |f_z^{(I)}(t)| \right] + G_{dj}^{(I,q)}(s) \left[|f^{(I,q)}(t)| \right] + |f_j^{(I)}(t)| \end{aligned} \quad (6.53)$$

and

$$\begin{aligned} \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) = & G_{0j}^{(I,q)}(s) \left((G_2^{(I,q)}(s) + 1) G_0^{(I,q)}(s) + 1 \right) \\ & \times \left(\lambda_{\eta_I} \left(G_1^{(I,q)}(s) + 1 \right) \left(G_0^{(I,q)}(s) \left[\lambda_{h_I} |f_z^{(I)}(t)| \right] \right. \right. \\ & \left. \left. + G_d^{(I,q)}(s) \left[|f^{(I,q)}(t)| \right] \right) + \lambda_{\eta_I} |f_z^{(I)}(t)| \right) \end{aligned} \quad (6.54)$$

where s denotes the Laplace operator.

Proof. Let us define $\varepsilon_{x_F}^{(I,q)}(t) = \hat{x}^{(I,q)}(t) - \hat{x}_H^{(I,q)}(t)$; then, according to

(6.30), we have

$$\begin{aligned} \left| \varepsilon_{x_F}^{(I,q)}(t) \right| &\leq \int_{T_0^{(I,q)}}^t \left| e^{A_L^{(I,q)}(t-\tau)} \right| \epsilon_1^{(I,q)}(\tau) d\tau \\ &\quad + \int_{T_0^{(I,q)}}^t \left| e^{A_L^{(I,q)}(t-\tau)} L^{(I,q)} \right| \left| f^{(I,q)}(\tau) \right| d\tau, \end{aligned} \quad (6.55)$$

$$\begin{aligned} \epsilon_1^{(I,q)} &= \left| h^{(I)}(\hat{x}^{(I,q)}, u^{(I)}, y_{z_H}^{(I)} + f_z^{(I)}) - h^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}, y_{z_H}^{(I)}) \right| \\ &\quad + \left| \gamma^{(I)}(\hat{x}^{(I,q)}, u^{(I)}) - \gamma^{(I)}(\hat{x}_H^{(I,q)}, u^{(I)}) \right| \end{aligned} \quad (6.56)$$

Taking into account conditions (4.11) and (4.12), and (6.44), (6.48), we derive

$$\begin{aligned} \left| \varepsilon_{x_F}^{(I,q)}(t) \right| &\leq \int_{T_0^{(I,q)}}^t G_0^{(I,q)}(t-\tau) \epsilon_1^{(I,q)}(\tau) d\tau \\ &\quad + \int_{T_0^{(I,q)}}^t G_d^{(I,q)}(t-\tau) \left| f^{(I,q)}(\tau) \right| d\tau \end{aligned} \quad (6.57)$$

Based on Assumptions 1-4 in Section 2.1.1, we obtain

$$\epsilon_1^{(I,q)}(t) \leq \epsilon_2^{(I,q)}(t), \quad (6.58)$$

$$\epsilon_2^{(I,q)}(t) = (\lambda_{\gamma_I} + \lambda_{h_I}) \left| \varepsilon_{x_F}^{(I,q)}(t) \right| + \lambda_{h_I} \left| f_z^{(I)}(t) \right| \quad (6.59)$$

Using (6.58) and (6.59) in (6.57), and applying the Bellman-Gronwall Lemma [Ioannou and Sun [1995]] results in

$$\left| \varepsilon_{x_F}^{(I,q)}(t) \right| \leq \bar{\varepsilon}_{x_F}^{(I,q)}(t) \quad (6.60)$$

$$\bar{\varepsilon}_{x_F}^{(I,q)}(t) = \int_{T_0^{(I,q)}}^t G_1^{(I,q)}(t-\tau) \Upsilon_z^{(I,q)}(\tau) d\tau + \Upsilon_z^{(I,q)}(t), \quad (6.61)$$

where $G_1^{(I,q)}$ is defined in (6.46) and

$$\begin{aligned} \Upsilon_z^{(I,q)}(t) &= \lambda_{h_I} \int_{T_0^{(I,q)}}^t G_0^{(I,q)}(t-\tau) \left| f_z^{(I)}(\tau) \right| d\tau \\ &\quad + \int_{T_0^{(I,q)}}^t G_d^{(I,q)}(t-\tau) \left| f^{(I,q)}(\tau) \right| d\tau \end{aligned} \quad (6.62)$$

Using linear filtering techniques, we have

$$\Upsilon_z^{(I,q)}(t) = G_0^{(I,q)}(s) \left[\lambda_{h_I} \left| f_z^{(I)}(t) \right| \right] + G_d^{(I,q)}(s) \left[\left| f^{(I,q)}(t) \right| \right] \quad (6.63)$$

and

$$\begin{aligned}\bar{\varepsilon}_{x_F}^{(I,q)}(t) &= G_1^{(I,q)}(s) \left[\Upsilon_z^{(I,q)}(t) \right] + \Upsilon_z^{(I,q)}(t) \\ &= (G_1^{(I,q)}(s) + 1) \left[\Upsilon_z^{(I,q)}(t) \right]\end{aligned}\quad (6.64)$$

where the notation $W(s)[z(t)]$ denotes the output of the filter $W(s)$ defined in Laplace domain with $z(t)$ as input, for any signal $z(t)$. By combining (6.63) and (6.64), it yields

$$\begin{aligned}\bar{\varepsilon}_{x_F}^{(I,q)}(t) &= \left(G_1^{(I,q)}(s) + 1 \right) \left(G_0^{(I,q)}(s) \left[\lambda_{h_I} \left| f_z^{(I)}(t) \right| \right] \right. \\ &\quad \left. + G_d^{(I,q)}(s) \left[\left| f^{(I,q)}(t) \right| \right] \right)\end{aligned}\quad (6.65)$$

Based on (6.26), a bound on $\left| \varepsilon_{y_{Fj}}^{(I,q)}(t) \right|$, denoted by $\bar{\varepsilon}_{x_F}^{(I,q)}(t)$, can be computed as follows:

$$\begin{aligned}\left| \varepsilon_{y_{Fj}}^{(I,q)}(t) \right| &\leq \int_{T_0^{(I,q)}}^t \left| C_j^{(I)} e^{A_L^{(I,q)}(t-\tau)} \right| \left| \epsilon_1^{(I,q)}(\tau) d\tau + \left| f_j^{(I)}(t) \right| \right. \\ &\quad \left. + \int_{T_0^{(I,q)}}^t \left| C_j^{(I)} e^{A_L^{(I,q)}(t-\tau)} L^{(I,q)} \right| \left| f^{(I,q)}(\tau) \right| d\tau\end{aligned}\quad (6.66)$$

Taking into account (4.72) and (4.73), as well as (6.45), (6.49) and (6.58), we derive

$$\begin{aligned}\left| \varepsilon_{y_{Fj}}^{(I,q)}(t) \right| &\leq \int_{T_0^{(I,q)}}^t G_{0j}^{(I,q)}(t-\tau) \epsilon_2^{(I,q)}(\tau) d\tau + \left| f_j^{(I)}(t) \right| \\ &\quad + \int_{T_0^{(I,q)}}^t G_{dj}^{(I,q)}(t-\tau) \left| f^{(I,q)}(\tau) \right| d\tau\end{aligned}\quad (6.67)$$

By using (6.59) and (6.60), we obtain (6.51) with

$$\begin{aligned}\bar{y}_{Fj}^{(I,q)}(t) &= \int_{T_0^{(I,q)}}^t G_{0j}^{(I,q)}(t-\tau) \left((\lambda_{\gamma_I} + \lambda_{h_I}) \bar{\varepsilon}_{x_F}^{(I,q)}(\tau) + \lambda_{h_I} \left| f_z^{(I)}(\tau) \right| \right) d\tau \\ &\quad + \left| f_j^{(I)}(t) \right| + \int_{T_0^{(I,q)}}^t G_{dj}^{(I,q)}(t-\tau) \left| f^{(I,q)}(\tau) \right| d\tau\end{aligned}\quad (6.68)$$

Using linear filtering techniques and (6.65), after some mathematical manipulations we have (6.53).

A bound for the sensor fault effects on the j -th adaptive threshold given in (6.27) can be determined as follows. Based on (6.38), (6.45) and (6.48), we have

$$\left| \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) \right| \leq \int_{T_0^{(I,q)}}^t G_{0j}^{(I,q)}(t-\tau) \left(\Lambda_I \left| \delta Z^{(I,q)}(\tau) \right| + \left| \delta \bar{\eta}_F^{(I,q)}(\tau) \right| \right) d\tau, \quad (6.69)$$

while

$$\left| \delta Z^{(I,q)} \right| \leq \left| \delta E^{(I,q)} \right| + \int_{T_0^{(I,q)}}^t G_2^{(I,q)}(t-\tau) \left| \delta E^{(I,q)}(\tau) \right| d\tau \quad (6.70)$$

A bound for $\left| \delta E^{(I,q)} \right|$ can be computed taking into account (4.13) and (6.43); i.e.

$$\left| \delta E^{(I,q)}(t) \right| \leq \int_{T_0^{(I,q)}}^t G_0^{(I,q)}(t-\tau) \left| \delta \bar{\eta}_F^{(I,q)}(\tau) \right| d\tau \quad (6.71)$$

while

$$\begin{aligned} \left| \delta \bar{\eta}_F^{(I,q)}(t) \right| &\leq \lambda_{\eta_I} \left| \varepsilon_{x_F}^{(I,q)}(t) \right| + \lambda_{\eta_I} \left| f_z^{(I)}(t) \right| \\ &\leq \lambda_{\eta_I} \bar{\varepsilon}_{x_F}^{(I,q)}(t) + \lambda_{\eta_I} \left| f_z^{(I)}(t) \right| = \tilde{\eta}_F^{(I,q)}(t) \end{aligned} \quad (6.72)$$

Given (6.65), the bound $\tilde{\eta}_F^{(I,q)}(t)$ can be written explicitly as:

$$\begin{aligned} \tilde{\eta}_F^{(I,q)}(t) &= \lambda_{\eta_I} \left| f_z^{(I)}(t) \right| + \lambda_{\eta_I} \left(G_1^{(I,q)}(s) + 1 \right) \left(G_0^{(I,q)}(s) \left[\lambda_{h_I} \left| f_z^{(I)}(t) \right| \right] \right. \\ &\quad \left. + G_d^{(I,q)}(s) \left[\left| f^{(I,q)}(t) \right| \right] \right) \end{aligned} \quad (6.73)$$

Thus, using linear filtering techniques an upper bound for $\left| \delta E^{(I,q)}(t) \right|$, denoted by $\tilde{E}^{(I,q)}(t)$, is determined as

$$\tilde{E}^{(I,q)}(t) = G_0^{(I,q)}(s) \left[\tilde{\eta}_F^{(I,q)}(t) \right] \quad (6.74)$$

Using (6.47) a bound for $\left| \delta Z^{(I,q)} \right|$, denoted by $\tilde{Z}^{(I,q)}(t)$, can be computed as

$$\tilde{Z}^{(I,q)}(t) = \tilde{E}^{(I,q)}(t) + G_2^{(I,q)}(s) \left[\tilde{E}^{(I,q)}(t) \right] = (G_2^{(I,q)}(s) + 1) \left[\tilde{E}^{(I,q)}(t) \right] \quad (6.75)$$

Taking into account (6.69), (6.74) and (6.75), we have

$$\begin{aligned}\bar{\epsilon}_{y_{Fj}}^{(I,q)}(t) &= G_{0j}^{(I,q)}(s)[\tilde{Z}^{(I,q)}(t)] + G_{0j}^{(I,q)}(s)[\tilde{\eta}_F^{(I,q)}(t)] \\ &= \left(G_{0j}^{(I,q)}(s)(G_2^{(I,q)}(s) + 1)G_0^{(I,q)}(s) + G_{0j}^{(I,q)}(s) \right) [\tilde{\eta}_F^{(I,q)}(t)]\end{aligned}\quad (6.76)$$

Combining (6.73) and (6.76) leads to (6.54). \square

Based on Lemma 6.2, we may infer that the effects of local sensor faults $f^{(I,q)}$ on the j -th residual $\varepsilon_{y_{Fj}}^{(I,q)}(t)$ and adaptive threshold $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t)$, $j \in \mathcal{J}^{(I,q)}$ depend on the design of the module $\mathcal{M}^{(I,q)}$, while the effects of propagated sensor faults $f_z^{(I)}$ depend explicitly on the Lipschitz interconnection dynamics $h^{(I)}$ and bound on the modeling uncertainty $\bar{\eta}^{(I)}$. The effects of local sensor faults $f^{(I,q)}$ on the j -th residual and adaptive threshold do also depend on the Lipschitz interconnection dynamics and bound on the modeling uncertainty, but in an implicit way, that is the design of the module $\mathcal{M}^{(I,q)}$ is realized taking into account the Lipschitz constants λ_{h_I} and λ_{η_I} , as analyzed in Theorem 4.1.

Moreover, based on (6.54) and (6.53), we may infer that the sensor fault propagation effects on the j -th residual and adaptive threshold depend on the interconnection function, represented by the Lipschitz constant λ_{h_I} , and the modeling uncertainty represented by the Lipschitz constant λ_{η_I} . If λ_{h_I} is very high due to strong interconnection dynamics or λ_{η_I} is very high due to very limited knowledge of the interconnections, propagated sensor faults can have high effects on the residuals and adaptive thresholds associated with the module $\mathcal{M}^{(I,q)}$. This means that propagated sensor faults of large magnitude may have low impact on the j -th residual and adaptive threshold or propagated sensor faults of small magnitude may have high impact on the j -th residual and adaptive threshold.

On the other hand, the effects of sensor faults $f^{(I)}$ on the j -th residual and adaptive threshold can be designed to be higher than the sensor fault propagation effects, because they depend on the magnitude of the observer gain $|L^{(I,q)}|$. This quantitative analysis is the basis of the design of the sensor fault signature matrix implemented in the global aggregation module \mathcal{G} , which may differentiate qualitatively the

sensitivity of $\mathcal{E}^{(I)}$ with respect to the propagated sensor faults $f_z^{(I)}$ and the local sensor faults $f^{(I)}$ in the I -th CPS.

Lemmas 6.1 and 6.2 along with Proposition 6.1 can be taken into account during the design of the module $\mathcal{M}^{(I,q)}$ for selecting the design parameters aiming at: (i) reducing the adaptive threshold under healthy conditions $\bar{\varepsilon}_{y_{Hj}}^{(I,q)}$, (ii) amplify the effects of local sensor faults on the j -th residual and adaptive threshold, and (iii) (if possible) attenuate the effects of propagated sensor faults on the j -th residual and adaptive threshold. Regarding the last specification, since the effects of propagated sensor faults $f_z^{(I)}$ depend explicitly on the Lipschitz interconnection dynamics $h^{(I)}$ and bound on the modeling uncertainty $\bar{\eta}^{(I)}$, a coordinate transformation may help in reducing the values of the Lipschitz constants λ_{h_I} and λ_{η_I} (see Proposition 6.1) [Reppa et al. [2015b]].

Since the conditions in Lemma 6.2 and Proposition 6.1 are parameterized with respect to $f^{(I,q)}$ and $f_z^{(I)}$, during the design stage someone can select several time profiles and fault functions (see Section 2.2) and either analytically or through simulations to find the characteristics of sensor faults that are detectable by the designed modules. For example, assuming the occurrence of abrupt, offset local and propagated sensor faults, someone can define the smallest magnitude of these faults that satisfy the condition 6.1 and compute the bound on the sensor fault effects through the conditions in Proposition 6.1. Moreover, someone can start by choosing an observer gain, and parameters of the adaptive thresholds according to the conditions presented in Theorem 4.1 and then to find the fault with e.g. the smallest magnitude. If the computed smallest magnitude is high, then this procedure should be repeated starting with new values for observer gain, and parameter of adaptive thresholds.

6.1.2 Decentralized Analysis

The decentralized effects of the sensor faults $f^{(I,q)}$ on the j -th residual and adaptive threshold can be analyzed following the procedure presented in Section 6.1.1. Given that $T_0^{(I,q)}$ is the first time instant of that a fault occurs in $\mathcal{S}^{(I,q)}$, the effects of sensor faults $f^{(I,q)}$ are described

in Proposition 6.2.

Proposition 6.2. In a decentralized architecture, the effects of sensor faults $f^{(I,q)}$ on the j -th residual and adaptive threshold, i.e. $\varepsilon_{y_{Fj}}^{(I,q)}$ and $\bar{\varepsilon}_{y_{Fj}}^{(I,q)}$, $j \in \mathcal{J}^{(I,q)}$, respectively, are defined as

$$\begin{aligned} \varepsilon_{y_{Fj}}^{(I,q)}(t) = & \int_{T_0^{(I,q)}}^t C_j^{(I)} e^{A_L^{(I,q)}(t-\tau)} \left(\gamma^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau)) \right. \\ & - \gamma^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau)) + h^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau)) \\ & - h^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau)) - L^{(I,q)} f^{(I,q)}(\tau) \Big) d\tau \\ & + f_j^{(I)}(t) \end{aligned} \quad (6.77)$$

$$\begin{aligned} \bar{\varepsilon}_{y_{Fj}}^{(I,q)}(t) = & \int_{T_1^{(I,q)}}^t \alpha_j^{(I,q)} e^{-\zeta_j^{(I,q)}(t-\tau)} \left(\Lambda_I E_F^{(I,q)}(\tau) \right. \\ & + \rho^{(I,q)} \Lambda_I^2 \int_{T_1^{(I,q)}}^\tau E_F^{(I,q)}(s) e^{-(\xi^{(I,q)} - \rho^{(I,q)} \Lambda_I)(\tau-s)} ds \\ & + \bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) \\ & \left. - \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) \right) d\tau, \end{aligned} \quad (6.78)$$

$$\begin{aligned} \delta E^{(I,q)}(t) = & \int_{T_1^{(I,q)}}^t \rho^{(I,q)} \left(\bar{\eta}^{(I)}(\hat{x}^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) \right. \\ & \left. - \bar{\eta}^{(I)}(\hat{x}_H^{(I,q)}(\tau), u^{(I)}(\tau), z_r^{(I)}(\tau), \tau) \right) e^{-\xi^{(I,q)}(t-\tau)} d\tau \end{aligned} \quad (6.80)$$

for $t \geq T_0^{(I,q)}$ and for at least one $j \in \mathcal{J}^{(I,q)}$.

Proof. The proof is similar to the proof in Proposition 6.1. \square

If $\mathcal{S}^{(I,q)}$ contains a single sensor, condition (6.3) guarantees the isolation of the fault affecting this sensor. Similarly, we may derive certain conditions under which a sensor fault $f_j^{(I)}$, $j \in \mathcal{J}^{(I,q)}$ that affects the local sensor set $\mathcal{S}^{(I,q)}$ and is included in the diagnosis set of

the module $\mathcal{M}^{(I,q)}$ is guaranteed to be isolated by a module $\mathcal{M}^{(I,q^*)}$, as described in Section 5.1.2, given that $\mathcal{S}^{(I,q^*)} \cap \mathcal{S}^{(I,q)} = \emptyset$ and $\mathcal{S}^{(I,q^*)}$ is healthy. In particular, the diagnosed sensor fault $f_j^{(I)}$, i.e. $f_j^{(I)} \in \mathcal{F}_{c_i}^{(I)}$, $i \in \mathcal{I}_D^{(I)}$, can be isolated by the module $\mathcal{M}^{(I,q^*)}$, if

$$\left| f_j^{(I)}(t) \right| > 2\bar{\varepsilon}_{y_{Hj}}^{(I,q^*)}(t), \quad j \in \{1, \dots, m_I\}, \quad j \notin \mathcal{J}^{(I,q^*)}. \quad (6.81)$$

This can be proved based on (6.3) and (6.77)-(6.80) for $q = q^*$ and taking into account that $f^{(I,q^*)} = 0$ and $\hat{x}_F^{(I,q^*)}(t) = \hat{x}_H^{(I,q^*)}(t)$ in (6.77)-(6.80), since $\mathcal{S}^{(I,q^*)}$ is healthy and the estimation model provided by $\mathcal{M}^{(I,q^*)}$ is healthy.

Based on Proposition 6.2, there are no effects of sensor faults that occur in neighboring subsystems on the j -th residual and adaptive threshold. The occurrence of sensor faults in neighboring subsystems may affect the tracking error $z^{(I)}(t) - z_r^{(I)}(t)$ in the sense that the actual signal will not track the reference signal. In this case, it may need to define a large bound $\bar{z}^{(I)}$ such that $\left| z^{(I)}(t) - z_r^{(I)}(t) \right| \leq \bar{z}^{(I)}$ under healthy and faulty conditions, which will lead to a larger adaptive threshold under healthy conditions $\bar{\varepsilon}_{y_{Hj}}^{(I,q)}$, making the detection of the local sensor faults more difficult according to Lemma 6.1. However, as previously mentioned, in the presence of "small" faults, which are the most difficult to detect, a robust controller will typically maintain a reasonably good tracking performance. The bound on the effects of sensor faults $f^{(I,q)}$ on the j -th residual and adaptive threshold can be defined based on Proposition 6.2 by setting $f_z^{(I)} = 0$.

6.1.3 Centralized Analysis

Following a centralized approach for diagnosing multiple sensor fault effects on a network of CPS that is treated as a monolithic system (see Section 3.1), the effects of the sensor faults $f^{(q)}$ on the j -th residual and adaptive threshold can be analyzed following the procedure presented in Section 6.1.1, summarized in the following proposition.

Proposition 6.3. In a centralized architecture, the effects of sensor faults $f^{(q)}$ on the j -th residual and adaptive threshold, i.e. $\varepsilon_{y_{Fj}}^{(q)}$ and $\bar{\varepsilon}_{y_{Fj}}^{(q)}$, $j \in \mathcal{J}^{(q)}$, respectively, for $t \geq T_0^{(q)}$, where for $T_0^{(q)}$ is the first time

instant of sensor fault occurrence in $\mathcal{S}^{(I,q)}$, are defined as

$$\begin{aligned} \varepsilon_{y_{Fj}}^{(q)}(t) = \int_{T_0^{(q)}}^t C_j e^{A_L^{(q)}(t-\tau)} & \left(\gamma(\hat{x}_H^{(q)}(\tau), u(\tau)) - \gamma(\hat{x}^{(q)}(\tau), u(\tau)) \right. \\ & \left. - L^{(q)} f^{(q)}(\tau) \right) d\tau + f_j(t) \end{aligned} \quad (6.82)$$

$$\begin{aligned} \bar{\varepsilon}_{y_{Fj}}^{(q)}(t) = \int_{T_0^{(q)}}^t \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-\tau)} & \left(\Lambda_I E_F^{(q)}(\tau) \right. \\ & + \rho^{(q)} \Lambda_I^2 \int_{T_0^{(q)}}^{\tau} E_F^{(q)}(s) e^{-(\xi^{(q)} - \rho^{(q)} \Lambda_I)(\tau-s)} ds \\ & \left. + \bar{\eta} \left(\hat{x}^{(q)}(\tau), u(\tau), \tau \right) - \bar{\eta} \left(\hat{x}_H^{(q)}(\tau), u(\tau), \tau \right) \right) d\tau, \end{aligned} \quad (6.83)$$

$$\begin{aligned} \delta E^{(q)}(t) = \int_{T_0^{(q)}}^t \rho^{(q)} & \left(\bar{\eta} \left(\hat{x}^{(q)}(\tau), u(\tau), z_r(\tau), \tau \right) \right. \\ & \left. - \bar{\eta} \left(\hat{x}_H^{(q)}(\tau), u(\tau), z_r(\tau), \tau \right) \right) e^{-\xi^{(q)}(t-\tau)} d\tau \end{aligned} \quad (6.84)$$

and for at least one $j \in \mathcal{J}^{(q)}$.

Proof. The proof is similar to the proof in Proposition 6.1. \square

The effects of sensor faults $f^{(q)}$ on the j -th residual and adaptive threshold, i.e. $\varepsilon_{y_{Fj}}^{(q)}$ and $\bar{\varepsilon}_{y_{Fj}}^{(q)}$, $j \in \mathcal{J}^{(q)}$, are uniformly bounded, i.e.

$$\left| \varepsilon_{y_{Fj}}^{(q)}(t) \right| \leq \bar{y}_{Fj}^{(q)}(t) \quad (6.85)$$

$$\left| \bar{\varepsilon}_{y_{Fj}}^{(q)}(t) \right| \leq \bar{\bar{\varepsilon}}_{y_{Fj}}^{(q)}(t) \quad (6.86)$$

where the bounds $\bar{y}_{Fj}^{(q)}(t)$ and $\bar{\bar{\varepsilon}}_{y_{Fj}}^{(q)}(t)$ are determined following the steps of the proof of Lemma 6.2 as

$$\bar{y}_{Fj}^{(q)}(t) = \left(\lambda_\gamma G_2^{(q)}(s) \left(G_1^{(q)}(s) + 1 \right) G_0^{(q)}(s) + 1 \right) |L^{(q)}| \left[\left| f^{(q)}(t) \right| \right] \quad (6.87)$$

$$\begin{aligned} \bar{\bar{\varepsilon}}_{y_{Fj}}^{(q)}(t) = & \left(G_2^{(q)}(s) \lambda_\eta (G_3^{(q)}(s) + 1) G_0^{(q)}(s) \left((G_1^{(q)}(s) + 1) G_0^{(q)}(s) \right) \right. \\ & \left. + \lambda_\eta G_2^{(q)}(s) \left((G_1^{(q)}(s) + 1) G_0^{(q)}(s) \right) \right) |L^{(q)}| \left[\left| f^{(q)}(t) \right| \right] \end{aligned} \quad (6.88)$$

6.2 Multiple Sensor Fault Isolability

The ability of the proposed decision logic to isolate multiple sensor faults in $\mathcal{S}^{(I)}$ for all $I \in \{1, \dots, M\}$ with $M \geq 1$ can be analyzed with respect to the design of the multiple sensor fault signature matrix that is used by the local and global aggregation modules $\mathcal{A}^{(I)}$ and \mathcal{G} , respectively. Note that the multiple sensor fault isolability of \mathcal{G} is analyzed assuming that $F_{I_k}^z = 1$ (instead of $F_{I_k}^z = *$), in the case that at least one sensor fault included in $\mathcal{F}_{c_k}^z$ is involved in $\mathcal{E}^{(I)}$ and affects $\mathcal{S}_z^{(I)}$, while none of the sensor faults included in $\mathcal{F}_{c_k}^z$ affects $\mathcal{S}^{(I)}$.

Based on the consistency test, a sensor fault combination is isolable, if its associated theoretical pattern of sensor faults is unique. Moreover, the number of the rows of the multiple sensor fault signature matrix is N_I , which is the number of the local sensor sets. The number of all possible observed patterns of sensor faults (defined in (5.1)) is equal to $2^{N_I} - 1$, while the number of all possible theoretical patterns of sensor faults is $2^{m_I} - 1$. The number of isolable sensor fault combinations N_I^* is maximized when all theoretical sensor fault patterns are distinct. Obtaining $2^{m_I} - 1$ distinct patterns can be achieved if every individual sensor satisfies the observer stability specifications given in Theorem 4.1. In this case, every sensor combination that includes these sensors will satisfy the observer stability conditions. On the other hand, if two or more sensor fault combinations have the same theoretical pattern of sensor faults, they are mutually non-isolable. Using the consistency test, all the mutually non-isolable sensor fault combinations are included in the diagnosis set $\mathcal{D}_s^{(I)}(t)$. For this reason, the smaller the number of mutually non-isolable sensor fault combinations the fewer the diagnosed combinations of multiple sensor faults that are possible to have occurred in $\mathcal{S}^{(I)}$.

The decomposition of the local sensor set $\mathcal{S}^{(I)}$ into smaller local sensor sets is realized using an auxiliary matrix associated with the multiple sensor fault signature matrix $F^{(I)}$. Particularly, the last basic step of Algorithm 1 (see lines 22-37) initiates with the determination of an auxiliary matrix, where its columns correspond to the index tuples of all possible combinations of sensors in $\mathcal{S}^{(I)}$, while the

rows correspond to the index tuples of $\mathcal{J}_s^{(I)}$, a set associated with the sensor combinations that satisfy the observer stability specifications. The number of distinct columns of this auxiliary matrix determines the maximum number of isolable sensor fault combinations, while the maximum number of mutually identical columns is the minimum number of mutually non-isolable sensor fault combinations. Aiming at reducing the computational complexity of the agent $\mathcal{M}^{(I)}$ (by reducing the number of monitoring modules) without attenuating its multiple sensor fault isolability, we seek to eliminate rows of the auxiliary matrix without modifying the number of distinct columns and maximum number of mutually identical columns (lines 22-37). In this way, we may obtain the minimum number of the local sensor sets without affecting the multiple sensor fault isolability of $\mathcal{M}^{(I)}$.

Remark 6.1. Taking into account the auxiliary matrices F_1^* and F_2^* presented in Tables 3.5 and 3.7 in Example 3.1, we can infer that if these matrices are used directly for isolation, the matrix F_1^* has more isolable sensor fault combinations compared to the isolable sensor fault combinations raised based on matrix F_2^* . Moreover, in case I the mutually non-isolable sensor fault combinations are four, which in case II the corresponding combinations are seven. Thus, in the latter case if these combinations are included in the diagnosis set, we infer that each of these combinations are equally possible to have occurred.

6.3 Illustrative Example

The goal of this section is to illustrate the performance of the monitoring modules with respect to the distributed detectability of local and propagated sensor faults and decentralized detectability, by simulating several levels of sensor noise and using various parameters for the design of the observer-based residuals and adaptive thresholds of the modules $\mathcal{M}^{(1,1)}$, $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,1)}$, $\mathcal{M}^{(2,2)}$ of the agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ that monitor subsystems $\Sigma^{(1)}$ and $\Sigma^{(2)}$, as described in Section 4.5. More specifically, we obtained simulation results for three different levels of noise, that is: (i) $d_q^{(I)} = 1\%x_1^\circ = 0.0628$, (ii) $d_q^{(I)} = 3\%x_1^\circ = 0.1885$, and (iii) $d_q^{(I)} = 5\%x_1^\circ = 0.3142$. In order to analyze the distributed

detectability of the monitoring modules, we have selected three different sets of eigenvalues of the matrix $A_L^{(I,q)} = A^{(I)} - L^{(I,q)}C^{(I,q)}$, $I, q = 1, 2$. In order to analyze the decentralized detectability of the modules $\mathcal{M}^{(1,2)}$, $\mathcal{M}^{(2,2)}$, which did not succeed to detect the occurrence of sensor faults (see Fig. 4.7), an additional set of eigenvalues has been used. All the design parameters for every monitoring module are provided in the Appendix A. Aiming at illustrating the methodology of detectability analysis, the results have been obtained by simulating the first four scenarios of single fault occurrence presented in Section 4.5.4, that is

Sensor fault scenario 1: Single occurrence of sensor fault $f_1^{(1)}$

Sensor fault scenario 2: Single occurrence of sensor fault $f_1^{(2)}$

Sensor fault scenario 3: Single occurrence of sensor fault $f_2^{(1)}$

Sensor fault scenario 4: Single occurrence of sensor fault $f_2^{(2)}$

where $f_q^{(I)}$ is an abrupt, permanent fault whose evolution rate is $\kappa_{f_q}^{(I)} = 10^4$, the time of fault occurrence is $T_{f_q}^{(I)} = 150$ sec and fault function is an offset $\phi_q^{(I)}(t) = 25\%x_1^\circ = \pi/2$, for all $I, q = 1, 2$.

Before analyzing the performance of the monitoring modules $\mathcal{M}^{(1,1)}$, $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,1)}$, $\mathcal{M}^{(2,2)}$ in this case study, it has to be noted that: (i) the monitoring modules $\mathcal{M}^{(1,1)}$ and $\mathcal{M}^{(2,1)}$ (correspondingly for $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,2)}$) have the same structure, that is the observer-based residuals and adaptive thresholds have the same design and (ii) no optimization procedure has been realized for selecting the values of the parameters for designing the adaptive thresholds, and particularly parameters $\rho^{(I,q)}$, $\xi^{(I,q)}$, $\rho_d^{(I,q)}$, $\xi_d^{(I,q)}$, $\alpha^{(I,q)}$, $\zeta^{(I,q)}$, $\alpha_d^{(I,q)}$ and $\zeta_d^{(I,q)}$, but the values have been selected such that (4.189)-(4.191), (4.198) and (4.208)-(4.209) are fulfilled, while keeping the ratios $\rho^{(I,q)}/\xi^{(I,q)}$, $\rho_d^{(I,q)}/\xi_d^{(I,q)}$, $\alpha^{(I,q)}/\zeta^{(I,q)}$, and $\alpha_d^{(I,q)}/\zeta_d^{(I,q)}$ small. The optimal selection of the aforementioned parameters is beyond of the scope of this tutorial.

Figures 6.1-6.8 present the results of the distributed detection process of every monitoring module when local (Figs. 6.1, 6.3, 6.5, and

6.7) and propagated (Figs. 6.2, 6.4, 6.6 and 6.8) sensor faults occur according to scenarios 1-4. Every figure consists of nine subfigures, where the results shown in the i -th row of subfigures, $i = 1, 2, 3$ have been derived by simulating sensor noise with the same bound, while the results shown in i -th column of subfigures, $i = 1, 2, 3$ have been derived by selecting the same set of eigenvalues of the matrix $A_L^{(I,q)}$. It is noted that the limits of the vertical axes of every subfigure are the same for comparison reasons, while in the horizontal axis only the time of sensor fault detection (if a sensor fault is detected) is shown along with the beginning and end time of simulations.

Figure 6.1 presents the results of the detection process of the monitoring module $\mathcal{M}^{(1,1)}$ in the first simulated scenario (occurrence of local sensor faults). In all cases, the local sensor fault $f_1^{(1)}$ was detected just after its occurrence. As observed, when the eigenvalues of $A_L^{(1,1)}$ are characterized by low magnitude and the sensor noise bound is low, the local sensor fault $f_1^{(1)}$ is strongly detectable by the module $\mathcal{M}^{(1,1)}$ in the sense that not only it is detected but also its effects on the residual are persistent and do not vanish with time [Chen and Patton [1999]]. For example, based on e.g. Fig. 6.1-1, after the occurrence (and detection) of the sensor fault, the magnitude of the residual remains above the adaptive threshold up to the end of the simulation time. According to e.g. 6.1-3 the magnitude of the residual exceeds the adaptive threshold for a short time interval. In general, the residual is attenuated as the magnitude of the eigenvalues of $A_L^{(1,1)}$ becomes larger. On the other hand, both the residual and the adaptive threshold are intensified when the sensor noise bound becomes larger. However, the increase or decrease of the adaptive threshold with respect to the increase of the magnitude of the eigenvalues is not straightforward, due to the way of selecting the parameters $\rho^{(1,1)}$, $\xi^{(1,1)}$, $\rho_d^{(1,1)}$, and $\xi_d^{(1,1)}$, which may not be optimal in any case.

For the second simulated scenario, the results of the distributed detection process of the module $\mathcal{M}^{(1,1)}$ are shown in Fig. 6.2. As observed, the occurrence of the propagated sensor fault $f_1^{(2)}$ is identified by the monitoring module in the case of the lower sensor noise bound and lower-magnitude eigenvalues of $A_L^{(1,1)}$ (see Figs. 6.2-1 and 6.2-2).

Taking into account Figs. 6.1-1-6.1-3 and 6.2-1-6.2-3, we may infer that selecting the eigenvalues of $A_L^{(1,1)}$ to have lower magnitude can amplify the sensitivity of the set of ARR_s $\mathcal{E}^{(1,1)}$ and consequently the decision of $\mathcal{M}^{(1,1)}$ to the occurrence of local and propagated sensor faults, that is $f_1^{(1)}$ and $f_1^{(2)}$, while selecting the eigenvalues of $A_L^{(1,1)}$ to have higher magnitude can make the module $\mathcal{M}^{(1,1)}$ sensitive to local sensor fault $f_1^{(1)}$ only. As observed in Figs. 6.1 and 6.2 and according to Proposition 6.1 and Lemma 6.2, the effects of sensor faults $f_1^{(1)}$ and $f_1^{(2)}$ on the decision of the monitoring module $\mathcal{M}^{(1,1)}$ are manifested in different ways, although the local and propagated sensor fault are characterized by the same time profile and fault function. Moreover, the effects of local sensor fault $f_1^{(1)}$ are more strongly related to the design parameters of the module $\mathcal{M}^{(1,1)}$ than the effects of propagated sensor fault $f_1^{(2)}$, which are explicitly related to the interconnection dynamics.

The distributed detectability performance of the monitoring module $\mathcal{M}^{(2,1)}$ is similar to the distributed detectability performance of $\mathcal{M}^{(1,1)}$ with respect to various sensor noise bounds and sets of eigenvalues of $A_L^{(2,1)}$, as shown in Figs. 6.3 (local sensor fault) and 6.4 (propagated sensor fault). In the second scenario where the occurrence of the local sensor fault $f_1^{(2)}$ has been simulated, the module $\mathcal{M}^{(2,1)}$ detected $f_1^{(2)}$ just after its occurrence irrespective of the magnitude of the eigenvalues of $A_L^{(2,1)}$ and the sensor noise bounds, while the occurrence of the propagated sensor fault $f_1^{(1)}$ has been captured by the module $\mathcal{M}^{(2,1)}$ in the case of selecting eigenvalues of $A_L^{(2,1)}$ with low magnitude and lower noise bound.

The results of the distributed detectability performance of $\mathcal{M}^{(1,2)}$ when the local sensor fault $f_2^{(1)}$ occurs (third sensor fault scenario) are illustrated in Fig. 6.5, while Fig. 6.6 shows the decision-making performance results of $\mathcal{M}^{(1,2)}$ when the local sensor fault $f_1^{(2)}$ occurs (second sensor fault scenario). The local sensor fault effects are more strong than the propagated sensor fault effects, leading to the detection of the local sensor fault in all cases. On the other hand, the propagated sensor fault $f_1^{(2)}$ was detected only in the case of lowest sensor noise bound and set of eigenvalues of lowest magnitude. For this particular case of sensor noise bound and eigenvalues, the module $\mathcal{M}^{(1,2)}$ becomes

more sensitive to the propagated sensor fault $f_1^{(2)}$ than the module $\mathcal{M}^{(1,1)}$ (see Figs. 6.2-1 and 6.6-1).

The performance of the module $\mathcal{M}^{(2,2)}$ to local and propagated sensor faults, that is $f_2^{(2)}$ and $f_1^{(1)}$ is presented in Figs. 6.7 and 6.8. In contrast to the other monitoring modules and especially $\mathcal{M}^{(2,1)}$, the ability of $\mathcal{M}^{(2,2)}$ to detect the local fault $f_2^{(2)}$ is quite weak, while $f_2^{(2)}$ is not detected in the case that the sensor noise bound is the highest one and the eigenvalues of $A_L^{(2,2)}$ have the lowest magnitude (see Fig. 6.7-9). When the eigenvalues of the lowest magnitude are selected for the design of $\mathcal{M}^{(2,2)}$, the magnitude of the residual after the occurrence of the propagated sensor fault is comparable to the magnitude of the residual after the occurrence of the local sensor fault (see Figs. 6.7-1, 6.7-4 and 6.7-7 and 6.8-1, 6.8-4 and 6.8-7). This may be justified based on the interconnection dynamics of $\Sigma^{(2)}$, the selected design parameters shown in Table A.4 and the bounds on propagated and sensor fault effects on the residual and adaptive threshold given in Lemma 6.2.

Since the occurrence of sensor faults $f_2^{(1)}$ and $f_2^{(2)}$ was not detected by the modules $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,2)}$, in the case that $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,2)}$ were developed following a decentralized approach (see Fig. 4.7 in Section 4.5.4), the detectability performance of these monitoring modules is analyzed by simulating the third and the fourth sensor fault scenario for various sensor noise bounds and eigenvalues of $A_L^{(1,2)}$ and $A_L^{(2,2)}$ with larger magnitudes. The results of this performance analysis are shown in Figs. 6.9 and 6.10. By respectively comparing Figs. 6.9 and 6.10 to Figs. 6.5 and 6.7, it can be observed that the decentralized adaptive thresholds are much higher than the distributed adaptive thresholds, weakening the detectability of the modules. Based on 6.9, if the magnitude of the eigenvalues of $A_L^{(1,2)}$ is high, the detectability of $\mathcal{M}^{(1,2)}$ is amplified. However, this is not observed in the performance analysis of $\mathcal{M}^{(2,2)}$ shown in Fig. 6.10. In none of the cases, $\mathcal{M}^{(2,2)}$ detected the fault $f_2^{(2)}$. The discrepancy in the detectability performance between $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,2)}$ stems from the fact that the interconnection dynamics of $\Sigma^{(2)}$ are stronger (i.e. the Lipschitz constant is higher) than the interconnection dynamics of $\Sigma^{(1)}$, leading to a higher adaptive threshold (see Section 4.5.2).

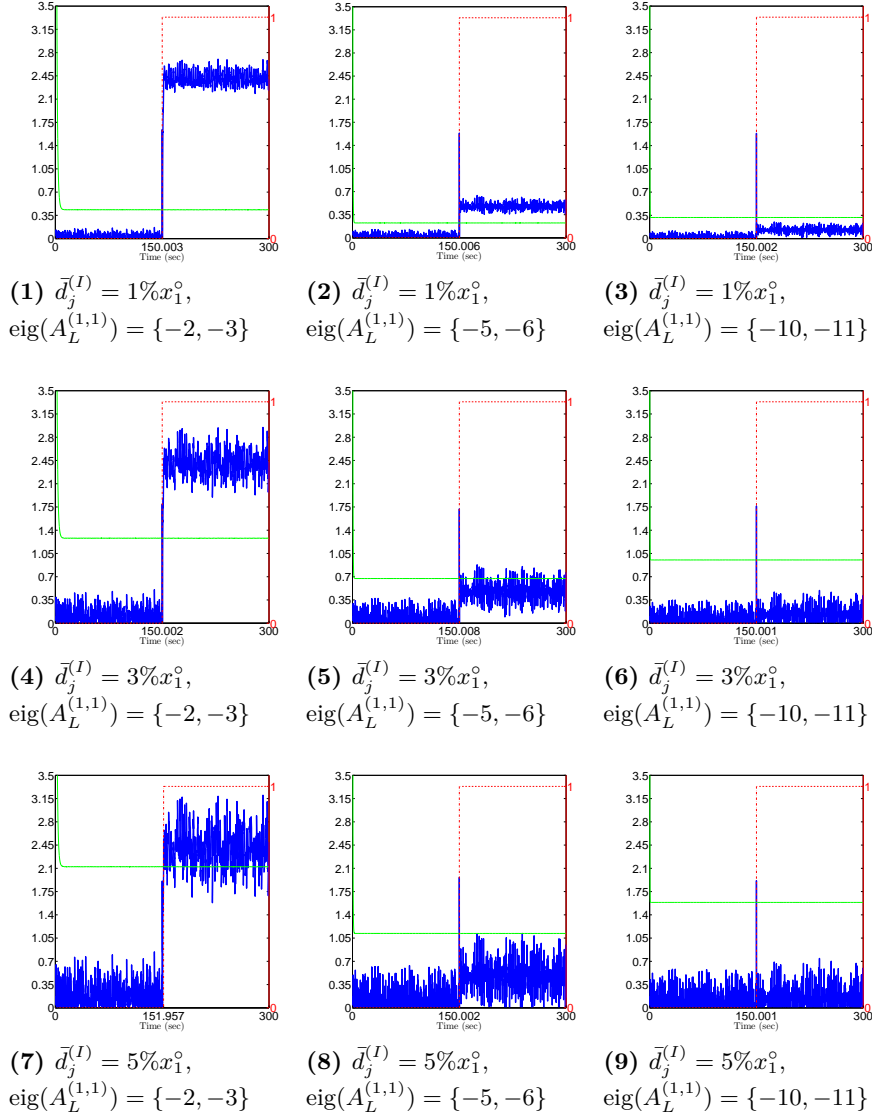


Figure 6.1: Distributed detectability performance of the module $\mathcal{M}^{(1,1)}$ when $f_1^{(1)}$ occurs at $T_{f_1}^{(1)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_1}^{(1,1)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_1}^{(1,1)}(t)$ (green line) and $D_1^{(1,1)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(1,1)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

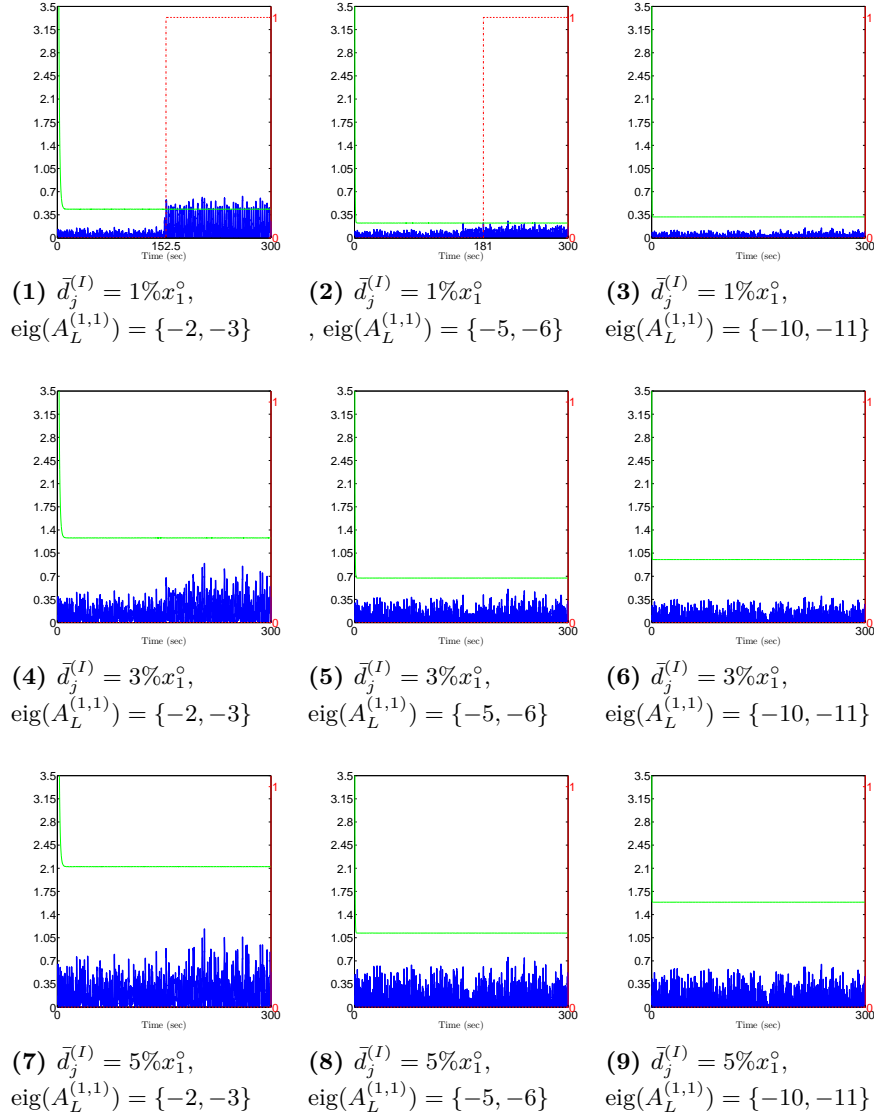


Figure 6.2: Distributed detectability performance of the module $\mathcal{M}^{(1,1)}$ when $f_1^{(2)}$ occurs at $T_{f_1}^{(2)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_1}^{(1,1)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_1}^{(1,1)}(t)$ (green line) and $D_1^{(1,1)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(1,1)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

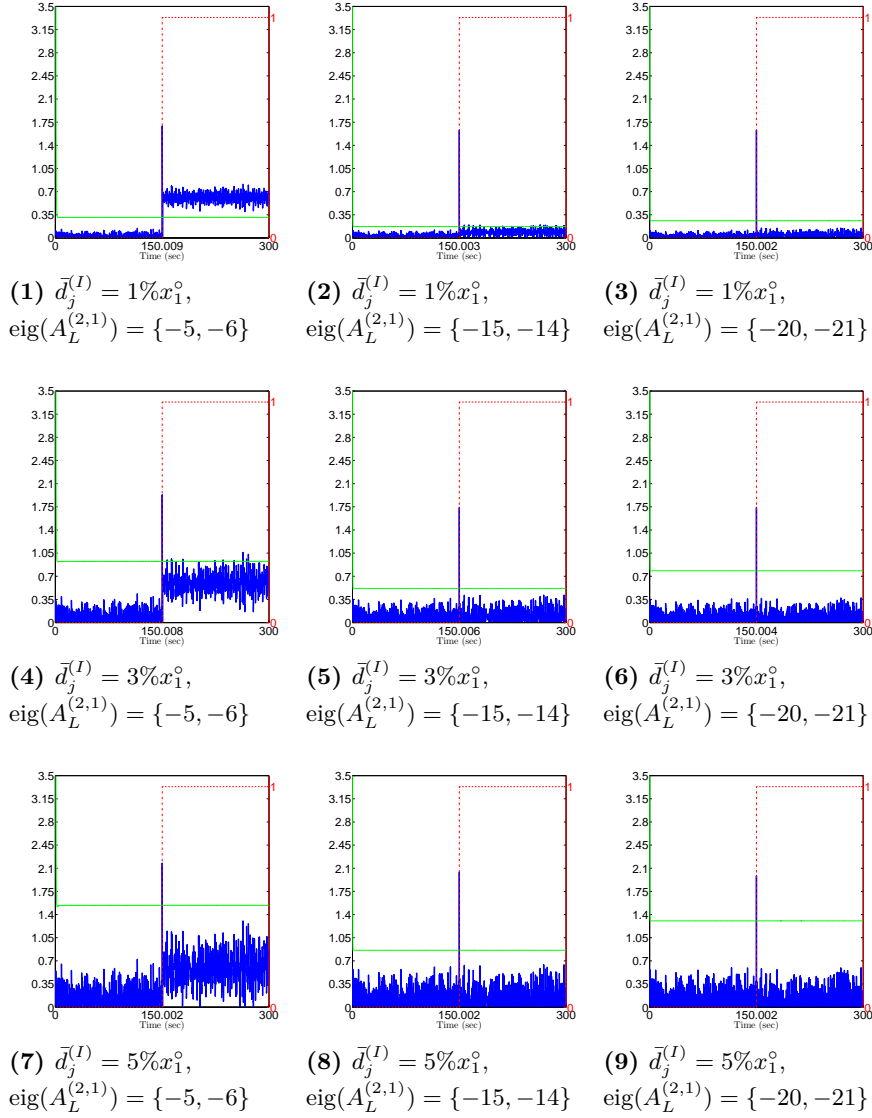


Figure 6.3: Distributed detectability performance of the module $\mathcal{M}^{(2,1)}$ when $f_1^{(2)}$ occurs at $T_{f_1}^{(2)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_1}^{(2,1)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_1}^{(2,1)}(t)$ (green line) and $D_1^{(2,1)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(2,1)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

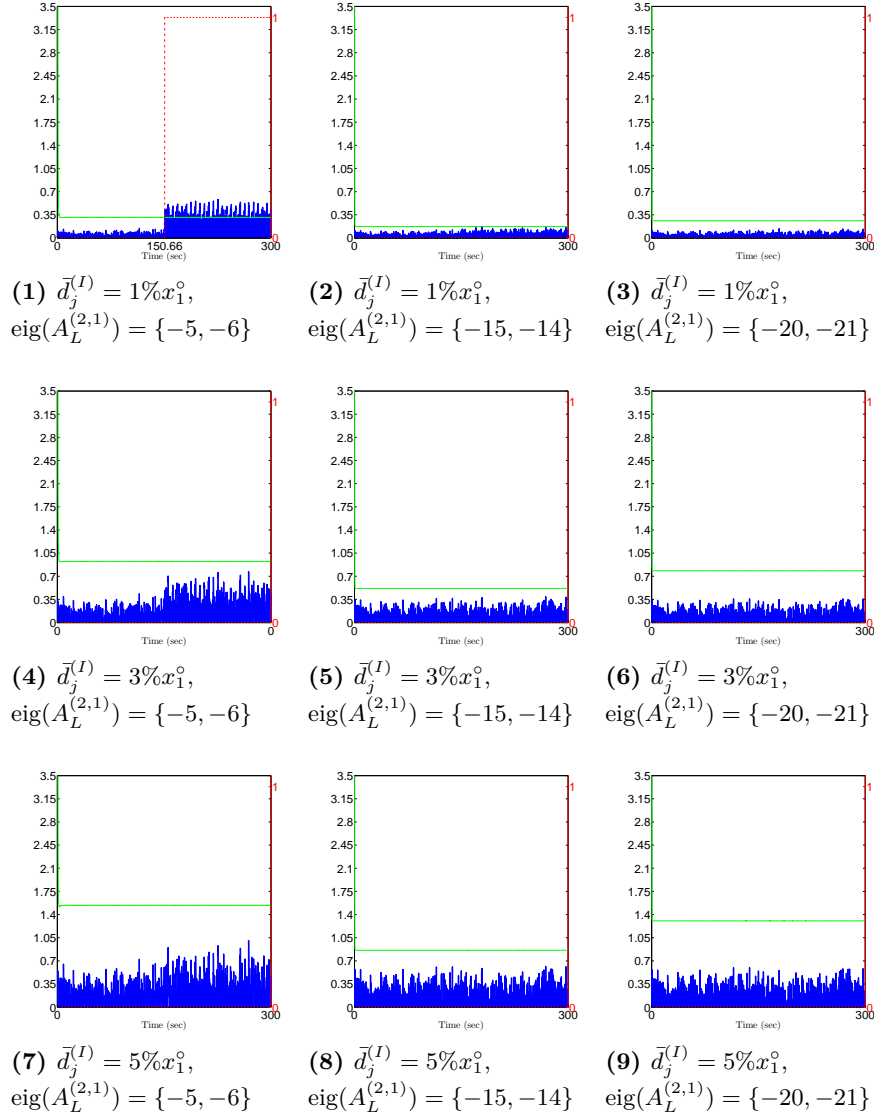


Figure 6.4: Distributed detectability performance of the module $\mathcal{M}^{(2,1)}$ when $f_1^{(1)}$ occurs at $T_{f_1}^{(1)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_1}^{(2,1)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_1}^{(2,1)}(t)$ (green line) and $D_1^{(2,1)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(2,1)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

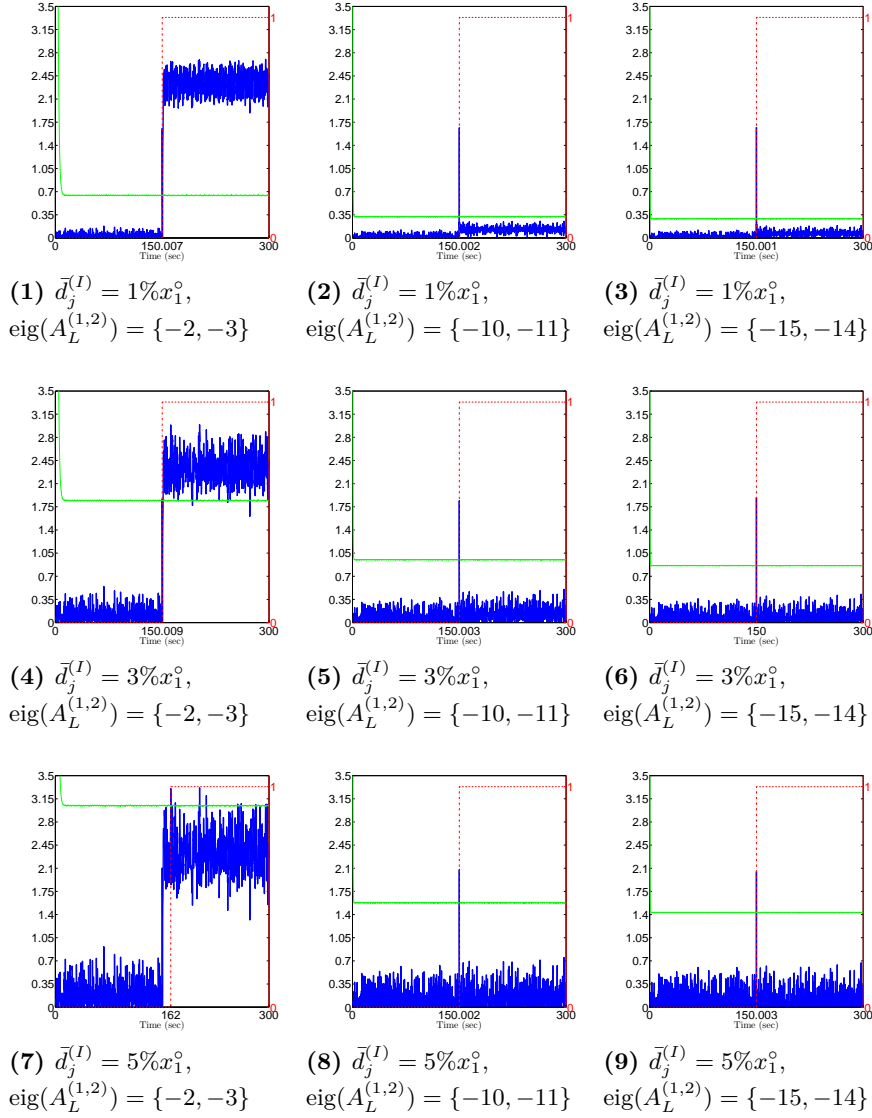


Figure 6.5: Distributed detectability performance of the module $\mathcal{M}^{(1,2)}$ when $f_2^{(1)}$ occurs at $T_{f_2}^{(1)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_2}^{(1,2)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_2}^{(1,2)}(t)$ (green line) and $D_2^{(1,2)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(1,2)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

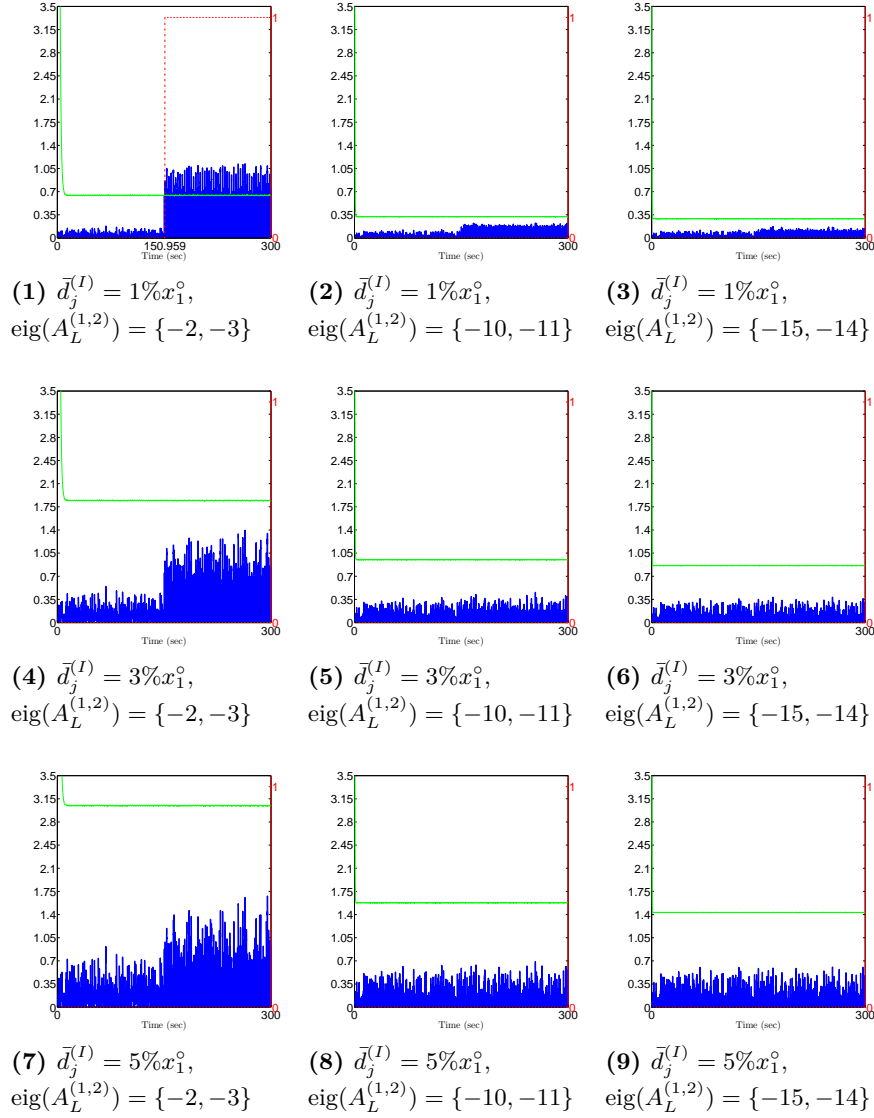


Figure 6.6: Distributed detectability performance of the module $\mathcal{M}^{(1,2)}$ when $f_1^{(2)}$ occurs at $T_{f_1}^{(2)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_2}^{(1,2)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_2}^{(1,2)}(t)$ (green line) and $D_2^{(1,2)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(1,2)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

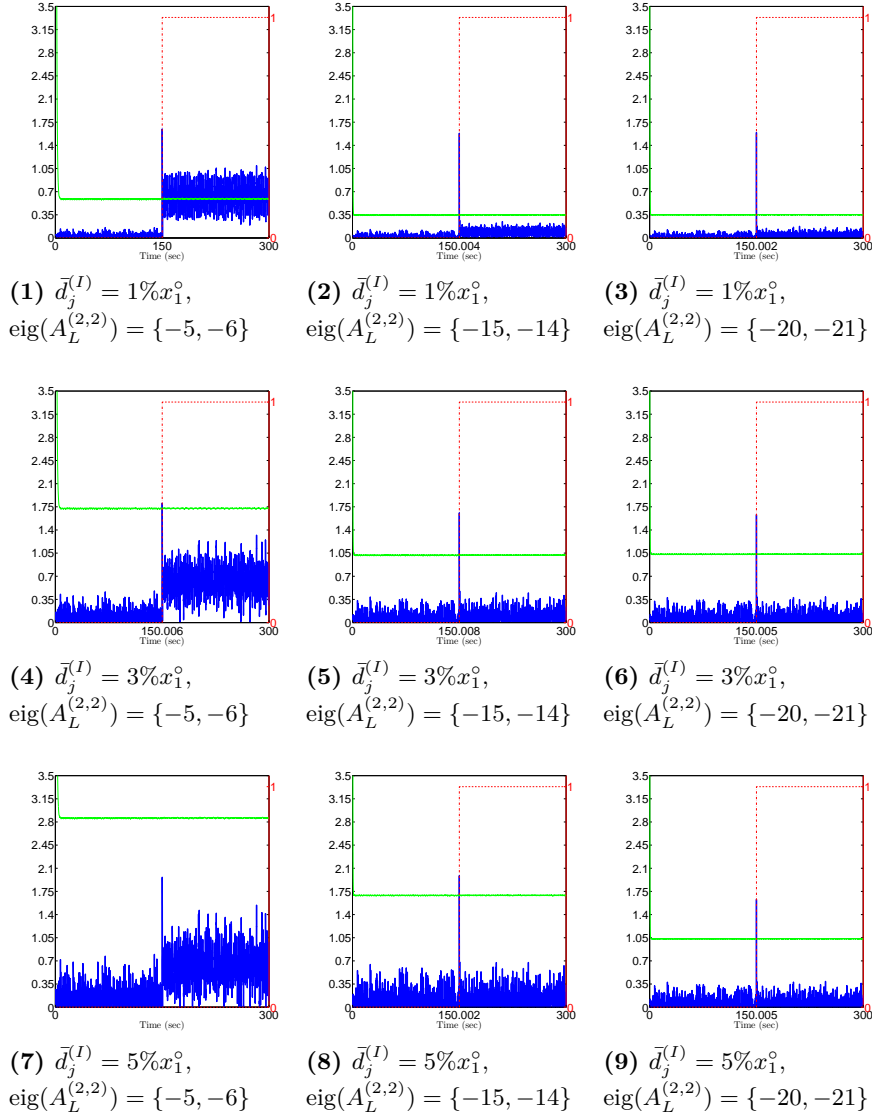


Figure 6.7: Distributed detectability performance of the module $\mathcal{M}^{(2,2)}$ when $f_2^{(2)}$ occurs at $T_{f_2}^{(2)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_2}^{(2,2)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_2}^{(2,2)}(t)$ (green line) and $D_2^{(2,2)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(2,2)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

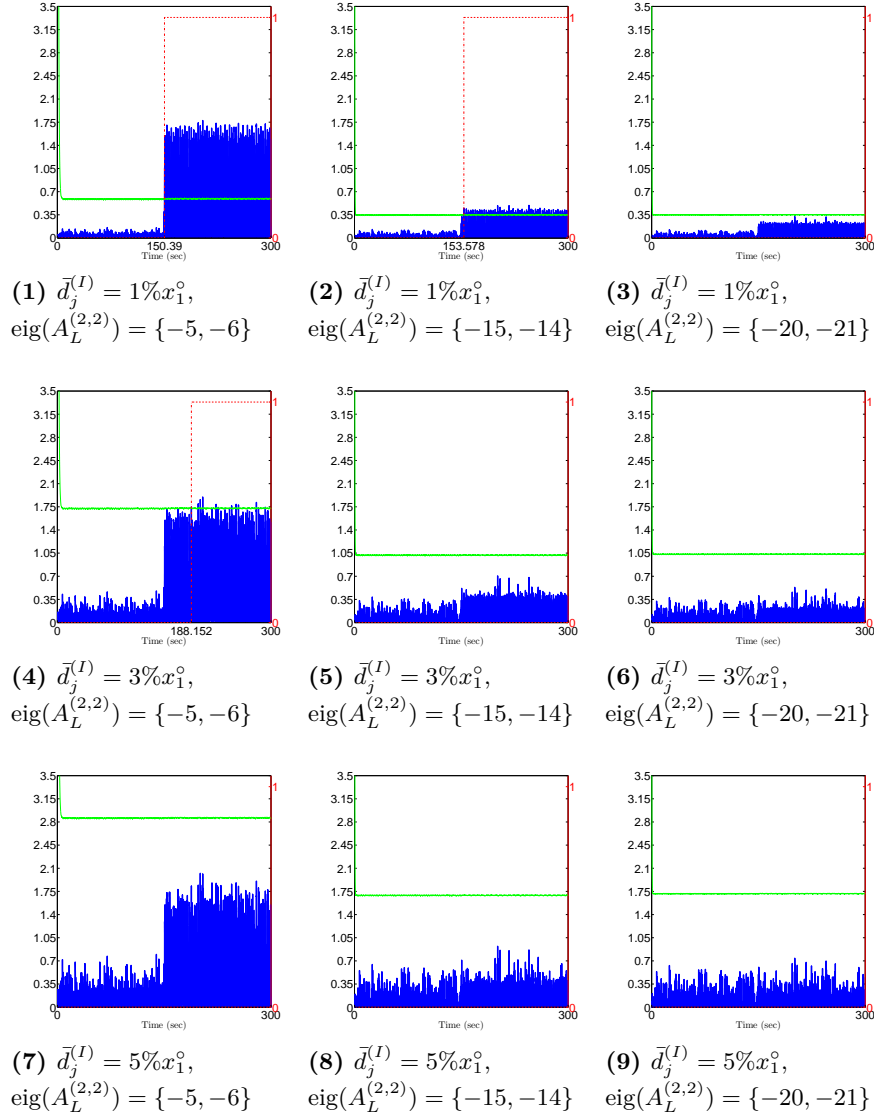


Figure 6.8: Distributed detectability performance of the module $\mathcal{M}^{(2,2)}$ when $f_1^{(1)}$ occurs at $T_{f_1}^{(1)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $\left| \varepsilon_{y_2}^{(2,2)}(t) \right|$ (blue line), $\bar{\varepsilon}_{y_2}^{(2,2)}(t)$ (green line) and $D_2^{(2,2)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(2,2)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

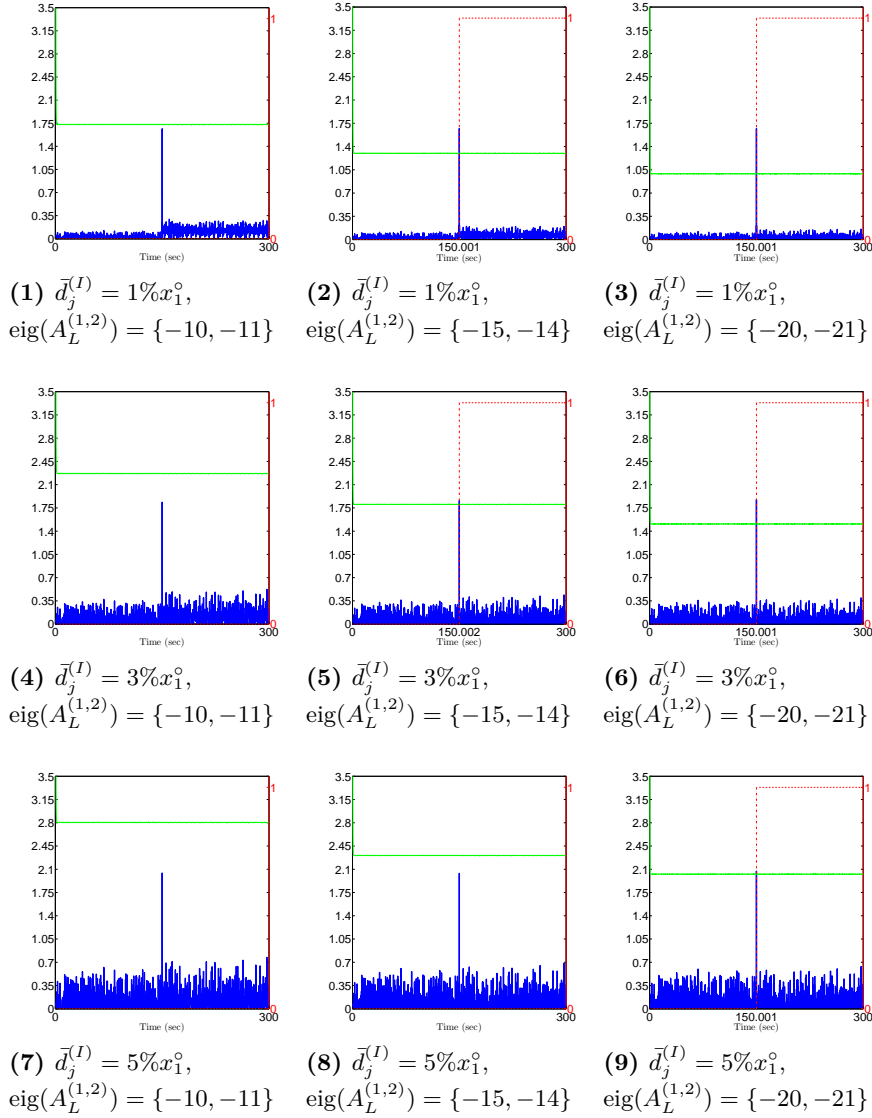


Figure 6.9: Decentralized detectability performance of the module $\mathcal{M}^{(1,2)}$ when $f_2^{(1)}$ occurs at $T_{f_2}^{(1)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $|\varepsilon_{y_2}^{(1,2)}(t)|$ (blue line), $\bar{\varepsilon}_{y_2}^{(1,2)}(t)$ (green line) and $D_2^{(1,2)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(1,2)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

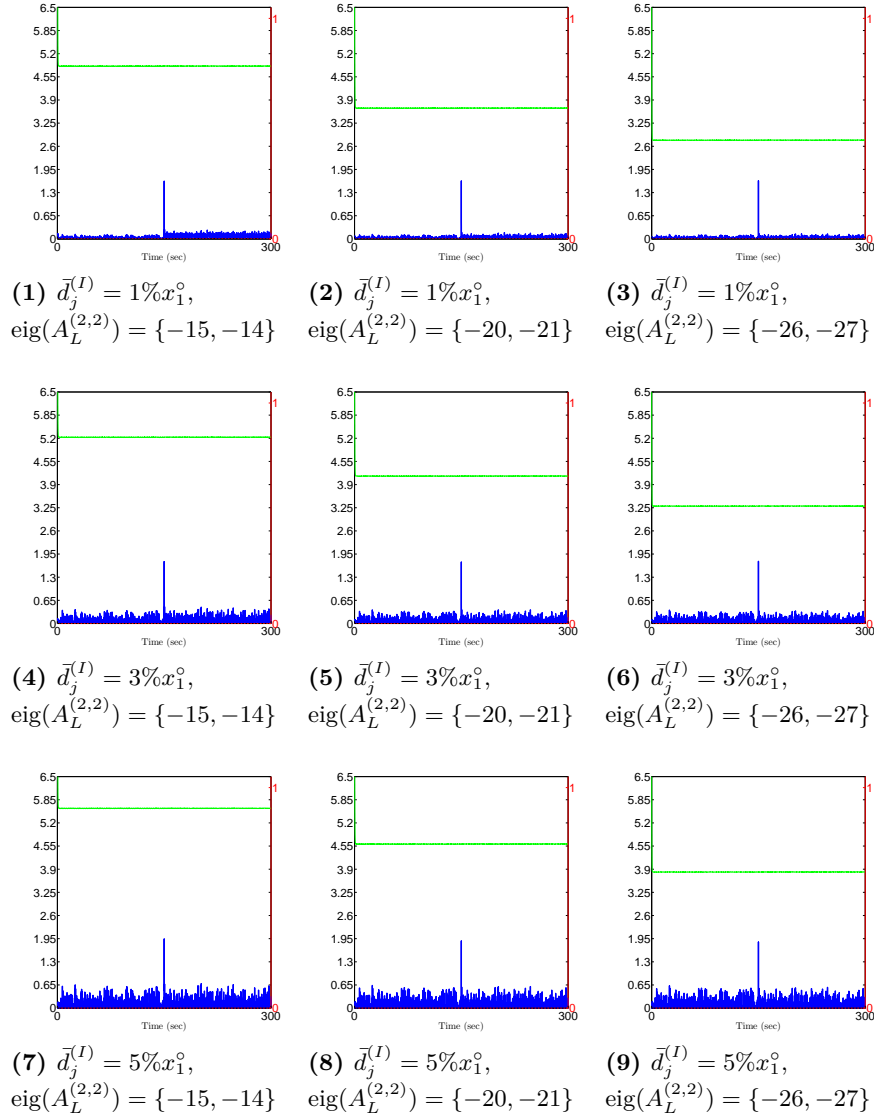


Figure 6.10: Decentralized detectability performance of the module $\mathcal{M}^{(2,2)}$ when $f_2^{(2)}$ occurs at $T_{f_2}^{(2)} = 150$ sec. Subfigures 1-9 show the temporal evolution of $|\varepsilon_{y_2}^{(2,2)}(t)|$ (blue line), $\bar{\varepsilon}_{y_2}^{(2,2)}(t)$ (green line) and $D_2^{(2,2)}(t)$ (red, dashed line), where the results in every subfigure are obtained by simulating three different sets of eigenvalues of $A_L^{(2,2)}$, and sensor noise with three different bounds $\bar{d}_q^{(I)}$, for all $I, q = 1, 2$.

6.4 Summary and Discussion

In this chapter, we analyzed the performance of a sensor fault diagnosis method that follows a distributed, decentralized or centralized approach for detecting and isolating multiple sensor faults that may occur in a network of cyber-physical systems. The performance was initially investigated with respect to sensor fault detectability, by characterizing the minimum effects of sensor faults on the residuals and adaptive thresholds, which are detectable by the monitoring modules. Secondly, we analyzed the performance of the monitoring agents in relation to multiple sensor fault isolability by characterizing the number of isolable, and mutually isolable sensor fault combinations. Both the sensor fault detectability and multiple sensor fault isolability analysis were based on the the properties of robustness and structured sensor fault sensitivity of the monitoring modules, discussed in Section 4.3. The added value of the performance analysis is that the design parameters of the nonlinear observer and the adaptive thresholds can be selected not only to satisfy the observer stability conditions only, but also to amplify the sensor fault effects on the residuals and the adaptive thresholds, and to contribute to distinguishing local and propagated sensor faults.

7

Learning Approaches for Sensor Fault Diagnosis

In chapter 6, it is shown that the detectability of a sensor fault diagnosis method depends on the bounds of the modeling uncertainty and measurement noise and the design parameters of the monitoring modules. Additionally, some rules of thumb were given on how to select the design parameters based on the analysis of sensor fault effects on the residual and adaptive thresholds. An alternative or complementary way to reduce the adaptive thresholds is to reduce the bound of the modeling uncertainty using learning techniques. The objective of this chapter is the design and analysis of an adaptive approximation methodology, which learns the modeling uncertainty for an initial time interval, assuming that no sensor is faulty. For the sake of simplicity, the design and analysis of the learning technique is realized following a centralized approach and taking into account system (3.1) and (3.2). The general architecture describing this methodology is illustrated in Fig. 7.1. The goal of the adaptive approximation is to learn for an initial time interval $[0, T_L]$ the unknown functional nonlinearity $\eta(x(t), u(t))$, thus reducing the uncertainty in the system model and optimizing the thresholds in the distributed sensor fault diagnosis.

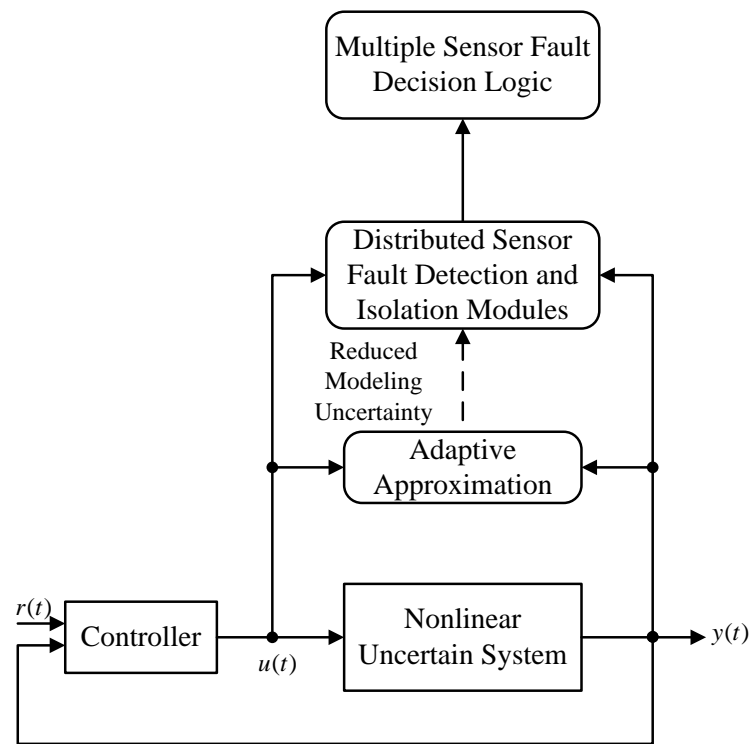


Figure 7.1: General architecture for sensor fault diagnosis with adaptive approximation in a centralized architecture.

7.1 Parametric Model

The adaptive approximation of the unknown $\eta(x, u)$ is realized taking into account the following assumption.

Assumption 6: During the time interval $[0, T_L]$, there are no sensor faults; i.e. $f(t) = 0$ for $t \in [0, T_L]$.

By considering the modeling uncertainty η as a function of the system state and input, i.e. $\eta(x, u)$, the state equation (3.1) can be re-written as

$$\dot{x}(t) = Ax(t) + \gamma(x(t), u(t)) + \hat{\eta}(x(t), u(t); \theta^*) + e_\eta(x(t), u(t)), \quad (7.1)$$

where $\hat{\eta} : \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^{q_\theta} \mapsto \mathbb{R}^n$ is an adaptive nonlinear approximation model (e.g. radial basis function, sigmoidal neural network, wavelet network etc.) and $\theta^* = [\theta_1^{*\top}, \dots, \theta_n^{*\top}]^\top \in \mathbb{R}^{q_\theta}$ is the (unknown) optimal parameter (weight) vector that minimizes a suitable cost function between η and $\hat{\eta}$ for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$. Each subvector $\theta_i^* \in \mathbb{R}^{q_{\theta_i}}$, $i = 1, \dots, n$ corresponds to the weights associated with the uncertainty η_i in the state x_i , and q_θ is defined by $q_\theta = \sum_{i=1}^n q_{\theta_i}$. The error term e_η , defined as

$$e_\eta(x(t), u(t)) = \eta(x(t), u(t)) - \hat{\eta}(x(t), u(t); \theta^*), \quad (7.2)$$

represents the (unknown) minimum functional approximation error (MFAE) [Farrell and Polycarpou [2006]], which is the minimum possible deviation between the unknown function η and the adaptive approximator $\hat{\eta}$ in the ∞ -norm sense over the compact set $\mathcal{X} \times \mathcal{U}$, with θ^* defined as:

$$\theta^* = \arg \left\{ \min_{\theta \in \Theta} \left\{ \sup_{x \in \mathcal{X}, u \in \mathcal{U}} |\eta(x, u) - \hat{\eta}(x, u; \theta)| \right\} \right\}. \quad (7.3)$$

It is noted that θ^* is constrained within a compact region $\Theta \subset \mathbb{R}^{q_\theta}$, which will be used by the projection operator later on. The set Θ is selected sufficiently large in order not to undermine the approximation power of $\hat{\eta}$ by increasing the minimum functional approximation error e_η . In the special case of linearly parametrized approximators, $\hat{\eta}$ is of

the form

$$\hat{\eta}(x, u, \theta^*) = G(x, u) \theta^*, \quad (7.4)$$

$$G(x, u) = (g_1(x, u))^T \oplus (g_2(x, u))^T \oplus \dots \oplus (g_n(x, u))^T, \quad (7.5)$$

where $G : \mathbb{R}^n \times \mathbb{R}^\ell \mapsto \mathbb{R}^{n \times q_\theta}$ is the regressor and $g_i : \mathbb{R}^n \times \mathbb{R}^\ell \mapsto \mathbb{R}^{q_{\theta_i}}$, for $i \in \{1, \dots, n\}$ is a bounded function with \oplus denoting the direct sum [Lipschutz and Lipson [2008]]. For simplicity, in this paper we consider linearly parametrized approximators, thus replacing $\hat{\eta}(x, u; \theta^*)$ by $G(x, u)\theta^*$ hereafter.

7.2 Adaptive Nonlinear Estimation Scheme

The adaptive nonlinear estimation scheme consists of a nonlinear observer, a filter for the regressor G and the adaptive law for updating the adjustable parameter vector $\hat{\theta}(t)$:

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + \gamma(\hat{x}(t), u(t)) + L(y(t) - C\hat{x}(t)) + G(\hat{x}(t), u(t))\hat{\theta}(t) \\ &\quad + \Omega(t)\dot{\hat{\theta}}(t), \end{aligned} \quad (7.6)$$

$$\dot{\Omega}(t) = (A - LC)\Omega(t) + G(\hat{x}(t), u(t)), \quad (7.7)$$

$$\dot{\hat{\theta}}(t) = \mathcal{P}_\Theta \left\{ \Gamma (C\Omega(t))^T (y(t) - C\hat{x}(t)) \right\}, \quad (7.8)$$

where $\hat{x} \in \mathbb{R}^n$ denotes the estimation of the state vector x with initial conditions set to $\hat{x}(0) = 0$ following Assumption 4. The matrix $L \in \mathbb{R}^{n \times m}$ is chosen such that the matrix $A - LC$ is Hurwitz (all the eigenvalues are in the left half complex plane). The vector $\hat{\theta}(t) = \left[\hat{\theta}_1(t)^T, \dots, \hat{\theta}_n(t)^T \right]^T \in \mathbb{R}^{q_\theta}$ contains the adjustable parameters of the adaptive approximator, with each subvector $\hat{\theta}_i \in \mathbb{R}^{q_{\theta_i}}$, $i = 1, \dots, n$ corresponding to the parameter estimate of θ_i^* . The initial parameter vector $\hat{\theta}(0)$ and matrix $\Omega(0)$ are chosen as $\hat{\theta}(0) = 0$ and $\Omega(0) = 0$, respectively. In the adaptive law (7.8), $\Gamma \in \mathbb{R}^{q_\theta \times q_\theta}$ is a symmetric, positive definite learning rate matrix, while the projection operator \mathcal{P}_Θ restricts the adjustable parameter vector $\hat{\theta}(t)$ to the predefined set Θ [Farrell and Polycarpou [2006]].

The reason for using the projection operator is to prevent parameter drift in the presence of nonzero minimum functional approximation

error $e_\eta(x, u)$. The matrix Ω is the filtered version of G , necessary for ensuring the stability property of the adaptive nonlinear estimation scheme, while the term $\Omega(t)\hat{\theta}(t)$ is added in order to compensate the estimation error due to $\hat{\theta}$. The idea of the linear filtering of G can also be found in adaptive nonlinear schemes used for fault diagnosis in both time invariant and time varying systems, which can be stochastic or deterministic [Trunov and Polycarpou [2000], Zhang et al. [2002], Xu and Zhang [2004b,a]].

The adaptive nonlinear estimation scheme is applied for the time interval $[0, T_L]$. In general, the training period T_L is a design constant, which should be sufficiently large to allow learning of the modeling uncertainty, while, at the same time, there should be no fault during the training period. Another approach for reducing the modeling error is off-line identification of uncertainties. In general, the adaptive approximation scheme provides a more flexible methodology for learning the uncertainties in the sense that the training time, T_L , can be adjusted on-line based on some criterion involving the estimation error. However, off-line and on-line techniques can be combined for enhancing the learning performance [Caccavale et al. [2009]].

Remark 7.1. In previous work [Zhang et al. [2002], Vemuri and Polycarpou [1997]], a similar adaptive nonlinear estimation scheme was used for estimating the fault function for the special case in which the unknown function depends on the sensor outputs y and not the state x . Such schemes can be used when the unknown nonlinearity η is a function of the measured, noiseless states. When this is not the case, the design and analysis of the adaptive nonlinear estimation scheme becomes more challenging.

7.3 Stability Analysis

The stability of the adaptive nonlinear estimation scheme is analyzed in the following theorem. We define the state estimation error as $\varepsilon_x(t) = x(t) - \hat{x}(t)$, the output estimation error as $\varepsilon_y(t) = y(t) - C\hat{x}(t)$ and the parameter estimation error as $\tilde{\theta}(t) = \hat{\theta}(t) - \theta^*$.

Theorem 7.1. Under Assumptions 1-5, the adaptive nonlinear estimation scheme (7.6)-(7.8) guarantees that:

a) The output estimation error $\varepsilon_y(t)$, the state estimation error $\varepsilon_x(t)$, and the parameter estimation error $\tilde{\theta}(t)$ are uniformly bounded; i.e. $\varepsilon_y(t), \varepsilon_x(t), \tilde{\theta}(t) \in L_\infty$.

b) There exists a positive constant w_1 and a bounded function $v(t)$ (that depends on the MFAE and the uncertainty terms) such that for all finite $t > 0$, the output estimation error satisfies

$$\int_0^t |\varepsilon_y(\tau)|^2 d\tau \leq w_1 + \int_0^t |v(\tau)|^2 d\tau, \quad (7.9)$$

c) In the absence of measurement noise ($d(t) = 0$), if the bounded function v is square integrable, it is implied that $\lim_{t \rightarrow \infty} \varepsilon_y(t) = 0$ and $\lim_{t \rightarrow \infty} \dot{\hat{\theta}}(t) = \lim_{t \rightarrow \infty} \dot{\tilde{\theta}}(t) = 0$.

Proof. a) Using (7.1) and (7.6), the state estimation error dynamics are described by:

$$\begin{aligned} \dot{\varepsilon}_x(t) = & A_0 \varepsilon_x(t) + \tilde{\gamma}(t) + G(x(t), u(t)) \theta^* + e_\eta(x(t), u(t)) \\ & - Ld(t) - G(\hat{x}(t), u(t)) \hat{\theta}(t) - \Omega(t) \dot{\hat{\theta}}(t), \end{aligned} \quad (7.10)$$

where $A_0 = A - LC$ and $\tilde{\gamma}(t) \triangleq \gamma(x(t), u(t)) - \gamma(\hat{x}(t), u(t))$. By adding and subtracting $G(\hat{x}(t), u(t)) \theta^*$ and using $G(\hat{x}(t), u(t)) = \dot{\Omega}(t) - A_0 \Omega(t)$, we obtain

$$\begin{aligned} \dot{\varepsilon}_x(t) = & A_0 \varepsilon_x(t) + \tilde{\gamma}(t) + (G(x(t), u(t)) - G(\hat{x}(t), u(t))) \theta^* \\ & + e_\eta(x(t), u(t)) - Ld(t) \\ & - (\dot{\Omega}(t) - A_0 \Omega(t)) (\hat{\theta}(t) - \theta^*) - \Omega(t) \dot{\hat{\theta}}(t) \end{aligned} \quad (7.11)$$

By letting $\tilde{G}(t) = G(x(t), u(t)) - G(\hat{x}(t), u(t))$ and $\tilde{\theta}(t) = \hat{\theta}(t) - \theta^*$, we obtain

$$\begin{aligned} \dot{\varepsilon}_x(t) = & A_0 (\varepsilon_x(t) + \Omega(t) \tilde{\theta}(t)) + \tilde{\gamma}(t) + \tilde{G}(t) \theta^* + e_\eta(x(t), u(t)) - Ld(t) \\ & - \frac{d}{dt} (\Omega(t) \tilde{\theta}(t)) \end{aligned} \quad (7.12)$$

Let $\tilde{\varepsilon}_x(t) = \varepsilon_x(t) + \Omega(t) \tilde{\theta}(t)$. Then (7.12) can be written as

$$\dot{\tilde{\varepsilon}}_x(t) = A_0 \tilde{\varepsilon}_x(t) + \tilde{\gamma}(t) + \tilde{G}(t) \theta^* + e_\eta(x(t), u(t)) - Ld(t). \quad (7.13)$$

Therefore, given $\Omega(0) = 0$, we obtain

$$\begin{aligned} \varepsilon_x(t) = & -\Omega(t)\tilde{\theta}(t) + e^{A_0 t}\varepsilon_x(0) + \int_0^t e^{A_0(t-\tau)}(\tilde{\gamma}(\tau) - Ld(\tau) \\ & + \tilde{G}(\tau)\theta^* + e_\eta(x(\tau), u(\tau)))d\tau. \end{aligned} \quad (7.14)$$

Since A_0 is Hurwitz, we can choose ρ_0, ξ_0 such that $|e^{A_0 t}| \leq \rho_0 e^{-\xi_0 t}$, for all $t \geq 0$. Therefore, based on Assumptions 1-5, we obtain:

$$|\varepsilon_x(t)| \leq E_0(t) + \rho_0 e^{-\xi_0 t} \lambda_\gamma \int_0^t e^{\xi_0 \tau} |\varepsilon_x(\tau)| d\tau \quad (7.15)$$

$$\begin{aligned} E_0(t) = & |\Omega(t)| |\tilde{\theta}(t)| + \rho_0 e^{-\xi_0 t} |\varepsilon_x(0)| + \int_0^t \rho_0 e^{-\xi_0(t-\tau)} (|L|\bar{d} \\ & + |\tilde{G}(\tau)| |\theta^*| + |e_\eta(x(\tau), u(\tau))|) d\tau \end{aligned} \quad (7.16)$$

and \bar{d} is a bound for $d(t)$ such that $|d(t)| \leq \bar{d}$ for all $t > 0$. Applying the Bellman-Gronwall Lemma [Ioannou and Sun [1995]] results in:

$$|\varepsilon_x(t)| \leq E_0(t) + \rho_0 \lambda_\gamma \int_0^t E_0(s) e^{-(\xi_0 - \rho_0 \lambda_\gamma)(t-s)} ds. \quad (7.17)$$

To guarantee that $\varepsilon_x(t)$ remains bounded, it is required that the constants ξ_0, ρ_0 are selected such that

$$\xi_0 > \rho_0 \lambda_\gamma. \quad (7.18)$$

Taking into account that i) the MFAE e_η and the regressor G are bounded functions in $\mathcal{X} \times \mathcal{U}$ and b) due to the use of parameter projection, $\hat{\theta}(t) \in L_\infty$ and $\tilde{\theta}(t) \in L_\infty$, it yields that the state estimation error is bounded, i.e. $\varepsilon_x(t) \in L_\infty$. The latter implies that the output estimation error $\varepsilon_y(t) = y(t) - C\hat{x}(t)$ is bounded; i.e. $\varepsilon_y(t) \in L_\infty$ since $\varepsilon_y(t) = C\varepsilon_x(t) + d(t)$.

b) Equation (7.14) can be written as

$$\varepsilon_x(t) = v_1(t) + v_2(t) - \Omega(t)\tilde{\theta}(t), \quad (7.19)$$

where $v_1(t)$ and $v_2(t)$ are the solutions of

$$\dot{v}_1(t) = A_0 v_1(t) + \tilde{\gamma}(t) - Ld(t) + \tilde{G}(t)\theta^* + e_\eta(x(t), u(t)), \quad (7.20)$$

$$\dot{v}_2(t) = A_0 v_2(t), \quad (7.21)$$

with initial conditions $v_1(0) = 0$ and $v_2(0) = \tilde{\varepsilon}_x(0) = \varepsilon_x(0)$.

Let the Lyapunov function candidate be $V(t) = \frac{1}{2}\tilde{\theta}(t)^\top \Gamma^{-1}\tilde{\theta}(t) + \int_t^\infty |Cv_2(\tau)|^2 d\tau$. The time derivative of V taking into account (7.8) is given by

$$\begin{aligned}\dot{V}(t) &= \tilde{\theta}(t)^\top \Gamma^{-1}\dot{\tilde{\theta}}(t) - |Cv_2(t)|^2 \\ &= \tilde{\theta}(t)^\top \Gamma^{-1}P_\Theta \left\{ \Gamma(C\Omega(t))^\top \varepsilon_y(t) \right\} - |Cv_2(t)|^2.\end{aligned}\quad (7.22)$$

The projection operator \mathcal{P}_Θ restricts the parameter estimation to a predefined compact set Θ . Since $\theta^* \in \Theta$, the projection operator cannot increase \dot{V} (as compared to the non-projection case) [Farrell and Polycarpou [2006]]. Therefore

$$\begin{aligned}\dot{V}(t) &\leq \tilde{\theta}(t)^\top (C\Omega(t))^\top \varepsilon_y(t) - |Cv_2(t)|^2 \\ &= \varepsilon_y(t)^\top C \left(\Omega(t)\tilde{\theta}(t) \right) - |Cv_2(t)|^2.\end{aligned}\quad (7.23)$$

Using (7.19) and by completing the squares, the time derivative of the Lyapunov function satisfies

$$\begin{aligned}\dot{V}(t) &\leq \varepsilon_y(t)^\top C (-\varepsilon_x(t) + v_1(t) + v_2(t)) - |Cv_2(t)|^2 \\ &\leq - \left(\frac{|\varepsilon_y(t)|^2}{4} - |\varepsilon_y(t)| |Cv_1(t)| + |Cv_1(t)|^2 \right) \\ &\quad - \left(\frac{|\varepsilon_y(t)|^2}{4} - |\varepsilon_y(t)| \bar{d} + \bar{d}^2 \right) \\ &\quad - \left(\frac{|\varepsilon_y(t)|^2}{4} - |\varepsilon_y(t)| |Cv_2(t)| + |Cv_2(t)|^2 \right) \\ &\quad - \frac{|\varepsilon_y(t)|^2}{4} + |Cv_1(t)|^2 + \bar{d}^2 \\ &\leq - \frac{|\varepsilon_y(t)|^2}{4} + |Cv_1(t)|^2 + \bar{d}^2.\end{aligned}\quad (7.24)$$

The above inequality guarantees that $\dot{V}(t) \leq 0$ for $2\left(|Cv_1(t)|^2 + \bar{d}^2\right)^{1/2} \leq |\varepsilon_y(t)|$.

Let $v(t) \triangleq 2\left(|Cv_1(t)|^2 + \bar{d}^2\right)^{1/2}$. Integrating (7.24) from 0 to t , we

obtain

$$\begin{aligned} \int_0^t |\varepsilon_y(\tau)|^2 d\tau &\leq 4[V(0) - V(t)] + \int_0^t |v(\tau)|^2 d\tau \\ &= w_1 + \int_0^t |v(\tau)|^2 d\tau \end{aligned} \quad (7.25)$$

where $w_1 = \sup_{t \geq 0} \{4[V(0) - V(t)]\}$ is a positive constant (since $V(t)$ is uniformly bounded).

c) Since $\varepsilon_y(t)$, $\Omega(t) \in L_\infty$, then based on (7.8), $\dot{\hat{\theta}}(t) \in L_\infty$. Given (7.10) and taking into account the boundedness of ε_x , G , d , e_η , θ^* , $\hat{\theta}$, it yields that $\dot{\varepsilon}_x \in L_\infty$. In the absence of measurement noise, $\varepsilon_y(t) = C\varepsilon_x(t)$, hence $\dot{\varepsilon}_y \in L_\infty$.

If the bounded function v is square integrable, i.e. $v \in L_2$, the outcome of the second part of theorem can be valid for $t \rightarrow \infty$, leading to $\varepsilon_y \in L_2$. Since ε_y , $\dot{\varepsilon}_y \in L_\infty$ and $\varepsilon_y \in L_2$, then according to Barbalat's Lemma, $\lim_{t \rightarrow \infty} \varepsilon_y(t) = 0$ and based on (7.8) $\lim_{t \rightarrow \infty} \dot{\hat{\theta}}(t) = \lim_{t \rightarrow \infty} \dot{\hat{\theta}}(t) = 0$. \square

7.4 Distributed Sensor Fault Detection using Nonlinear Uncertainty Approximators

The outcome of the learning process executed within the time interval $[0, T_L]$, that is the approximated model of the unknown functional nonlinearity $\eta(x(t), u(t))$ is exploited in the sensor fault detection process. Particularly, the estimation model associated with the module $\mathcal{M}^{(q)}$ is generated by a nonlinear observer that is similar to the observer given in (4.49) but it includes also the vector field $\hat{\eta}_L(\hat{x}^{(q)}(t), u(t)) \triangleq G(\hat{x}^{(q)}(t), u(t))\hat{\theta}(T_L)$, which is the nonlinear functional approximator with the parameter vector derived at the end of the learning period $[0, T_L]$. Moreover, the adaptive thresholds are derived following the methodology in Section 4.2.3, but they are designed taking into account the deviation between the approximated model and the actual modeling uncertainty, satisfying the following assumption:

Assumption 7: The error between the modeling uncertainty $\eta(x, u)$ and the functional approximator $\hat{\eta}_L(x, u)$ is bounded for all $x \in \mathcal{X}$,

$u \in \mathcal{U}$ and $t > T_L$ by some known functional $\bar{\eta}$:

$$|\eta(x(t), u(t)) - \hat{\eta}_L(x(t), u(t))| \leq \bar{\eta}(x(t), u(t)), \quad (7.26)$$

where $\bar{\eta}$ is locally Lipschitz in $x \in \mathcal{X}$ for all $u \in \mathcal{U}$ and $t > 0$. The time-varying function $\bar{\eta}$ can be computed by applying a set-membership technique [Milanese and Novara [2004], Yu and de Jesús Rubio [2009]] to dynamical nonlinear systems that satisfy Assumption 2 and can be transformed in or approximated by linear-in-the-parameters (or linear-in-the-weights) models [Kosmatopoulos et al. [1995]].

The generation of residuals and adaptive thresholds using the nonlinear uncertainty approximator is provided in the following sections.

7.4.1 Residual Generation

The nonlinear observer $\mathcal{O}^{(q)}$ is described by:

$$\begin{aligned} \dot{\hat{x}}^{(q)}(t) &= A\hat{x}^{(q)}(t) + \gamma(\hat{x}^{(q)}(t), u(t)) + \hat{\eta}_L(\hat{x}^{(q)}(t), u(t)) \\ &\quad + L^{(q)} \left(y^{(q)}(t) - C^{(q)}\hat{x}^{(q)}(t) \right), \end{aligned} \quad (7.27)$$

where $\hat{x}^{(q)} \in \mathbb{R}^n$ denotes the estimation of the state vector x , with initial conditions $\hat{x}^{(q)}(T_L) = \hat{x}(T_L)$ and $L^{(q)} \in \mathbb{R}^{n \times m_I}$ is the observer gain matrix.

Under healthy conditions, the estimation error dynamics $\varepsilon_{x_H}^{(q)} = x - \hat{x}_H^{(q)}$, where $\hat{x}_H^{(q)}$ is generated by (7.27) taking into account healthy measurements, i.e.

$$\begin{aligned} \dot{\varepsilon}_{x_H}^{(q)}(t) &= A^{(q)}\varepsilon_{x_H}^{(q)}(t) + \gamma(x(t), u(t)) - \gamma(\hat{x}_H^{(q)}(t), u(t)) + \eta(x(t), u(t)) \\ &\quad - \hat{\eta}_L(\hat{x}_H^{(q)}(t), u(t)) - L^{(q)} \left(d^{(q)}(t) + f^{(q)}(t) \right), \end{aligned} \quad (7.28)$$

$$A^{(q)} = A - L^{(q)}C^{(q)}. \quad (7.29)$$

The solution of (7.28) for all $t \geq T_L$ satisfies the following equation:

$$\begin{aligned} \varepsilon_{x_H}^{(q)}(t) &= e^{A^{(q)}(t-T_L)}\varepsilon_{x_H}^{(q)}(T_L) + \int_{T_L}^t e^{A^{(q)}(t-\tau)} \left(\gamma(x(\tau), u(\tau)) \right. \\ &\quad \left. - \gamma(\hat{x}_H^{(q)}(\tau), u(\tau)) + \eta(x(\tau), u(\tau)) - \hat{\eta}_L(\hat{x}_H^{(q)}(\tau), u(\tau)) \right. \\ &\quad \left. - L^{(q)}d^{(q)}(\tau) \right) d\tau \end{aligned} \quad (7.30)$$

Let us define

$$s\Lambda = \lambda_\gamma + \lambda_{\hat{\eta}} + \lambda_{\bar{\eta}} \quad (7.31)$$

where λ_γ , $\lambda_{\hat{\eta}}$, $\lambda_{\bar{\eta}}$ are the Lipschitz constants of γ , $\hat{\eta}_L$ and $\bar{\eta}$, respectively, and \bar{x}_L is a bound for $\varepsilon_x^{(q)}(T_L)$ such that $|\varepsilon_x^{(q)}(T_L)| \leq \bar{x}_L$, for all $x \in \mathcal{X}$. The stability of the error dynamics given in (7.28) for all $t \geq T_L$ is analyzed in Theorem 7.2, taking into account Assumption 7.

Theorem 7.2. Suppose that the observer gain $L^{(q)}$ is chosen such that: (a) the matrix $A_L^{(q)} = A - L^{(q)}C^{(q)}$ is stable, and (b) the positive constants $\rho^{(q)}$, $\xi^{(q)}$ satisfy $\xi^{(q)} > \Lambda\rho^{(q)}$; then the state estimation error under healthy conditions, $\varepsilon_{x_H}^{(q)}(t)$, is uniformly bounded and satisfies

$$|\varepsilon_{x_H}^{(q)}(t)| \leq Z_H^{(q)}(t) \quad (7.32)$$

where

$$Z_H^{(q)}(t) = E_H^{(q)}(t) + \rho^{(q)}\Lambda \int_{T_L}^t E_H^{(q)}(\tau) e^{-\nu^{(q)}(t-\tau)} d\tau, \quad (7.33)$$

$$\begin{aligned} E_H^{(q)}(t) &= \rho^{(q)} e^{-\xi^{(q)}(t-T_L)} \bar{x}_L + \frac{\rho_d^{(q)}}{\xi_d^{(q)}} \bar{d}^{(q)} \left(1 - e^{-\xi^{(q)}(t-T_L)}\right) \\ &\quad + \int_{T_L}^t \rho^{(q)} e^{-\xi^{(q)}(t-\tau)} \bar{\eta}(\hat{x}_H^{(q)}(\tau), u(\tau)) d\tau. \end{aligned} \quad (7.34)$$

Proof. For $t \geq T_L$, the bound of $\varepsilon_x^{(q)}(t)$ given in (7.30), can be determined as:

$$\begin{aligned} |\varepsilon_{x_H}^{(q)}(t)| &\leq \left| e^{A^{(q)}(t-T_L)} \right| |\varepsilon_{x_H}^{(q)}(T_L)| + \int_{T_L}^t \left| e^{A^{(q)}(t-\tau)} L^{(q)} \right| |d^{(q)}(\tau)| d\tau \\ &\quad + \int_{T_L}^t \left| e^{A^{(q)}(t-\tau)} \right| \left(\left| \gamma(x(\tau), u(\tau)) - \gamma(\hat{x}_H^{(q)}(\tau), u(\tau)) \right| \right. \\ &\quad \left. + \left| \eta(x(\tau), u(\tau)) - \hat{\eta}_L(\hat{x}_H^{(q)}(\tau), u(\tau)) \right| \right) d\tau \end{aligned} \quad (7.35)$$

By choosing $\rho^{(q)}$, $\xi^{(q)} > 0$ such that $|e^{A^{(q)}t}| \leq \rho^{(q)} e^{-\xi^{(q)}t}$ for $t > 0$ and

$\rho_d^{(q)}, \xi_d^{(q)} > 0$ such that $|e^{A^{(q)}t}L^{(q)}| \leq \rho_d^{(q)}e^{-\xi_d^{(q)}t}$, for $t > 0$ we obtain

$$\begin{aligned} |\varepsilon_{x_H}^{(q)}(t)| &\leq \rho^{(q)}e^{-\xi^{(q)}(t-T_L)}\bar{x}_L + \frac{\rho_d^{(q)}}{\xi_d^{(q)}}\bar{d}^{(q)}\left(1 - e^{-\xi^{(q)}(t-T_L)}\right) \\ &\quad + \int_{T_L}^t \rho^{(q)}e^{-\xi^{(q)}(t-\tau)}\left(\lambda_\gamma|\varepsilon_{x_H}^{(q)}(\tau)| + |\tilde{\eta}^{(q)}(\tau)|\right)d\tau \end{aligned} \quad (7.36)$$

where

$$\tilde{\eta}^{(q)}(t) = \eta(x(t), u(t)) - \hat{\eta}_L(\hat{x}_H^{(q)}(t), u(t)). \quad (7.37)$$

By adding and subtracting $\hat{\eta}_L(x(t), u(t))$ in (7.37) and using (7.26), a bound for $|\tilde{\eta}^{(q)}(t)|$ can be computed as follows:

$$|\tilde{\eta}^{(q)}(t)| \leq \bar{\eta}(x(t), u(t)) + |\hat{\eta}_L(x(t), u(t)) - \hat{\eta}_L(\hat{x}_H^{(q)}(t), u(t))| \quad (7.38)$$

Assuming that the functional approximator $\hat{\eta}_L$ is locally Lipschitz in $x \in \mathcal{X}$, for all $u \in \mathcal{U}$ and $t > 0$ [Uykan et al. [2000]], i.e. $|\hat{\eta}_L(x(t), u(t)) - \hat{\eta}_L(\hat{x}_H^{(q)}(t), u(t))| \leq \lambda_{\hat{\eta}}|\varepsilon_{x_H}^{(q)}(t)|$, where $\lambda_{\hat{\eta}}$ is the known Lipschitz constant, we have

$$|\tilde{\eta}^{(q)}(t)| \leq \bar{\eta}(x(t), u(t)) + \lambda_{\hat{\eta}}|\varepsilon_{x_H}^{(q)}(t)|. \quad (7.39)$$

By adding and subtracting $\bar{\eta}(\hat{x}_H^{(q)}(t), u(t))$, inequality (7.39) can be rewritten as:

$$\begin{aligned} |\tilde{\eta}^{(q)}(t)| &\leq \bar{\eta}(\hat{x}_H^{(q)}(t), u(t)) + |\bar{\eta}(x(t), u(t)) - \bar{\eta}(\hat{x}_H^{(q)}(t), u(t))| \\ &\quad + \lambda_{\hat{\eta}}|\varepsilon_{x_H}^{(q)}(t)| \end{aligned} \quad (7.40)$$

As stated in Assumption 7, $\bar{\eta}$ is locally Lipschitz, i.e. $|\bar{\eta}(x(t), u(t)) - \bar{\eta}(\hat{x}_H^{(q)}(t), u(t))| \leq \lambda_{\bar{\eta}}|\varepsilon_{x_H}^{(q)}(t)|$, where $\lambda_{\bar{\eta}}$ is the known Lipschitz constant. Thus, it yields

$$|\tilde{\eta}^{(q)}(t)| \leq \bar{\eta}(\hat{x}_H^{(q)}(t), u(t)) + (\lambda_{\hat{\eta}} + \lambda_{\bar{\eta}})|\varepsilon_{x_H}^{(q)}(t)|. \quad (7.41)$$

Using (7.41) in (7.36) results in

$$|\varepsilon_{x_H}^{(q)}(t)| \leq E^{(q)}(t) + \int_{T_L}^t \rho^{(q)}e^{-\xi^{(q)}(t-\tau)}\Lambda|\varepsilon_{x_H}^{(q)}(\tau)|d\tau \quad (7.42)$$

where $E_H^{(q)}$ is defined in (7.34). By applying the Bellman-Gronwall Lemma [Ioannou and Sun [1995]], we derive (7.32) and (7.33). \square

Under healthy conditions, the j -th residual is described by:

$$\varepsilon_{y_{jH}}^{(q)}(t) = C_j \varepsilon_{x_H}^{(q)}(t) + d_j(t), \quad j \in \mathcal{J}^{(q)} \quad (7.43)$$

7.4.2 Computation of Adaptive Thresholds

Under healthy conditions, the bound of $\varepsilon_{y_{jH}}^{(q)}(t)$, $j \in \mathcal{J}^{(q)}$, defined in (7.43) using (7.30) can be determined as:

$$\begin{aligned} \left| \varepsilon_{y_{jH}}^{(q)}(t) \right| &\leq \left| C_j e^{A^{(q)}(t-T_L)} \right| \left| \varepsilon_x^{(q)}(T_L) \right| \\ &\quad + \int_{T_L}^t \left| C_j e^{A^{(q)}(t-\tau)} \right| \left(\left| \tilde{\gamma}^{(q)}(\tau) \right| + \left| \tilde{\eta}^{(q)}(\tau) \right| \right) d\tau \\ &\quad + \int_{T_L}^t \left| C_j e^{A^{(q)}(t-\tau)} L^{(q)} \right| \left| d^{(q)}(\tau) \right| d\tau + |d_j(t)| \end{aligned} \quad (7.44)$$

where $\tilde{\gamma}^{(q)}(t) = \gamma(x(t), u(t)) - \gamma(x_H^{(q)}(t), u(t))$ and $\tilde{\eta}$ is defined in (7.37). Consider that : i) $\alpha_j^{(q)}$, $\zeta_j^{(q)}$ are positive constants chosen such that $\left| C_j e^{A^{(q)}t} \right| \leq \alpha_j^{(q)} e^{-\zeta_j^{(q)}t}$ for $t > 0$ and ii) $\alpha_{d_j}^{(q)}$, $\zeta_{d_j}^{(q)}$ are positive constants chosen such that $\left| C_j e^{A^{(q)}t} L^{(q)} \right| \leq \alpha_{d_j}^{(q)} e^{-\zeta_{d_j}^{(q)}t}$, for $t > 0$. Then, based on Assumptions 1-3 and given (7.41), (7.44) is described by:

$$\begin{aligned} \left| \varepsilon_{y_{jH}}^{(q)}(t) \right| &\leq \int_{T_L}^t \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-\tau)} \Lambda \left| \varepsilon_x^{(q)}(\tau) \right| d\tau \\ &\quad + \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-T_L)} \bar{x}_L + \frac{\alpha_{d_j}^{(q)}}{\zeta_{d_j}^{(q)}} \bar{d}^{(q)} \left(1 - e^{-\zeta_{d_j}^{(q)}(t-T_L)} \right) + \bar{d}_j \\ &\quad + \int_{T_L}^t \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-\tau)} \tilde{\eta}(\hat{x}_H^{(q)}(\tau), u(\tau)) d\tau. \end{aligned} \quad (7.45)$$

Taking into account healthy conditions (7.45), the adaptive threshold is defined in (4.79).

$$\bar{\varepsilon}_{y_j}^{(q)}(t) = \int_{T_L}^t \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-\tau)} \left(\Lambda Z^{(q)}(\tau) + \tilde{\eta}(\hat{x}^{(q)}(\tau), u(\tau)) \right) d\tau + Y_j^{(q)}(t), \quad (7.46)$$

$$Y_j^{(q)}(t) = \alpha_j^{(q)} e^{-\zeta_j^{(q)}(t-T_L)} \bar{x}_L + \frac{\alpha_{d_j}^{(q)}}{\zeta_{d_j}^{(q)}} \bar{d}^{(q)} \left(1 - e^{-\zeta_{d_j}^{(q)}(t-T_L)} \right) + \bar{d}_j \quad (7.47)$$

7.4.3 Distributed Fault Diagnosis Decision Logic

Using the learned approximator of the modeling uncertainty can enhance the fault detectability of the distributed fault diagnosis scheme, but it may complicate its fault isolability. The reason for this is that the adaptive threshold $\bar{\varepsilon}_j^{(q)}$ is computed taking into account a filtered version of the bound on the deviation of the approximation model η_L from the modeling uncertainty η under healthy conditions, i.e. the bound $\bar{\eta}$. However, under faulty conditions the approximation model η_L , and consequently the bound $\bar{\eta}$ may not be valid. More specifically, for the computation of the adaptive thresholds we consider Assumption 7 and (7.26) for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$, where \mathcal{X} and \mathcal{U} are compact regions, which are selected for the adaptive learning of the modeling uncertainty under healthy conditions, as shown in Section 7.2. The domain of the approximation $\mathcal{X} \times \mathcal{U}$ may change under faulty conditions, if the faulty sensor measurements are used by the controller to generate the control input u . In this case, the approximation model η_L and the bound $\bar{\eta}$ in (7.26) may not be valid, as well as the filtered version of the bound $\bar{\eta}$ used in the computation of the adaptive thresholds.

The fault signature matrix F , based on which the faulty sensors are isolated, is designed such that the q -th element, $q \in \{1, \dots, N\}$, of the column F_i , defined as $F_i = [F_{1i}, \dots, F_{Ni}]^\top$, is equal to 1, i.e. $F_{qi} = 1$ if at least one sensor fault that belongs to the combination \mathcal{F}_{c_i} , $i \in \{1, \dots, N_c\}$ affects the local sensor set $\mathcal{S}^{(q)}$ and can provoke the violation of (or else is involved in) $\mathcal{E}^{(q)}$, and $F_{qi} = *$ if at least one sensor fault that belongs to the combination \mathcal{F}_{c_i} , $i \in \{1, \dots, N_c\}$ may provoke the violation of condition (7.26) and consequently the violation of $\mathcal{E}^{(q)}$, but none of the sensor faults in \mathcal{F}_{c_i} affects the local sensor set $\mathcal{S}^{(q)}$. To exemplify this, let us consider the Example 5.1 for $I = 1$; the distributed sensor fault diagnosis scheme is similar to the scheme presented in Fig. 5.1. By using an approximator η_L in every local monitoring module, the fault isolation decision logic relies on the fault signature matrix shown in Table 7.1. The equation $F_{qi} = *$ means that the set $\mathcal{E}^{(q)}$, $q \in \{1, \dots, N\}$, may be violated when the bound functional $\bar{\eta}(x, u)$ is not valid due to the occurrence of a fault in a sensor that does not belong to the local set $\mathcal{S}^{(q)}$ but can cause the

	\mathcal{F}_{c_1}	\mathcal{F}_{c_2}	\mathcal{F}_{c_3}	\mathcal{F}_{c_4}	\mathcal{F}_{c_5}	\mathcal{F}_{c_6}	\mathcal{F}_{c_7}
$\mathcal{E}^{(1)}$	1	*	*	1	1	*	1
$\mathcal{E}^{(2)}$	*	1	1	1	1	1	1
$\mathcal{E}^{(3)}$	*	*	1	*	1	1	1

Table 7.1: Multiple sensor fault signature matrix F for a distributed sensor fault diagnosis scheme.

deviation of x and u from the desired domain of approximation $\mathcal{X} \times \mathcal{U}$.

It is worth to note that the guaranteed validity of the domain of approximation $\mathcal{X} \times \mathcal{U}$, within which the approximation model η_L is trained under *any* conditions is an open, challenging issue. It is always possible to select a domain of approximation large enough to be valid for any condition, but then the approximation efficiency will be reduced, as well as the benefit of using the approximation model of the modeling uncertainty. If some sensor measurements are not used for control but only for monitoring, then a possible fault affecting them cannot cause the deviation of the domain of approximation. In this case, the fault signature matrix will include this information by replacing $F_{qi} = *$ with $F_{qi} = 0$ in case that all the sensor faults in the combination \mathcal{F}_{c_i} , $i \in \{1, \dots, Nc\}$, are used only for monitoring and are not involved in the set $\mathcal{E}^{(q)}$.

7.5 Illustrative Example

The objective of this section is to illustrate the use of the proposed adaptive approximation technique in the design of a distributed sensor fault diagnosis method applied to the interconnected nonlinear subsystems presented in Example 2.2. Let us consider that the interconnection dynamics of subsystem $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are respectively expressed as $h^{(1)}(z^{(1)}) = b^{(1)} \sin(z^{(1)})$ and $h^{(2)}(z^{(2)}) = b^{(2)} \cos(z^{(2)})$, while the modeling uncertainty $\eta_2^{(1)}$ and $\eta_2^{(2)}$ are nonlinear functions of $x_1^{(1)}$ and $x_1^{(2)}$, which can be modeled as:

$$\eta_2^{(1)}(x_1^{(1)}) = G^{(1)}(x_1^{(1)})\theta^{(1)} + e_\eta^{(1)}(x_1^{(1)}) \quad (7.48)$$

$$\eta_2^{(2)}(x_1^{(2)}) = G^{(2)}(x_1^{(2)})\theta^{(2)} + e_\eta^{(2)}(x_1^{(2)}) \quad (7.49)$$

where $\theta^{(1)} \in \mathbb{R}^{q_{\theta_1}}$, $\theta^{(2)} \in \mathbb{R}^{q_{\theta_2}}$ are the unknown optimal parameter vectors that respectively minimize the unknown minimum functional approximation errors $e_{\eta}^{(1)} \in \mathbb{R}$ and $e_{\eta}^{(2)} \in \mathbb{R}$, and $G^{(1)} \in \mathbb{R}^{1 \times q_{\theta_1}}$, $G^{(2)} \in \mathbb{R}^{1 \times q_{\theta_2}}$ are the regressors. Using (7.48) and (7.49), $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are described by

$$\begin{aligned} \Sigma^{(1)} : \dot{x}^{(1)} = & A^{(1)}x^{(1)} + Q_1(g_1^{(1)}u^{(1)}) + Q_1(g_2^{(1)}\cos(x_1^{(1)})) \\ & + Q_1(b^{(1)}\sin(z^{(1)})) + Q_1(G^{(1)}(x_1^{(1)})\theta^{(1)} + e_{\eta}^{(1)}(x_1^{(1)})), \end{aligned} \quad (7.50)$$

$$\begin{aligned} \Sigma^{(2)} : \dot{x}^{(2)} = & A^{(2)}x^{(2)} + Q_1(g_1^{(2)}u^{(2)}) + Q_1(g_2^{(2)}\sin(x_1^{(2)})) \\ & + Q_1(b^{(2)}\cos(z^{(2)})) + Q_1(G^{(2)}(x_1^{(2)})\theta^{(2)} + e_{\eta}^{(2)}(x_1^{(2)})), \end{aligned} \quad (7.51)$$

where $z^{(1)} = x_1^{(2)}$, $z^{(2)} = x_1^{(1)}$, $A^{(1)}$ and $A^{(2)}$ are defined through (2.10), and $Q_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$. Based on (7.48)-(7.51), an adaptive nonlinear estimation scheme is designed for each subsystem following a distributed approach as shown next.

7.5.1 Distributed Adaptive Nonlinear Estimation Scheme

The adaptive nonlinear observer that generates the estimation model of $\Sigma^{(1)}$ is defined as:

$$\begin{aligned} \dot{\hat{x}}_H^{(1)} = & A^{(1)}\hat{x}_H^{(1)} + Q_1(g_1^{(1)}u^{(1)}) + Q_1(g_2^{(1)}\cos(\hat{x}_{H1}^{(1)})) + Q_1(b^{(1)}\sin(y_1^{(2)})) \\ & + L^{(1)}(y^{(1)} - C^{(1)}\hat{x}_H^{(1)}) + Q_1G^{(1)}(\hat{x}_{H1}^{(1)})\hat{\theta}^{(1)} + \Omega^{(1)}\dot{\hat{\theta}}^{(1)} \end{aligned} \quad (7.52)$$

$$\dot{\Omega}^{(1)} = (A^{(1)} - L^{(1)}C^{(1)})\Omega^{(1)} + Q_1G^{(1)}(\hat{x}_{H1}^{(1)}) \quad (7.53)$$

$$\dot{\hat{\theta}}^{(1)} = \mathcal{P}_{\Theta}^{(1)} \left\{ \Gamma^{(1)} (C^{(1)}\Omega^{(1)})^T (y^{(1)} - C^{(1)}\hat{x}_H^{(1)}) \right\} \quad (7.54)$$

where $\hat{x}_H^{(1)} \in \mathbb{R}^2$ denotes the estimation of the state vector $x^{(1)}$ (using the measurements $y^{(1)} = [y_1^{(1)}, y_2^{(1)}]^T$ with initial conditions set to $\hat{x}^{(1)}(0) = 0$) following Assumption 6 stated in Section 7.1. The matrix $L^{(1)} \in \mathbb{R}^{2 \times 2}$ is chosen such that the matrix $A^{(1)} - L^{(1)}C^{(1)}$ is stable. The vector $\hat{\theta}^{(1)}(t) \in \mathbb{R}^{q_{\theta_1}}$ contains the adjustable parameters of the adaptive

approximator. The initial parameter vector $\hat{\theta}^{(1)}(0)$ and matrix $\Omega^{(1)}(0)$ are chosen as $\hat{\theta}^{(1)}(0) = 0$ and $\Omega^{(1)}(0) = 0$, respectively. In the adaptive law (7.54), $\Gamma^{(1)} \in \mathbb{R}^{q_{\theta_1} \times q_{\theta_1}}$ is a symmetric, positive definite learning rate matrix, and $\mathcal{P}_{\Theta}^{(1)}$ is the projection operator. Taking into account (7.50) and (7.52), the estimation error $\varepsilon_{x_H}^{(1)} = x^{(1)} - \hat{x}_H^{(1)}$ under healthy conditions is described by

$$\begin{aligned} \dot{\varepsilon}_{x_H}^{(1)} = & A_L^{(1)} \varepsilon_{x_H}^{(1)} + Q_1 g_2^{(1)} \left(\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1)}) \right) \\ & + b^{(1)} \left(\sin(x_1^{(2)}) - \sin(y_1^{(2)}) \right) + Q_1 e_{\eta}^{(1)}(x_1^{(1)}) - \Omega^{(1)}(t) \dot{\hat{\theta}}^{(1)}(t) \\ & + Q_1 \left(G^{(1)}(x_1^{(1)}) \theta^{(1)} - G^{(1)}(\hat{x}_{H1}^{(1)}) \hat{\theta}^{(1)}(t) \right) - L^{(1)} d^{(1)}, \end{aligned} \quad (7.55)$$

$$A_L^{(1)} = A^{(1)} - L^{(1)} C^{(1)} \quad (7.56)$$

By adding and subtracting $G^{(1)}(\hat{x}_{H1}^{(1)}) \theta^{(1)}$ and using $Q_1 G^{(1)}(\hat{x}_{H1}^{(1)}) = \dot{\Omega}^{(1)} - A^{(1)} \Omega^{(1)}$, we have

$$\begin{aligned} \dot{\varepsilon}_{x_H}^{(1)} = & A_L^{(1)} \varepsilon_{x_H}^{(1)} + Q_1 g_2^{(1)} \left(\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1)}) \right) \\ & + b^{(1)} \left(\cos(x_1^{(2)}) - \cos(y_1^{(2)}) \right) + Q_1 e_{\eta}^{(1)}(x_1^{(1)}) - \Omega^{(1)}(t) \dot{\hat{\theta}}^{(1)}(t) \\ & + \left(\dot{\Omega}^{(1)} - A^{(1)} \Omega^{(1)} \right) \left(\theta^{(1)} - \hat{\theta}^{(1)}(t) \right) - L^{(1)} d^{(1)} \\ & + Q_1 \left(G^{(1)}(x_1^{(1)}) - G^{(1)}(\hat{x}_{H1}^{(1)}) \right) \theta^{(1)}, \end{aligned} \quad (7.57)$$

Let us define

$$\tilde{\theta}^{(1)}(t) = \theta^{(1)} - \hat{\theta}^{(1)}(t), \quad (7.58)$$

$$\tilde{G}^{(1)}(t) = G^{(1)}(x_1^{(1)}) - G^{(1)}(\hat{x}_{H1}^{(1)}), \quad (7.59)$$

$$\tilde{\varepsilon}_{x_H}^{(1)}(t) = \varepsilon_{x_H}^{(1)}(t) - \Omega^{(1)}(t) \tilde{\theta}^{(1)}(t) \quad (7.60)$$

Then, (7.57) can be re-written as

$$\begin{aligned} \dot{\tilde{\varepsilon}}_{x_H}^{(1)} = & A_L^{(1)} \tilde{\varepsilon}_{x_H}^{(1)} + Q_1 g_2^{(1)} \left(\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1)}) \right) - L^{(1)} d^{(1)} + Q_1 \tilde{G}^{(1)} \theta^{(1)} \\ & + b^{(1)} \left(\cos(x_1^{(2)}) - \cos(y_1^{(2)}) \right) + Q_1 e_{\eta}^{(1)}(x_1^{(1)}), \end{aligned} \quad (7.61)$$

Following the analysis presented in Section 7.3 in order to find a bound on the solution of (7.61) we obtain

$$\begin{aligned}
\left| \varepsilon_{x_H}^{(1)} \right| &\leq \left| \Omega^{(1)} \right| \left| \tilde{\theta}^{(1)} \right| + \left| e^{A_L^{(1)} t} \right| \left| \varepsilon_{x_H}^{(1)}(0) \right| + \int_0^t \left| e^{A_L^{(1)}(t-\tau)} L^{(1)} \right| \left| \bar{d}^{(1)} \right| d\tau \\
&\quad + \int_0^t \left| e^{A_L^{(1)}(t-\tau)} Q_1 \right| \left(\left| g_2^{(1)} \right| \left| \varepsilon_{x_{H1}}^{(1)} \right| + \left| b^{(1)} \right| \left| \bar{d}_1^{(2)} \right| \right) d\tau \\
&\quad + \int_0^t \left| e^{A_L^{(1)}(t-\tau)} Q_1 \right| \left(\left| e_\eta^{(1)}(x_1^{(1)}) \right| + \left| \tilde{G}^{(1)} \right| \left| \theta^{(1)} \right| \right) d\tau \quad (7.62)
\end{aligned}$$

In order to obtain a bound for $\left| \varepsilon_{x_H}^{(1)} \right|$, we need to compute a bound on $\left| \varepsilon_{x_{H1}}^{(1)} \right|$. Taking into account (4.187), it yields

$$\begin{aligned}
\left| \varepsilon_{x_{H1}}^{(1)} \right| &\leq \left| Q_2 \Omega^{(1)} \right| \left| \tilde{\theta}^{(1)} \right| + \left| Q_2 e^{A_L^{(1)} t} \right| \left| \varepsilon_{x_H}^{(1)}(0) \right| + \int_0^t \left| Q_2 e^{A_L^{(1)}(t-\tau)} L^{(1)} \right| \left| \bar{d}^{(1)} \right| d\tau \\
&\quad + \int_0^t \left| Q_2 e^{A_L^{(1)}(t-\tau)} Q_1 \right| \left(\left| g_2^{(1)} \right| \left| \varepsilon_{x_{H1}}^{(1)} \right| + \left| b^{(1)} \right| \left| \bar{d}_1^{(2)} \right| \right) d\tau \\
&\quad + \int_0^t \left| Q_2 e^{A_L^{(1)}(t-\tau)} Q_1 \right| \left(\left| e_\eta^{(1)}(x_1^{(1)}) \right| + \left| \tilde{G}^{(1)} \right| \left| \theta^{(1)} \right| \right) d\tau \quad (7.63)
\end{aligned}$$

By choosing $L^{(1)}$ such that $A_L^{(1)}$ is stable and positive constants $\rho_0^{(1)}$, $\xi_0^{(1)}$, $\rho^{(1)}$, $\xi^{(1)}$, $\rho_d^{(1)}$, $\xi_d^{(1)}$ such that

$$\left| Q_2 e^{A_L^{(1)} t} \right| \leq \rho_0^{(1)} e^{-\xi_0^{(1)} t}, \quad (7.64)$$

$$\left| Q_2 e^{A_L^{(1)} t} Q_1 \right| \leq \rho^{(1)} e^{-\xi^{(1)} t} \quad (7.65)$$

$$\rho^{(1)} |g_2^{(1)}| < \xi^{(1)} \quad (7.66)$$

$$\left| Q_2 e^{A_L^{(1)} t} L^{(1)} \right| \leq \rho_d^{(1)} e^{-\xi_d^{(1)} t}, \quad (7.67)$$

and using the Bellman-Gronwall Lemma, it is ensured that $\left| \varepsilon_{x_{H1}}^{(1)} \right|$ is bounded by a finite function, that is

$$\left| \varepsilon_{x_{H1}}^{(1,q)} \right| \leq \rho^{(1)} |g_2^{(1)}| \int_0^t e^{-\left(\xi^{(1)} - \rho^{(1)} |g_2^{(1)}| \right) (t-\tau)} E_H^{(1)} d\tau + E_H^{(1)} \quad (7.68)$$

with

$$\begin{aligned} E_H^{(1)} = & \left| Q_2 \Omega^{(1)} \right| \left| \tilde{\theta}^{(1)} \right| + \rho_0^{(1)} e^{-\xi_0^{(1)} t} \left| \varepsilon_{x_H}^{(1)}(0) \right| + \int_0^t \rho_d^{(1)} e^{-\xi_d^{(1)}(t-\tau)} \bar{d}^{(1)} d\tau \\ & + \int_0^t \rho^{(1)} e^{-\xi^{(1)}(t-\tau)} \left(\left| b^{(1)} \right| \bar{d}_1^{(2)} + \left| e_\eta^{(1)}(x_1^{(1)}) \right| + \left| \tilde{G}^{(1)} \right| \left| \theta^{(1)} \right| \right) d\tau \end{aligned} \quad (7.69)$$

where $\left| \tilde{\theta}^{(1)} \right|$ is bounded due to the use of the projection operator $\mathcal{P}_\Theta^{(1)}$, while $\tilde{G}^{(1)}$ and $e_\eta^{(1)}(x_1^{(1)})$ are bounded in \mathcal{X}_1 .

Similarly, the adaptive nonlinear observer that generates the estimation model of $\Sigma^{(2)}$ is defined as:

$$\begin{aligned} \dot{\hat{x}}_H^{(2)} = & A^{(2)} \hat{x}_H^{(2)} + Q_1 \left(g_1^{(2)} u^{(2)} \right) + Q_1 \left(g_2^{(2)} \sin(\hat{x}_{H1}^{(2)}) \right) + Q_1 \left(b^{(2)} \cos(y_1^{(1)}) \right) \\ & + L^{(2)} \left(y^{(2)} - C^{(2)} \hat{x}_H^{(2)} \right) + Q_1 G^{(2)} \left(\hat{x}_{H1}^{(2)} \right) \hat{\theta}^{(2)} + \Omega^{(2)} \dot{\hat{\theta}}^{(2)} \end{aligned} \quad (7.70)$$

$$\dot{\Omega}^{(2)} = (A^{(2)} - L^{(2)} C^{(2)}) \Omega^{(2)} + Q_1 G^{(2)} \left(\hat{x}_{H1}^{(2)} \right) \quad (7.71)$$

$$\dot{\hat{\theta}}^{(2)} = \mathcal{P}_\Theta^{(2)} \left\{ \Gamma^{(2)} \left(C^{(2)} \Omega^{(2)}(t) \right)^\top \left(y^{(2)} - C^{(2)} \hat{x}_H^{(2)} \right) \right\} \quad (7.72)$$

where $\hat{x}_H^{(2)} \in \mathbb{R}^2$ denotes the estimation of the state vector $x^{(2)}$ (using the measurements $y^{(2)} = [y_1^{(2)}, y_2^{(2)}]^\top$ with initial conditions set to $\hat{x}^{(2)}(0) = 0$) taking into account Assumption 6 stated in Section 7.1. The matrix $L^{(2)} \in \mathbb{R}^{2 \times 2}$ is chosen such that the matrix $A^{(2)} - L^{(2)} C^{(2)}$ is stable. The vector $\hat{\theta}^{(2)}(t) \in \mathbb{R}^{q_{\theta_2}}$ contains the adjustable parameters of the adaptive approximator. The initial parameter vector $\hat{\theta}^{(2)}(0)$ and matrix $\Omega^{(2)}(0)$ are chosen as $\hat{\theta}^{(2)}(0) = 0$ and $\Omega^{(2)}(0) = 0$, respectively. In the adaptive law (7.54), $\Gamma^{(2)} \in \mathbb{R}^{q_{\theta_2} \times q_{\theta_2}}$ is a symmetric, positive definite learning rate matrix, and $\mathcal{P}_\Theta^{(2)}$ is the projection operator. The stability of the estimation scheme (7.70)-(7.72) is ensured by choosing $L^{(2)}$ such that $A_L^{(2)}$ is stable and there are positive constants $\rho^{(2)}$, $\xi^{(2)}$, such that

$$\left| Q_2 e^{A_L^{(2)} t} Q_1 \right| \leq \rho^{(2)} e^{-\xi^{(2)} t} \quad (7.73)$$

$$\rho^{(2)} |g_2^{(2)}| < \xi^{(2)} \quad (7.74)$$

The learning time for the I -th adaptive estimation scheme, denoted by $T_L^{(I)}$, $I = 1, 2$, is defined when the magnitude of the q -th output estimation error $\varepsilon_{y_q}^{(I)}$, $q = 1, 2$ is below a certain threshold.

Remark 7.2. In this simulation example, the projection operator $\mathcal{P}_\Theta^{(I)}$ is used to restrict $\hat{\theta}^{(I)}$ in the set $\{\hat{\theta} \in \mathbb{R}^{q_{\theta I}} : \hat{\theta}^\top \hat{\theta} \leq M_{\theta_I}^2\}$. Thus the adaptive laws in (7.54) and (7.72) are implemented as in [Trunov and Polycarpou [2000], Farrell and Polycarpou [2006]]; i.e.,

$$\dot{\hat{\theta}}^{(I)} = \Gamma^{(I)} \left(C^{(I)} \Omega^{(I)} \right)^\top \varepsilon_y^{(I)} - \chi^{(I)} \Gamma^{(I)} \frac{\hat{\theta}^{(I)} \left(\hat{\theta}^{(I)} \right)^\top}{\left(\hat{\theta}^{(I)} \right)^\top \Gamma^{(I)} \hat{\theta}^{(I)}} \Gamma^{(I)} \left(C^{(I)} \Omega^{(I)} \right)^\top \varepsilon_y^{(I)} \quad (7.75)$$

where $\varepsilon_y^{(I)} = y^{(I)} - C^{(I)} \hat{x}_H^{(I)}$ and $\chi^{(I)} = 0$ if

$$\left| \hat{\theta}^{(I)} \right| \leq M_{\theta_I}, \quad (7.76)$$

or

$$\left| \hat{\theta}^{(I)} \right| = M_{\theta_I} \ \& \ \left(\hat{\theta}^{(I)} \right)^\top \Gamma^{(I)} \left(C^{(I)} \Omega^{(I)} \right)^\top \left(y^{(I)} - C^{(I)} \hat{x}_H^{(I)} \right) \leq 0, \quad (7.77)$$

and $\chi^{(I)} = 1$ if

$$\left| \hat{\theta}^{(I)} \right| = M_{\theta_I} \ \& \ \left(\hat{\theta}^{(I)} \right)^\top \Gamma^{(I)} \left(C^{(I)} \Omega^{(I)} \right)^\top \left(y^{(I)} - C^{(I)} \hat{x}_H^{(I)} \right) > 0 \quad (7.78)$$

7.5.2 Distributed Sensor Fault Diagnosis Scheme based on Nonlinear Uncertainty Approximators

The nonlinear functional approximator of the modeling uncertainty $\eta_2^{(I)}$, $I = 1, 2$, denoted by $\hat{\eta}_L^{(I)}$, is obtained using the estimated parameter vector computed in the end of the learning period, that is

$$\hat{\eta}_L^{(I)}(x) = G^{(I)}(x) \hat{\theta}^{(I)}(T_L^{(I)}). \quad (7.79)$$

The approximator $\hat{\eta}_L^{(I)}$ is used in the observer-based residual generation and the computation of adaptive thresholds of the monitoring modules $\mathcal{M}^{(I,q)}$, $q = 1, 2$ presented in Section 4.5.1 and 4.5.2. Particularly, the distributed observer presented in (4.184) with $z_\star^{(1)} = y_1^{(2)}$ is properly

customized to exploit the nonlinear uncertainty approximator defined in (7.79); i.e.

$$\begin{aligned}\dot{\hat{x}}^{(1,q)} = & A^{(1)}\hat{x}^{(1,q)} + Q_1\left(g_1^{(1)}u^{(1)}\right) + Q_1\left(g_2^{(1)}\cos(\hat{x}_1^{(1,q)})\right) + \hat{\eta}_L^{(1)}(\hat{x}^{(1)}) \\ & + Q_1\left(b^{(1)}\sin(y_1^{(2)})\right) + L^{(1,q)}\left(y_q^{(1)} - C_q^{(1)}\hat{x}^{(1,q)}\right)\end{aligned}\quad (7.80)$$

where $\hat{x}^{(1,q)}(T_L^{(1)}) = \hat{x}^{(1)}(T_L^{(1)})$. The state estimation error $\varepsilon_{x_H}^{(1,q)} = x^{(1)}(t) - \hat{x}_H^{(1,q)}(t)$, where $\hat{x}_H^{(1,q)}(t)$ is the estimation model under healthy conditions for $t \geq T_L^{(1)}$, is described by

$$\begin{aligned}\varepsilon_{x_H}^{(1,q)} = & A_L^{(1,q)}\varepsilon_{x_H}^{(1,q)} + Q_1g_2^{(1)}\left(\cos(x_1^{(1)}) - \cos(\hat{x}_{H1}^{(1,q)})\right) \\ & + b^{(1)}\left(\sin(x_1^{(2)}) - \sin(y_1^{(2)})\right) + Q_1\left(\eta_2^{(1)}(x_1^{(1)}) - \hat{\eta}_L^{(1)}(\hat{x}_{H1}^{(1,q)})\right) \\ & - L^{(1,q)}d_q^{(1)}\end{aligned}\quad (7.81)$$

In this illustrative example, a worst case bound on the difference $\eta_2^{(1)}(x_1^{(1)}) - \hat{\eta}_L^{(1)}(\hat{x}_{H1}^{(1,q)})$ for all $t \geq T_L^{(1)}$ is used; i.e.

$$\left|\eta_2^{(1)}(x_1^{(1)}(t)) - \hat{\eta}_L^{(1)}(\hat{x}_{H1}^{(1,q)}(t))\right| \leq \bar{\eta}_L^{(1)}, \quad (7.82)$$

where $\bar{\eta}_L^{(1)}$ is known. As discussed in Section 4.5.1, in order for $\varepsilon_{x_H}^{(1,q)}$ to be bounded, $\varepsilon_{x_{H1}}^{(1,q)}$ should be bounded. Taking into consideration that $\varepsilon_{x_{H1}}^{(1,q)} = Q_2\varepsilon_{x_H}^{(1,q)}$ and using the solution of (7.81) and (7.82), a bound on $\left|\varepsilon_{x_{H1}}^{(1,q)}\right|$ satisfies

$$\begin{aligned}\left|\varepsilon_{x_{H1}}^{(1,q)}\right| \leq & \rho_0^{(1,q)}e^{-\xi_0^{(1,q)}(t-T_L^{(1)})}\bar{x}_L^{(1,q)} + \int_{T_L^{(1)}}^t \rho^{(1,q)}e^{-\xi^{(1,q)}(t-\tau)}\left(\left|g_2^{(1)}\right|\left|\varepsilon_{x_{H1}}^{(1,q)}\right| \right. \\ & \left. + \left|b^{(1)}\right|\bar{d}_1^{(2)} + \bar{\eta}_L^{(1)}\right)d\tau + \int_{T_L^{(1)}}^t \rho_d^{(1,q)}e^{-\xi_d^{(1,q)}(t-\tau)}\bar{d}_q^{(1)}d\tau\end{aligned}\quad (7.83)$$

By re-writing (7.83), we obtain

$$\left|\varepsilon_{x_{H1}}^{(1,q)}\right| \leq E_H^{(1,q)} + \int_{T_L^{(1)}}^t \rho^{(1,q)}\left|g_2^{(1)}\right|e^{-\xi^{(1,q)}(t-\tau)}\left|\varepsilon_{x_{H1}}^{(1,q)}\right|d\tau \quad (7.84)$$

where

$$E_H^{(1,q)} = \rho_0^{(1,q)} e^{-\xi_0^{(1,q)}(t-T_L^{(1)})} \bar{x}_L^{(1,q)} + \int_{T_L^{(1)}}^t \rho^{(1,q)} e^{-\xi^{(1,q)}(t-\tau)} \left(|b^{(1)}| \bar{d}_1^{(2)} \right. \quad (7.85)$$

$$\left. + \bar{\eta}_L^{(1)} \right) d\tau + \int_{T_L^{(1)}}^t \rho_d^{(1,q)} e^{-\xi_d^{(1,q)}(t-\tau)} \bar{d}_q^{(1)} d\tau \quad (7.86)$$

By applying the Bellman-Gronwall Lemma [Ioannou and Sun [1995]], an upper bound on $|\varepsilon_{x_1}^{(1,q)}|$ such that

$$|\varepsilon_{x_{H1}}^{(1,q)}| \leq Z_{H1}^{(1,q)} \quad (7.87)$$

is determined as

$$Z_{H1}^{(1,q)} = \rho^{(1,q)} |g_2^{(1)}| \int_0^t e^{-\left(\xi^{(1,q)} - \rho^{(1,q)} |g_2^{(1)}|\right)(t-\tau)} E_H^{(1,q)} d\tau + E_H^{(1,q)} \quad (7.88)$$

which is finite if $\xi^{(1,q)} > \rho^{(1,q)} |g_2^{(1)}|$. Thus, the observer $\mathcal{O}^{(1,q)}$ is stable, if the gain $L^{(1,q)}$ is designed such that the matrix $A^{(1)} - L^{(1,q)}C^{(1,q)}$ is stable and there are positive constants $\rho^{(1,q)}, \xi^{(1,q)}$ for which $|Q_2 e^{(A^{(1)} - L^{(1,q)}C^{(1,q)})t} Q_1| \leq \rho^{(1,q)} e^{-\xi^{(1,q)}t}$ and $\xi^{(1,q)} > \rho^{(1,q)} |g_2^{(1)}|$. Similarly for the monitoring modules $\mathcal{M}^{(2,q)}$, the distributed nonlinear observer is structured as

$$\begin{aligned} \dot{\hat{x}}^{(2,q)} = & A^{(2)} \hat{x}^{(2,q)} + Q_1 \left(g_1^{(2)} u^{(2)} \right) + Q_1 \left(g_2^{(2)} \sin(\hat{x}_1^{(2,q)}) \right) + \hat{\eta}_L^{(2)}(\hat{x}^{(2)}) \\ & + Q_1 \left(b^{(2)} \cos(y_1^{(2)}) \right) + L^{(2,q)} \left(y_q^{(2)} - C_q^{(2)} \hat{x}^{(2,q)} \right) \end{aligned} \quad (7.89)$$

where the gain $L^{(2,q)}$ is designed such that the matrix $A^{(2)} - L^{(2,q)}C^{(2,q)}$ is stable and there are positive constants $\rho^{(2,q)}, \xi^{(2,q)}$ for which $|Q_2 e^{(A^{(2)} - L^{(2,q)}C^{(2,q)})t} Q_1| \leq \rho^{(2,q)} e^{-\xi^{(2,q)}t}$ and $\xi^{(2,q)} > \rho^{(2,q)} |g_2^{(2)}|$.

The nonlinear function approximator of the modeling uncertainty is used in the design of the adaptive thresholds associated with the monitoring modules $\mathcal{M}^{(I,q)}$. Particularly, taking into account $\varepsilon_{x_{H1}}^{(I,q)} = Q_2 \varepsilon_{x_H}^{(I,q)} = C_1^{(I)} \varepsilon_{x_H}^{(I,q)}$, as well as that $|\varepsilon_{y_{Hj}}^{(I,q)}| \leq |C_j^{(I)} \varepsilon_{x_H}^{(I,q)}| + \bar{d}_j^{(I,q)}$, the adaptive threshold $\bar{\varepsilon}_{y_1}^{(I,1)}$ of the module $\mathcal{M}^{(I,1)}$ is determined as

$$\begin{aligned}\bar{\varepsilon}_{y_1}^{(I,1)} &= \rho^{(I,1)} |g_2^{(I)}| \int_{T_L^{(I)}}^t e^{-\left(\xi^{(I,1)} - \rho^{(I,1)} |g_2^{(I)}|\right)(t-\tau)} E^{(I,1)} d\tau \\ &\quad + E^{(I,1)} + \bar{d}_1^{(I)},\end{aligned}\tag{7.90}$$

$$\begin{aligned}E^{(I,1)} &= \int_{T_L^{(I)}}^t \rho^{(I,1)} e^{-\xi^{(I,1)}(t-\tau)} \bar{\eta}_L^{(1)} d\tau \\ &\quad + \rho_0^{(I,1)} e^{-\xi_0^{(I,1)}(t-T_L^{(I)})} \bar{x}_L^{(I,q)} + \frac{\rho_d^{(I,1)}}{\xi_d^{(I,1)}} \bar{d}_1^{(I)} \left(1 - e^{-\xi_d^{(I,1)}(t-T_L^{(I)})}\right) \\ &\quad + \frac{\rho^{(I,1)}}{\xi^{(I,1)}} \left|b^{(I)}\right| \bar{d}_1^{(Q)} \left(1 - e^{-\xi^{(I,1)}(t-T_L^{(I)})}\right),\end{aligned}\tag{7.91}$$

where $\bar{e}_\eta^{(I)}$ is a known bound such that $\left|e_\eta^{(I)}(x_1^{(I)})\right| \leq \bar{e}_\eta^{(I)}$, and $\bar{d}_1^{(Q)} = \bar{d}_1^{(2)}$ if $I = 1$ and $\bar{d}_1^{(Q)} = \bar{d}_1^{(1)}$ if $I = 2$.

The adaptive threshold $\bar{\varepsilon}_{y_2}^{(I,2)}$ is computed by taking into account (4.201), and the solution of (7.81), and by selecting $\alpha_0^{(I,2)}$, $\zeta_0^{(I,2)}$, $\alpha^{(I,2)}$, $\zeta^{(I,2)}$, $\alpha_d^{(I,2)}$, $\zeta_d^{(I,2)}$ such that (4.207)-(4.209) are satisfied; i.e. a bound on $\left|\varepsilon_{y_{H2}}^{(I,2)}\right|$ is defined as:

$$\begin{aligned}\bar{\varepsilon}_{y_2}^{(I,2)} &= \alpha_0^{(I,q)} e^{-\zeta_0^{(I,q)}(t-T_L^{(I)})} \bar{x}_L^{(I,q)} + \int_{T_L^{(I)}}^t \alpha^{(I,2)} e^{-\zeta^{(I,2)}(t-\tau)} \left(|g_2^{(I)}| Z_1^{(I,2)} \right. \\ &\quad \left. + \bar{\eta}_L^{(I)}\right) d\tau + \frac{\alpha_d^{(I,2)}}{\zeta_d^{(I,2)}} \bar{d}_2^{(I)} \left(1 - e^{-\zeta_d^{(I,2)}(t-T_L^{(I)})}\right) \\ &\quad + \frac{\alpha^{(I,2)}}{\zeta^{(I,2)}} \left|b^{(I)}\right| \bar{d}_1^{(Q)} \left(1 - e^{-\zeta^{(I,2)}(t-T_L^{(I)})}\right) + \bar{d}_2^{(I)}\end{aligned}\tag{7.92}$$

$$Z_1^{(I,2)} = \rho^{(I,2)} |g_2^{(I)}| \int_{T_L^{(I)}}^t e^{-\left(\xi^{(I,2)} - \rho^{(I,2)} |g_2^{(I)}|\right)(t-\tau)} E^{(I,2)} d\tau + E^{(I,2)}\tag{7.93}$$

and

$$\begin{aligned}
E^{(I,2)} = & \int_{T_L^{(I)}}^t \rho^{(I,2)} e^{-\xi^{(I,2)}(t-\tau)} \bar{\eta}_L^{(I)} d\tau + \rho_0^{(I,2)} e^{-\xi_0^{(I,2)}(t-T_L^{(I)})} \bar{x}_L^{(I,q)} \\
& + \frac{\rho_d^{(I,2)}}{\xi_d^{(I,2)}} \bar{d}_2^{(I)} \left(1 - e^{-\xi_d^{(I,2)}(t-T_L^{(I)})}\right) \\
& + \frac{\rho^{(I,2)}}{\xi^{(I,2)}} |b^{(I)}| \bar{d}_1^{(Q)} \left(1 - e^{-\xi^{(I,2)}(t-T_L^{(I)})}\right)
\end{aligned} \tag{7.94}$$

The adaptive thresholds are implemented using linear filters, as described in Section 4.5.2.

7.5.3 Distributed Sensor Fault Decision Logic

The local distributed sensor fault isolation decision logic, which is implemented in each monitoring agent $\mathcal{M}^{(I,q)}$, $I = 1, 2$, $q = 1, 2$ is based on the fault signature matrices shown in Tables 7.2 and 7.3. These matrices are designed based on the discussion in Section 7.4.3, that is the occurrence of sensor faults may change the domain of approximation of the modeling uncertainty, and consequently the bound on the deviation of the approximator from the actual modeling uncertainty. The global distributed sensor fault isolation is realized based on the fault signature matrix presented in Table 5.7.

	$\{f_1^{(1)}\}$	$\{f_2^{(1)}\}$	$\{f_1^{(1)}, f_2^{(1)}\}$	$\{f_1^{(2)}\}$	$\bigcup_{i \in \{1,2,3\}} \{f_1^{(2)}, \mathcal{F}_{c_i}^{(1)}\}$
$\mathcal{E}^{(1,1)}$	1	*	1	1	1
$\mathcal{E}^{(1,2)}$	*	1	1	1	1

Table 7.2: Distributed configuration of sensor fault signature matrix $F^{(1)}$ ($\mathcal{F}_{c_1}^{(1)} = \{f_1^{(1)}\}$, $\mathcal{F}_{c_2}^{(1)} = \{f_2^{(1)}\}$ and $\mathcal{F}_{c_3}^{(1)} = \{f_1^{(1)}, f_2^{(1)}\}$).

7.5.4 Simulation Results

This section presents the simulation results of the application of the distributed adaptive nonlinear scheme presented in Section 7.5.1, and the distributed sensor fault diagnosis scheme using the nonlinear functional approximator of the modeling uncertainty described in Section

	$\{f_1^{(2)}\}$	$\{f_2^{(2)}\}$	$\{f_1^{(2)}, f_2^{(2)}\}$	$\{f_1^{(1)}\}$	$\bigcup_{i \in \{1,2,3\}} \{f_1^{(1)}, \mathcal{F}_{c_i}^{(2)}\}$
$\mathcal{E}^{(2,1)}$	1	*	1	1	1
$\mathcal{E}^{(2,2)}$	*	1	1	1	1

Table 7.3: Distributed configuration of sensor fault signature matrix $F^{(2)}$ ($\mathcal{F}_{c_1}^{(2)} = \{f_1^{(2)}\}$, $\mathcal{F}_{c_2}^{(2)} = \{f_2^{(2)}\}$ and $\mathcal{F}_{c_3}^{(2)} = \{f_1^{(2)}, f_2^{(2)}\}$).

7.5.2. For simulation purposes, the modeling uncertainty affecting $\Sigma^{(1)}$ and $\Sigma^{(2)}$ described by (2.6) and (2.7) is:

$$\eta_2^{(1)}(x_1^{(1)}) = \tilde{g}_2^{(1)} \cos(x_1^{(1)}) = 10 \cos(x_1^{(1)}), \quad (7.95)$$

$$\eta_2^{(2)}(x_1^{(2)}) = \tilde{g}_2^{(2)} \sin(x_1^{(2)}) = 20 \sin(x_1^{(2)}), \quad (7.96)$$

with $\bar{g}_2^{(1)} = 12$ and $\bar{g}_2^{(2)} = 24$ such that $\bar{g}_2^{(I)} \geq |\tilde{g}_2^{(I)}|$, $I = 1, 2$. The remainder system parameters are given in Section 4.5.4. The nonlinear adaptive approximation model $\hat{\eta}^{(I)}$ is implemented by using a network of 12 radial basis functions (RBF), with fixed centers, denoted by $c_{G_i}^{(I)}$, $i \in \{1, \dots, 12\}$ that are evenly distributed in the interval $\mathcal{X}_G^{(I)}$, and width, denoted by w_{G_I} , which is the same for all $i \in \{1, \dots, 12\}$; i.e. the regressors $G^{(1)}(x_1^{(1)})$ and $G^{(2)}(x_1^{(2)})$ given in (7.48) and (7.49) are expressed as

$$G^{(I)}(x_1^{(I)}) = \left[G_1^{(I)}(x_1^{(I)}), \dots, G_{12}^{(I)}(x_1^{(I)}) \right] \quad (7.97)$$

$$G_i^{(I)}(x_1^{(I)}) = e^{-\frac{(x_1^{(I)} - c_{G_i}^{(I)})^2}{w_{G_I}^2}}, \quad I = 1, 2 \quad (7.98)$$

In order to learn the modeling uncertainty of each system, $\eta_2^{(1)}$ and $\eta_2^{(2)}$, we apply the distributed adaptive nonlinear estimation scheme presented in Section 7.5.1, using the following parameters: $\text{eig}(A_L^{(1)}) = \{-10, -11\}$, $\text{eig}(A_L^{(2)}) = \{-10, -11\}$, $L^{(1)} = [101; 99]$, $L^{(2)} = [151; 123]$, $\Gamma^{(1)} = \Gamma^{(2)} = 10 * I_{12}$ where I_{12} is the identity matrix of length 12 and the projection operator $\mathcal{P}_{\Theta}^{(I)}$ is implemented as described in Remark 7.2 with $M_{\theta_I} = 100$, $I = 1, 2$. In order to analyze the performance of the distributed adaptive learning scheme, we performed three simulation

scenarios with different sensor noise levels, as shown in Table 7.4. For each scenario, the interval $\mathcal{X}_G^{(I)}$ (consequently the RBF centers that are evenly distributed in $\mathcal{X}_G^{(I)}$) and the RBF width vary. The learning period for both subsystems is $[0, T_L^{(1)}] \equiv [0, T_L^{(2)}]$ with $T_L^{(1)} = T_L^{(2)} = 1000$ sec.

	Simulation Scenarios		
	1	2	3
$\bar{d}_i^{(I)}, i = 1, 2, I = 1, 2$	0	$3\%x_1^\circ$	$5\%x_1^\circ$
$\mathcal{X}_G^{(1)}$	$[-6, 6.6]$	$[-6.3, 7]$	$[-6.5, 7.2]$
$\mathcal{X}_G^{(2)}$	$[-5.2, 5.2]$	$[-5.7, 5.7]$	$[-6, 6]$
w_{G_1}	0.9	0.9	1
w_{G_2}	1	0.9	0.8

Table 7.4: Simulation scenarios for distributed adaptive learning process ($x_1^\circ = 2\pi$).

Figures 7.2 and 7.3 present the results of the distributed learning process for the three simulation scenarios, described in Table 7.4. Particularly, the first row of subfigures in Fig. 7.2 (correspondingly for Fig. 7.3) illustrates the estimation of the unknown weights $\hat{\theta}^{(1)}$ in relation to the optimal weights $\theta^{(1)}$, which are different in each simulation scenario, since the regressors (i.e. RBFs) are different (see Table 7.4). Moreover, the presence of noise has an impact on the parameter estimation error. The second row of subfigures in Fig. 7.2 (correspondingly for Fig. 7.3) depicts the absolute value of the adaptive approximation error $|\eta_2^{(1)}(x_1^{(1)}(t)) - G^{(1)}(\hat{x}_1^{(1)}(t))\hat{\theta}^{(1)}(t)|$, where $\hat{x}_1^{(1)}$ is the estimation model generated based on (7.52). The effects of the noise on the learning process are noticeable in subfigures 7.2-5 and 7.2-6. The third row of subfigures in Fig. 7.2 (correspondingly for Fig. 7.3) presents the modeling uncertainty $\eta_2^{(1)}(x_1^{(1)})$ ($\eta_2^{(2)}(x_1^{(2)})$) and the adaptive approximator $G^{(1)}(\hat{x}_1^{(1)}(t))\hat{\theta}^{(1)}(t)$ ($G^{(2)}(\hat{x}_1^{(2)}(t))\hat{\theta}^{(2)}(t)$) in the end of the learning process. The effects of noise are more visible in the curvaceous part of the modeling uncertainty, where the learning ability of the adaptive approximator is reduced. The learning capability of the distributed adaptive approximation scheme is further illustrated in Figure 7.4, which shows the actual modeling uncertainty and its corresponding nonlinear

approximator derived in the end of the learning period for the three different simulation scenarios. Also in this figure, the noise effects in both adaptive learning procedures are apparent in subfigures 7.4-3 – 7.4-6.

After applying the distributed adaptive approximation scheme for learning the modeling uncertainty associated with $\Sigma^{(1)}$ and $\Sigma^{(2)}$, we use the resultant approximators $\hat{\eta}_L^{(1)}$ and $\hat{\eta}_L^{(2)}$ defined in (7.79) in the distributed diagnostic scheme as described in Section 7.5.2. In order to illustrate the benefit of learning the modeling uncertainty, we implemented the distributed diagnostic scheme with [(A)-DSFD scheme] and without [(B)-DSFD scheme] the nonlinear approximator.

Without loss of generality the starting time for simulating the two subsystems with the two DSFD schemes is 0. In the (A)-DSFD scheme, we implemented the nonlinear observers $\mathcal{O}^{(1,q)}$ and $\mathcal{O}^{(2,q)}$ given in (7.80) and (7.89) respectively, while in the (B)-DSFD scheme, we implemented the nonlinear observers $\mathcal{O}^{(1,q)}$ and $\mathcal{O}^{(2,q)}$ given in (4.184) with $z_*^{(1)} = y_1^{(2)}$ and (4.199) with $z_*^{(2)} = y_1^{(1)}$. In the (A)-DSFD scheme, the implementation of the adaptive thresholds are based on (7.90) and (7.91) with $T_L^{(I)} = 0$ and is realized using linear filters; i.e. for $I = 1, 2$ and $Q = 1, 2$ with $Q \neq I$, and $j = 1, 2$

$$\bar{\varepsilon}_{y_1}^{(I,1)} = \left(\frac{\rho^{(I,1)} |g_2^{(I)}|}{s + \xi^{(I,1)} - \rho^{(I,1)} |g_2^{(I)}|} + 1 \right) [E^{(I,1)}] + \bar{d}_1^{(I)}, \quad (7.99)$$

$$\begin{aligned} \bar{\varepsilon}_{y_2}^{(I,2)} = & \frac{\alpha^{(I,2)}}{\zeta^{(I,2)}} |b^{(I)}| \bar{d}_1^{(Q)} (1 - e^{-\zeta^{(I,2)} t}) + \frac{\alpha^{(I,2)}}{s + \zeta^{(I,2)}} [\bar{\eta}_L^{(I)} + |g_2^{(1)}| Z_1^{(I,2)}] \\ & + \alpha_0^{(I,2)} e^{-\zeta_0^{(I,2)} t} \bar{x}_L^{(I,2)} + \frac{\alpha_d^{(I,2)}}{\zeta_d^{(I,2)}} \bar{d}_2^{(I)} (1 - e^{-\zeta_d^{(I,2)} t}) + \bar{d}_2^{(I)} \end{aligned} \quad (7.100)$$

$$Z_1^{(I,2)} = \left(\frac{\rho^{(I,2)} |g_2^{(I)}|}{s + \xi^{(I,2)} - \rho^{(I,2)} |g_2^{(I)}|} + 1 \right) [E^{(I,2)}] \quad (7.101)$$

$$\begin{aligned} E^{(I,j)} = & \frac{\rho^{(I,j)}}{\xi^{(I,j)}} |b^{(I)}| \bar{d}_1^{(Q)} (1 - e^{-\xi^{(I,j)} t}) + \frac{\rho^{(I,j)}}{s + \xi^{(I,j)}} [\bar{\eta}_L^{(I)}] \\ & + \rho_0^{(I,j)} e^{-\xi_0^{(I,j)} t} \bar{x}_L^{(I,j)} + \frac{\rho_d^{(I,j)}}{\xi_d^{(I,j)}} \bar{d}_j^{(I)} (1 - e^{-\xi_d^{(I,j)} t}). \end{aligned} \quad (7.102)$$

In (B)-DSFD scheme, the adaptive thresholds are implemented based

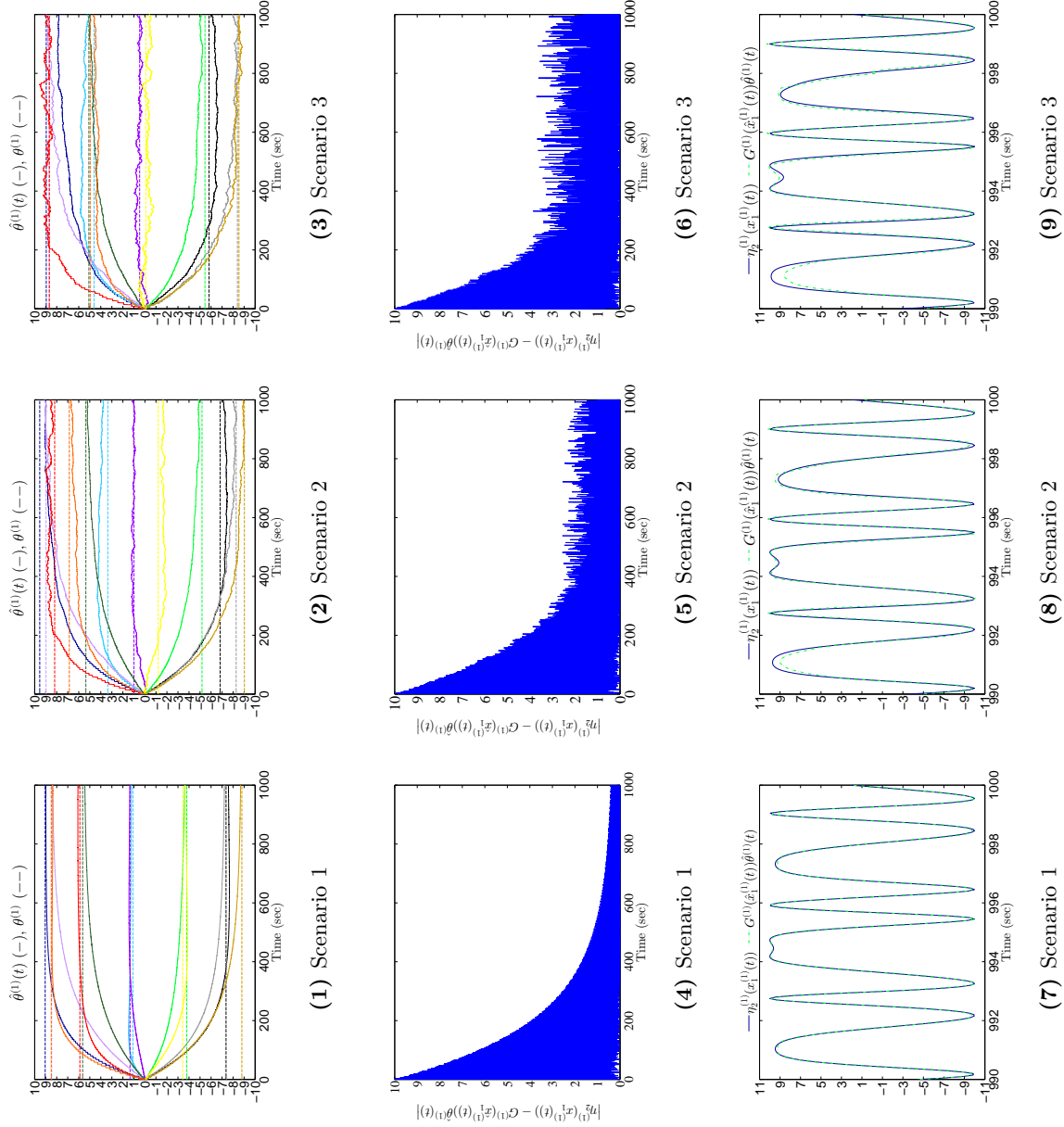


Figure 7.2: Adaptive approximation process for learning $\eta_2^{(1)}$ associated with $\Sigma^{(1)}$ within the time interval $[0, T_L^{(1)}]$ with $T_L^{(1)} = 1000$ sec. Subfigures 7.2-1-7.2-3 show the estimation of the weights $\hat{\theta}^{(1)}(t)$ in relation to the nominal parameters $\theta^{(1)}$. Subfigures 7.2-4-7.2-6 present the absolute value of the adaptive approximation error $|\eta_2^{(1)}(x_1^{(1)}(t)) - G^{(1)}(\hat{x}_1^{(1)}(t))\hat{\theta}^{(1)}(t)|$. Subfigures 7.2-7-7.2-9 illustrate the modeling uncertainty $\eta_2^{(1)}(x_1^{(1)}(t))$ and adaptive approximator $G^{(1)}(\hat{x}_1^{(1)}(t))\hat{\theta}^{(1)}(t)$ for $t \in [990, 1000]$ sec.

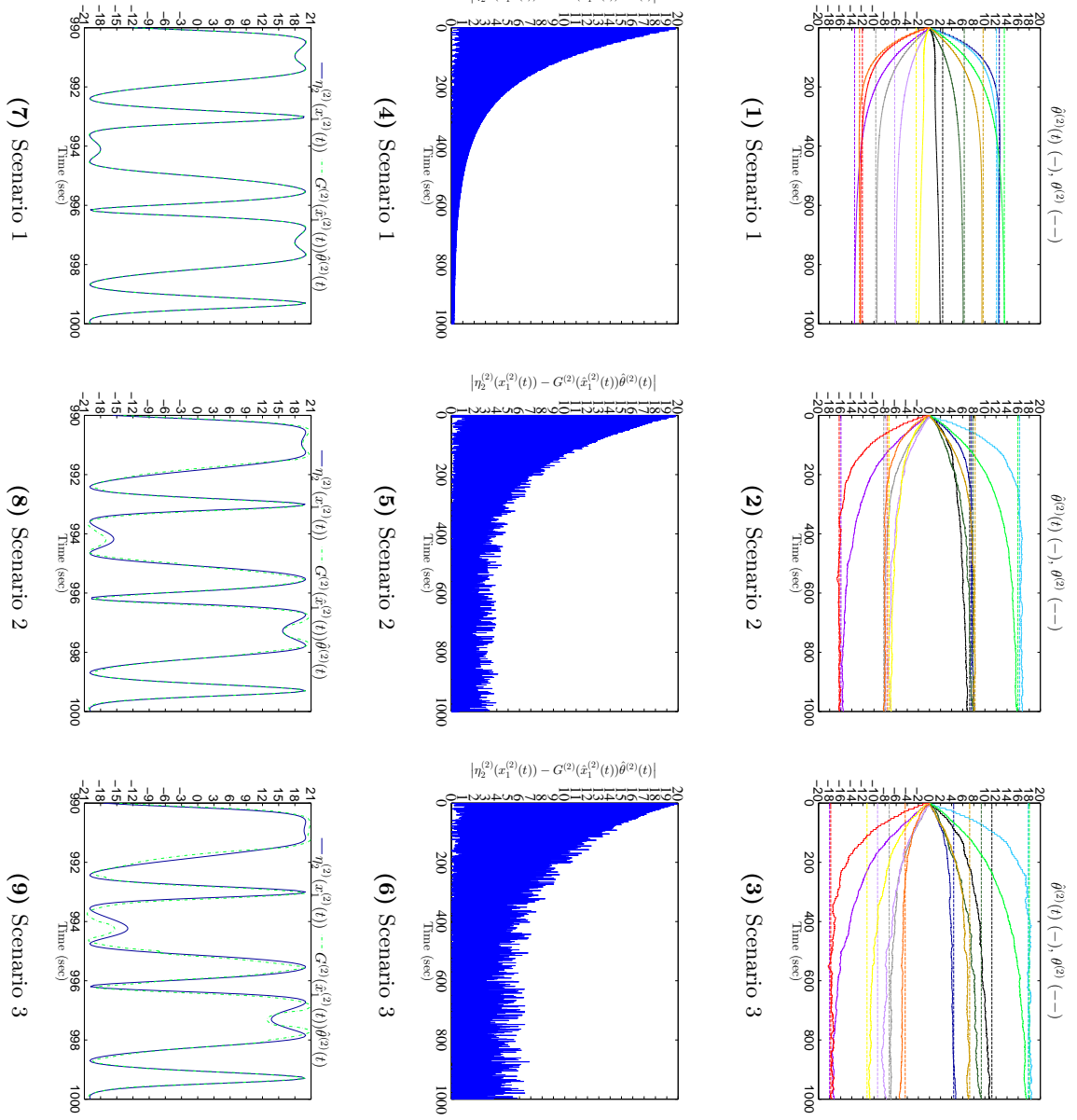
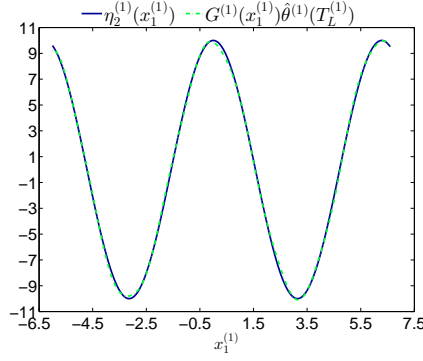
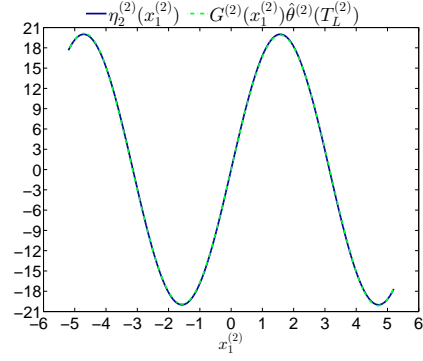


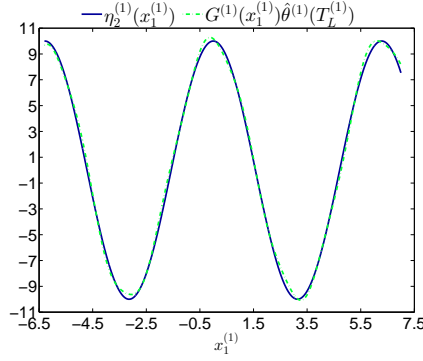
Figure 7.3: Adaptive approximation process for learning $\eta_2^{(2)}$ associated with $\Sigma^{(2)}$ within the time interval $[0, T_L^{(2)}]$ with $T_L^{(2)} = 1000$ sec. Subfigures 1-3 show the estimation of the weights $\hat{\theta}^{(2)}(t)$ in relation to the nominal parameters $\theta^{(2)}$. Subfigures 4-6 present the absolute value of the adaptive approximation error $|\eta_2^{(2)}(x_1^{(2)}(t)) - G^{(2)}(\hat{x}_1^{(2)}(t))\hat{\theta}^{(2)}(t)|$. Subfigures 7-9 illustrate the modeling uncertainty $\eta_2^{(2)}(x_1^{(2)}(t))$ and adaptive approximator $G^{(2)}(\hat{x}_1^{(2)}(t))\hat{\theta}^{(2)}(t)$ for $t \in [990, 1000]$ sec.



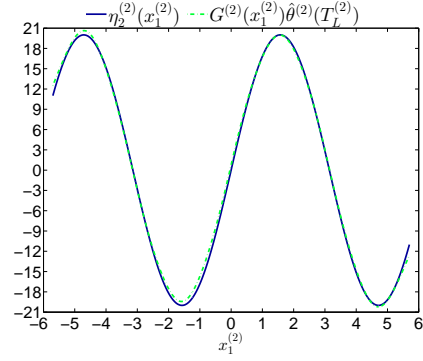
(1) Scenario 1



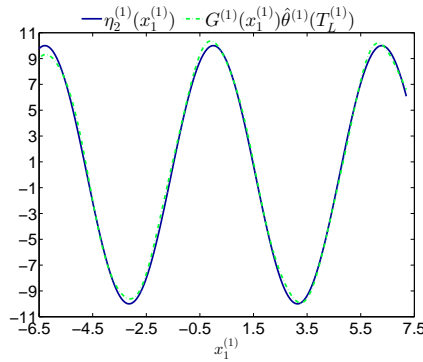
(2) Scenario 1



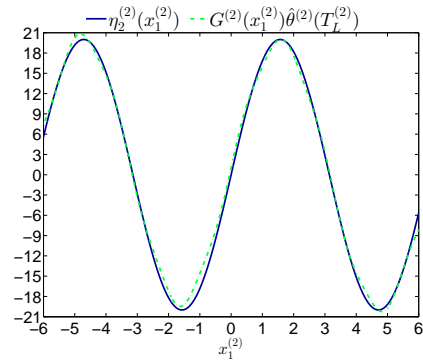
(3) Scenario 2



(4) Scenario 2



(5) Scenario 3



(6) Scenario 3

Figure 7.4: Modeling uncertainty $\eta_2^{(I)}(x_1^{(I)})$ and nonlinear approximator $G^{(I)}(x_1^{(I)})\hat{\theta}^{(I)}(T_L^{(I)})$ ($T_L^{(I)} = 1000$ sec) for $x_1^{(I)} \in \mathcal{X}_{G_I}$. Subfigures 7.4-1, 7.4-3 and 7.4-5 are associated to $\Sigma^{(1)}$ ($I = 1$), while subfigures 7.4-2, 7.4-4 and 7.4-6 are associated to $\Sigma^{(2)}$ ($I = 2$).

on (7.99),(7.101) for $I, j = 1, 2$

$$\begin{aligned} \bar{\varepsilon}_{y_2}^{(I,2)} &= \frac{\alpha^{(I,2)}}{\zeta^{(I,2)}} \left| b^{(I)} \right| \bar{d}_1^{(Q)} (1 - e^{-\zeta^{(I,2)}t}) + \frac{\alpha^{(I,2)}}{s + \zeta^{(I,2)}} \left[\bar{g}_2^{(I)} + \left| g_2^{(1)} \right| Z_1^{(I,2)} \right] \\ &\quad + \alpha_0^{(I,2)} e^{-\zeta_0^{(I,2)}t} \bar{x}_L^{(I,2)} + \frac{\alpha_d^{(I,2)}}{\zeta_d^{(I,2)}} \bar{d}_2^{(I)} (1 - e^{-\zeta_d^{(I,2)}t}) + \bar{d}_2^{(I)} \end{aligned} \quad (7.103)$$

$$\begin{aligned} E^{(I,j)} &= \rho_0^{(I,j)} e^{-\xi_0^{(I,j)}t} \bar{x}_L^{(I,j)} + \frac{\rho_d^{(I,j)}}{\xi_d^{(I,j)}} \bar{d}_j^{(I)} (1 - e^{-\xi_d^{(I,j)}t}) \\ &\quad + \frac{\rho^{(I,j)}}{\xi^{(I,j)}} \left| b^{(I)} \right| \bar{d}_1^{(Q)} (1 - e^{-\xi^{(I,j)}t}) + \frac{\rho^{(I,j)}}{s + \xi^{(I,j)}} \left[\bar{g}_2^{(I)} \right]. \end{aligned} \quad (7.104)$$

Remark 7.3. The computation of the adaptive thresholds in (B-DSFD scheme) is realized following the procedure described in Sections 4.5.1 and 4.5.2, where however instead of using the bound in (4.192), we use

$$\bar{g}_2^{(1)} |\cos(x_1^{(1)})| \leq \bar{g}_2^{(1)}. \quad (7.105)$$

Using (7.105) in (4.193) leads to

$$\left| \varepsilon_{x_{H1}}^{(1,q)} \right| \leq E_H^{(1,q)} + \int_0^t \rho^{(1,q)} e^{-\xi^{(1,q)}(t-\tau)} |g_2^{(1)}| \left| \varepsilon_{x_{H1}}^{(1,q)} \right| d\tau \quad (7.106)$$

with

$$\begin{aligned} E_H^{(1,q)} &= \rho_0^{(1,q)} e^{-\xi_0^{(1,q)}t} \bar{x}^{(1,q)} + \int_0^t \rho^{(1,q)} e^{-\xi^{(1,q)}(t-\tau)} \bar{g}_2^{(1)} d\tau \\ &\quad + \frac{\rho_d^{(1,q)}}{\xi_d^{(1,q)}} \bar{d}_q^{(1)} (1 - e^{-\xi_d^{(1,q)}t}) + \frac{\rho^{(1,q)}}{\xi^{(1,q)}} \left| b^{(1)} \right| \bar{d}_1^{(2)} (1 - e^{-\xi^{(1,q)}t}) \end{aligned} \quad (7.107)$$

For simulation purposes, we used the design parameters of the 3rd, 1st, 2nd and 1st set shown in Tables A.1, A.2, A.3 and A.4, respectively. Moreover, for the implementation of the adaptive thresholds in the (A)-DSFD scheme, we determined the bounds on the deviation of the nonlinear approximator $\eta_L^{(I)}(\hat{x}_1^{(I,j)})$ from the modeling uncertainty $\eta_2^{(I)}(x_1^{(I)})$, $I = 1, 2$ and $j = 1, 2$ for different noise bounds based on the simulation results shown in Figs. 7.5 and 7.6 (in each simulation scenario the noise bound is the same for all sensors). The

bounds are summarized in Table 7.5. As observed, the monitoring module $\mathcal{M}^{(2,2)}$ is more sensitive to the presence of noise that increases the error $\eta_2^{(2)}(x_1^{(2)}) - \eta_L^{(I)}(\hat{x}_1^{(2,2)})$ affecting the state estimation.

	Noise level $\bar{d}_j^{(I)}$		
	0	3% x_1°	5% x_1°
$\bar{\eta}_L^{(1,1)}$	0.45	3	4
$\bar{\eta}_L^{(1,2)}$	0.45	3.2	4.5
$\bar{\eta}_L^{(2,1)}$	0.25	5	8
$\bar{\eta}_L^{(2,2)}$	0.25	8	15

Table 7.5: Simulation scenarios for distributed adaptive learning process ($x_1^\circ = 2\pi$).

We simulated abrupt, single permanent faults affecting the sensors of $\Sigma^{(1)}$ at the time instant $T_{f_1}^{(1)} = T_{f_1}^{(2)} = 150$ sec with evolution rate $\kappa_{f_q}^{(I)} = 10^4$. Six sensor fault scenarios have been simulated with different fault functions of $f_1^{(1)}$ and $f_2^{(1)}$ and noise bounds, summarized in Table 7.6: The local distributed sensor fault isolation decision logic for the

	Sensor fault scenario			
	1	2	3	4
$\bar{d}_j^{(I)}$	3% x_1°	3% x_1°	5% x_1°	5% x_1°
$\phi_1^{(1)}(t)$	25% x_1°	0	25% x_1°	0
$\phi_2^{(1)}(t)$	0	25% x_1°	0	50% x_1°

Table 7.6: Simulated sensor fault scenarios ($x_1^\circ = 2\pi$).

monitoring agents of the (A)-DSFD scheme relies on the fault signature matrices shown in Tables 7.2 and 7.3, while in the (B)-DSFD scheme the local isolation is realized based on the matrices shown in Tables 5.5 and 5.6.

Figures 7.7 – 7.10 present the performance of the monitoring agents $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ in the six scenarios summarized in Table 7.6. As ob-

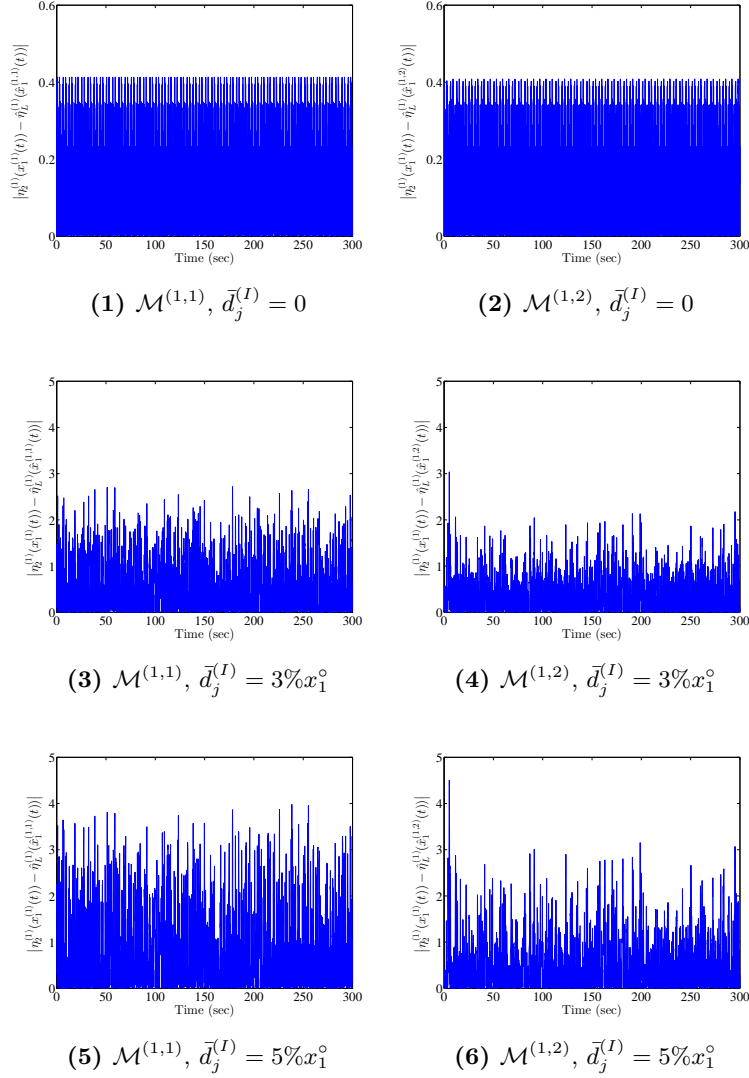


Figure 7.5: Absolute value of the deviation of the nonlinear approximator $\eta_L^{(I)}(\hat{x}_1^{(I,j)})$ from the modeling uncertainty $\eta_2^{(I)}(x_1^{(1)})$ for $I = 1$ and $j = 1, 2$. Each row of subfigures presents the simulation results with the same noise level, while each column illustrates the simulation results obtained by the same monitoring module.

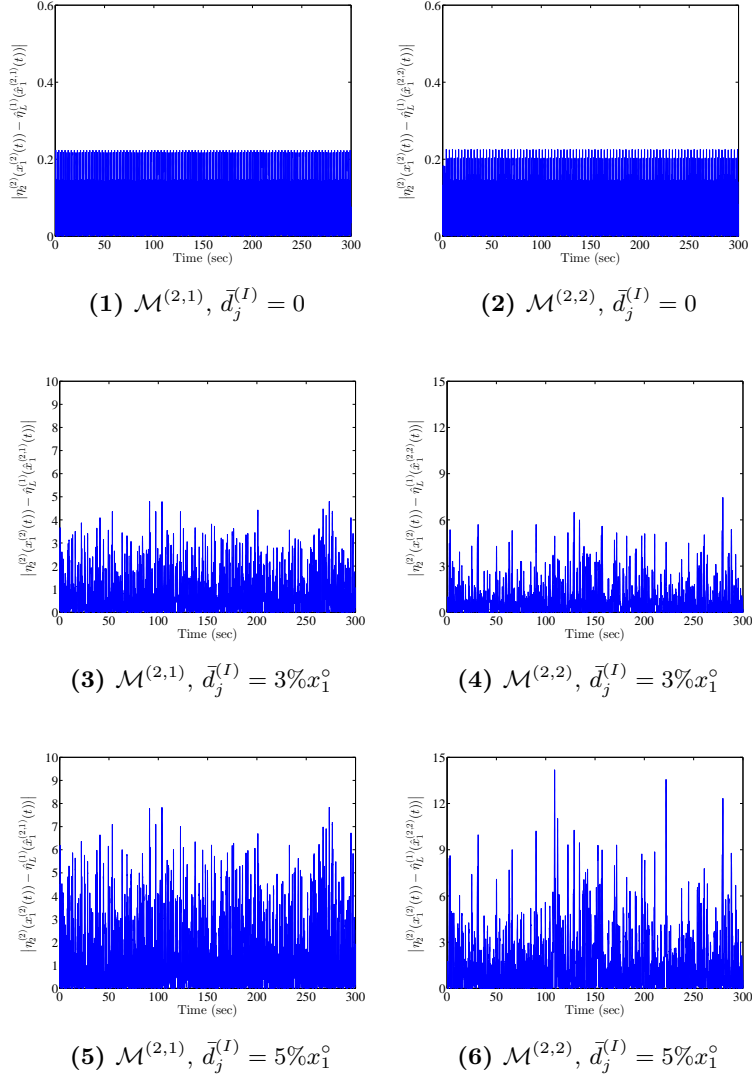


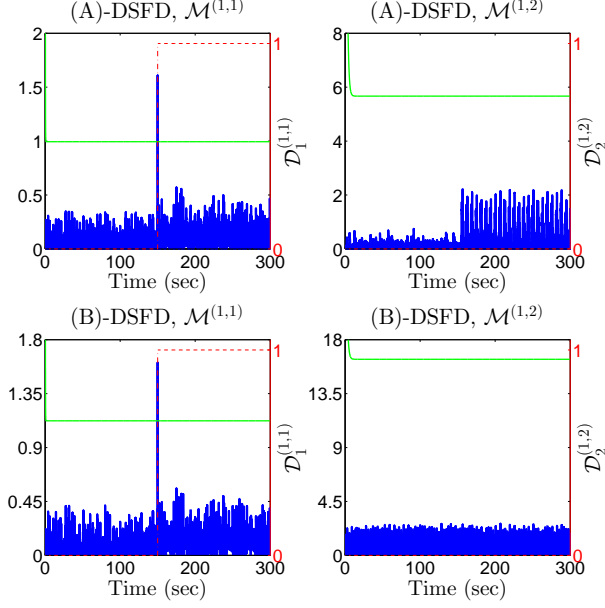
Figure 7.6: Absolute value of the deviation of the nonlinear approximator $\eta_L^{(I)}(\hat{x}_1^{(I,j)})$ from the modeling uncertainty $\eta_2^{(I)}(x_1^{(1)})$ for $I = 2$ and $j = 1, 2$. Each row of subfigures presents the simulation results with the same noise level, while each column illustrates the simulation results obtained by the same monitoring module.

served in all figures, the thresholds of the monitoring modules deployed in the (A)-DSFD scheme are lower than the thresholds in the (B)-DSFD scheme with the thresholds of the modules $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,2)}$ in the (B)-DSFD scheme being much higher than the thresholds of these modules in the (A)-DSFD scheme. Based on Fig. 7.7, in the first simulation scenario, the agent $\mathcal{M}^{(1)}$ detected the occurrence of sensor fault $f_1^{(1)}$ in both (A)- and (B)-DSFD schemes. Particularly in the (A)-DSFD scheme, based on the fault signature matrix shown in Table 7.2 and the observed pattern $D^{(1)} = [1, 0]^\top$, the diagnosis set of the agent $\mathcal{M}^{(1)}$ is $\mathcal{D}_s^{(1)} = \{f_1^{(1)}\}$, leading to the decision $D_z^{(1)} = 1$. The same observed pattern was generated in the (B)-DSFD scheme, which was compared to the fault signature matrix presented in Table 5.5, leading to the same $\mathcal{D}_s^{(1)}$ and $D_z^{(1)}$ as in the (A)-DSFD scheme. In both DSFD schemes, the agent $\mathcal{M}^{(2)}$ did not detect any fault. Thus, the global observed pattern is $D_z = [1, 0]^\top$, which is compared to the columns of the fault signature matrix shown in Table 5.7 leading to the diagnosis set $\mathcal{D}_s = \{f_1^{(1)}\}$.

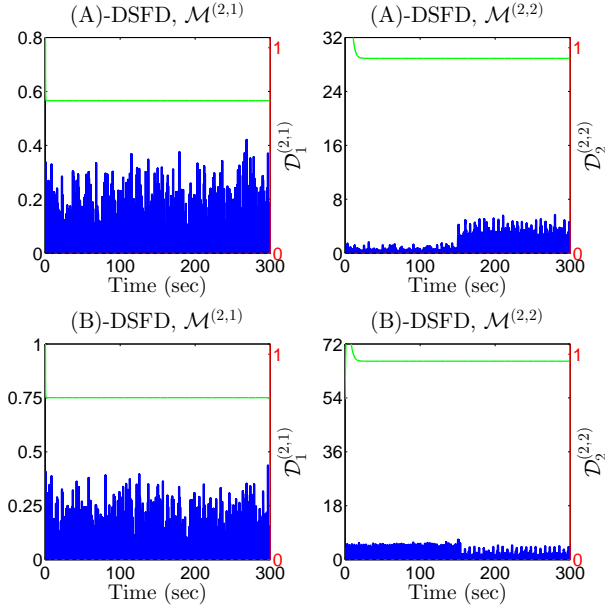
However, according to the results of second simulation scenario shown in Fig. 7.8, the agent $\mathcal{M}^{(1)}$ deployed in the (B)-DSFD scheme could not detect the occurrence of sensor fault $f_1^{(2)}$. On the contrary, the agent $\mathcal{M}^{(1)}$ in the (A)-DSFD scheme detected and isolated sensor fault $f_1^{(2)}$ by comparing the observed pattern $D^{(1)} = [0, 1]^\top$ to the columns of the fault signature matrix presented Table 7.2. Based on Figs. 7.9 and 7.10 that present the results of the third and fourth simulation scenario, the agent $\mathcal{M}^{(1)}$ deployed in the (B)-DSFD scheme could detect neither sensor fault $f_1^{(1)}$ nor sensor fault $f_1^{(2)}$, while the agent $\mathcal{M}^{(1)}$ deployed in the (A)-DSFD scheme detected and isolated both of them.

7.6 Summary and Discussion

The objective of this chapter was the design and analysis of an adaptive approximation methodology, which aims at learning the modeling uncertainty under healthy conditions. By considering that the modeling uncertainty may be a nonlinear function of the system state and input,

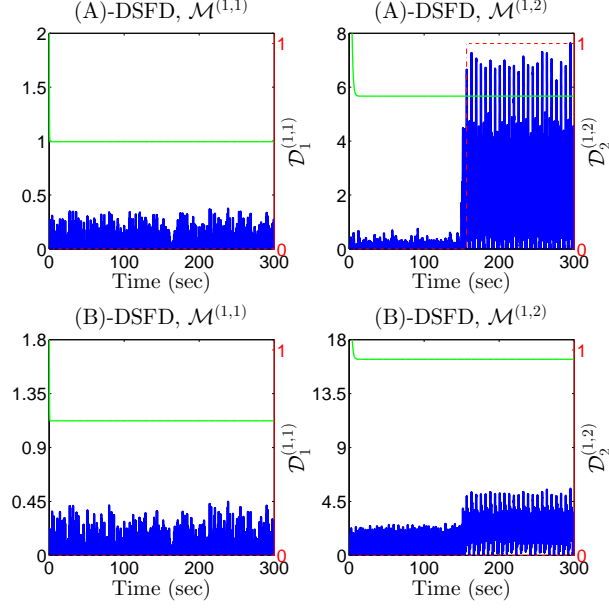


(1) Agent $\mathcal{M}^{(1)}$ deployed in (A)- and (B)-DSFD scheme.

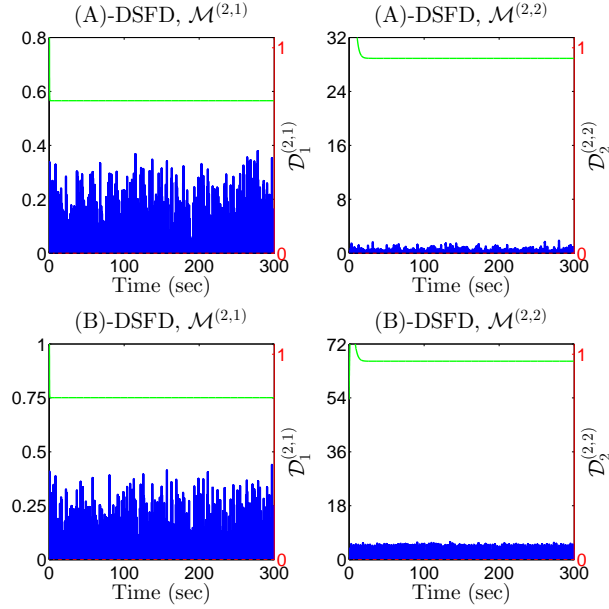


(2) Agent $\mathcal{M}^{(2)}$ deployed in (A)- and (B)-DSFD scheme.

Figure 7.7: Sensor fault diagnosis for the simulated scenario 1. Subfigures 1(a)-2(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (green line), $\varepsilon_{y_q}^{(I,q)}(t)$ (blue line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.

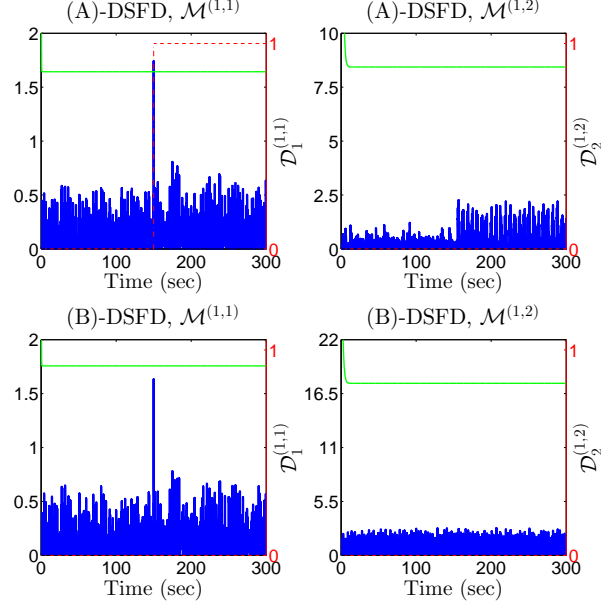


(1) Agent $\mathcal{M}^{(1)}$ deployed in (A)- and (B)-DSFD scheme.

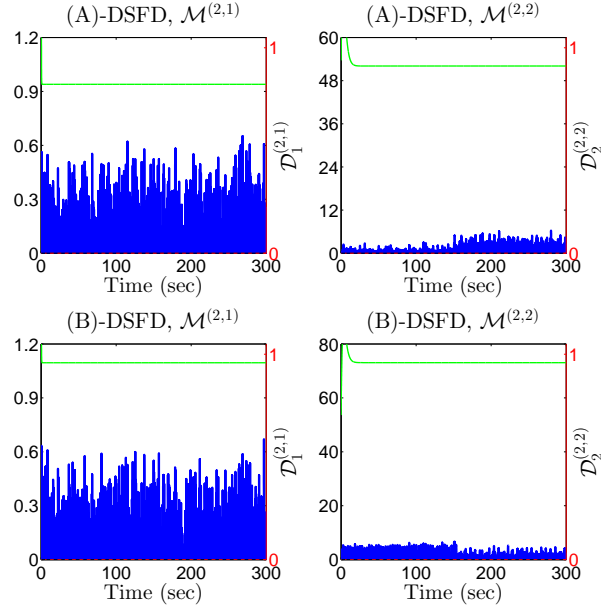


(2) Agent $\mathcal{M}^{(2)}$ deployed in (A)- and (B)-DSFD scheme.

Figure 7.8: Sensor fault diagnosis for the simulated scenario 2. Subfigures 1(a)-2(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (green line), $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (blue line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.

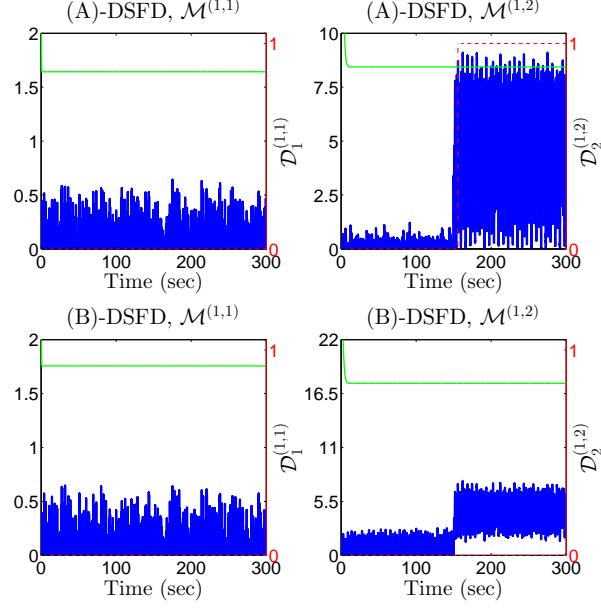


(1) Agent $\mathcal{M}^{(1)}$ deployed in (A)- and (B)-DSFD scheme.

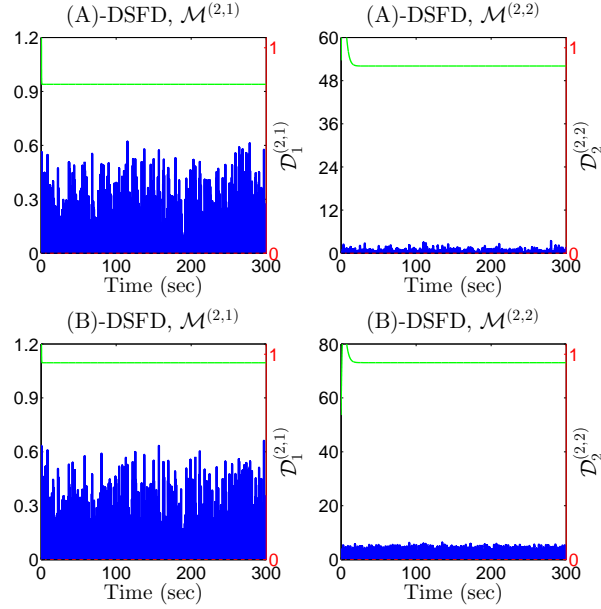


(2) Agent $\mathcal{M}^{(2)}$ deployed in (A)- and (B)-DSFD scheme.

Figure 7.9: Sensor fault diagnosis for the simulated scenario 3. Subfigures 1(a)-2(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (green line), $\bar{\varepsilon}_{y_q}^{(I,q)}(t)$ (blue line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.



(1) Agent $\mathcal{M}^{(1)}$ deployed in (A)- and (B)-DSFD scheme.



(2) Agent $\mathcal{M}^{(2)}$ deployed in (A)- and (B)-DSFD scheme.

Figure 7.10: Sensor fault diagnosis for the simulated scenario 4. Subfigures 1(a)-2(d) show the temporal evolution of $\left| \varepsilon_{y_q}^{(I,q)}(t) \right|$ (green line), $\hat{\varepsilon}_{y_q}^{(I,q)}(t)$ (blue line) and $D_q^{(I,q)}(t)$ (red, dashed line), $I, q = 1, 2$.

we formulated a parametric model for the approximation of the modeling uncertainty and designed a nonlinear adaptive estimation scheme for estimating both the state and weights of the selected approximator (e.g. linearly parametrized approximator). Then, the estimated weights obtained in the end of the learning period, and consequently the approximation model of the uncertainty are used in the distributed sensor fault detection scheme. Specifically, the trained approximator of the modeling uncertainty is incorporated in the structure of the nonlinear observers used in the local monitoring modules. Then the adaptive thresholds are designed by using a bound on the deviation of the approximator, which was trained under healthy conditions, and the actual modeling uncertainty. The distributed sensor fault isolation decision logic was properly customized in order to handle the issue that the bound on the deviation of the approximator from the actual uncertainty, which is computed under healthy conditions, as well as the filtered version of this bound used in the computation of the adaptive thresholds, may not be valid under faulty conditions. This issue appears due to the fact that the domain of approximation of the modeling uncertainty may change under faulty conditions (especially under the occurrence of multiple sensor faults), when the system is operating in closed-loop and the faulty sensors are used for the implementation of the control scheme.

8

Concluding Remarks and Open Issues

The objective of this tutorial was the presentation of a methodology for designing and analyzing a sensor fault diagnosis method for a network of interconnected cyber-physical systems (CPS), where each system may be characterized by simple dynamics, but the overall dynamics can be large-scale and complex. In this work we considered that a large number of sensors is used by the cyber core to the monitor, control and coordinate physical part (interconnected CPS). Giving emphasis on the large-scale, nonlinear nature of the physical part and the isolation of multiple sensor faults, the methodology was developed following a distributed and decentralized architecture, while a centralized architecture was proposed for the case that the network of CPS is treated as a monolithic entity. In a decentralized architecture, a monitoring agent is designed for each of the interconnected CPS, while the agents do not exchange any data-driven information (sensor measurements, decisions), but use *a priori known* reference signals related to the system interconnections, taking advantage of the available knowledge that stems from the control design. In a distributed approach, we proposed the deployment of monitoring agents, where every agent is dedicated to one of the interconnected CPS, while it is allowed to exchange sensor

information with its neighboring monitoring agents.

The common design feature of the three architectures is the decomposition of the monitoring of the set of sensors of each CPS into several modules that monitor subsets of sensors that may be disjoint or overlapping. The decomposition is realized to facilitate the isolation of multiple sensor faults, which can be extremely difficult by using a single module, especially when the number of sensors is large. The sensor fault detection process is executed by a monitoring agent through the monitoring modules designed to detect the occurrence of sensor faults in their associated sensor sets. A monitoring module makes a decision by comparing some observer-based residuals to their corresponding adaptive thresholds. The monitoring agent can isolate multiple sensor faults, which affect a CPS locally, using an aggregation module that collects and processes combinatorially the decisions of the monitoring modules. In the distributed architecture, a global aggregation module is used to isolate multiple sensor faults that are propagated in the network of CPS, by combining the decisions of monitoring agents.

In this research work, the main criteria for the decomposition of the set of all sensors used in the network of CPS into subsets of sensors are related to certain conditions that guarantee the stability of the nonlinear estimation model used in the residuals generated in every monitoring module, and the isolability of multiple sensor faults of the aggregation module in the monitoring agent. The decomposition procedure in a centralized scheme can be more complicated compared to the decomposition procedure in a non-centralized scheme, due to the possible large number of sensors and the large scale of the system. Note that in a non-centralized scheme, the decomposition is applied to the set of sensors used in each of the interconnected CPS, and not to all sensors of the network of CPS.

The decomposition of the sensor set for large-scale diagnosis is an optimization problem for which it is very hard to find the optimal solution. The analysis of the structural model of CPS can be very useful, and several algorithms have been proposed for sensor placement that could be used for the decomposition problem as well [Commault and Dion [2007], Krysander and Frisk [2008], Frisk et al. [2012], Blanke et al.

[2016]]. The structural model represents only the known relationships between variables without taking into account the value of the parameters, or analytical expressions. The decomposition of the sensor set in a CPS could be combined with the decomposition of the CPS itself [Krysander et al. [2008]]. In this case, the monitoring modules could be designed assuming several decomposed subsystems. However, the effectiveness of the decomposition of the sensor set for robust fault diagnosis can be attenuated due to modeling errors, measurement noise and system disturbances. Thus, it is necessary to perform the decomposition aiming at optimizing the fault detectability and isolability under uncertainties [Ferrari et al. [2009], Eriksson et al. [2013], Boem et al. [2015a]]. This was the motivation for designing the decomposition algorithm proposed in this tutorial. From a practical point of view, the decomposition process could be enhanced by exploiting further information related to the sensors, such as their accuracy and reliability [Figueras et al. [2006]], as well as their spatial distribution, especially in a centralized architecture where the sensor data should be transmitted to the monitoring modules. For instance, sensors that are located in the same geographical region can be grouped together.

In the proposed methodology, the main role of the monitoring modules is to detect the occurrence of sensor faults in subsets of sensors. The fault detection decision making process relies on checking the validity of a set of analytical redundancy relations of observer-based residuals and adaptive thresholds. The residuals are designed as the difference between the measured system outputs and their estimations provided by a nonlinear observer. The adaptive thresholds are designed to bound the residuals under healthy conditions. In a centralized scheme the residuals are generated based on a nonlinear Lipschitz observer. In a distributed or decentralized scheme, this observer is properly customized to take into account the known nonlinear interconnection dynamics.

In a centralized scheme, the conditions for guaranteeing the nonlinear observer stability can be more strict due to the possible large system scale and large number of sensors. This issue may impact the decomposition procedure as previously discussed. The conditions are relaxed following a non-centralized architecture. In a decentralized scheme, the

state estimation error (and consequently the residual) under healthy conditions may be increased in the case that the controllers of the neighboring subsystems do not preserve a good tracking performance under healthy and faulty conditions, i.e. the tracking error is high. The state estimation error can be lower in a distributed architecture, due to the sensor information transmitted by the neighboring agents. Nevertheless, the estimation process can be directly affected by sensor faults in neighboring cyber-physical subsystems. An alternative distributed observer structure would be a Lipschitz observer, where the interconnection dynamics are function of the estimated interconnection state vector transmitted by neighboring agents. In this case, given that every neighboring agent consists of several monitoring modules that can provide an estimation of the subsystem's state. Therefore, an additional mechanism would be needed to determine the state estimate with the smaller estimation error, or to fuse the state estimates before transmitting to neighboring agents.

Irrespective of the adopted architecture, the sensor fault detection decision logic is based on the design properties of the analytical redundancy relations of residuals and adaptive thresholds, that is structured robustness and fault sensitivity. The structure robustness implies that every monitoring module is insensitive to modeling uncertainties and the measurement noise affecting the associated local sensor set, and the transmitted sensor information in the case of a distributed scheme. The structured sensor fault sensitivity entails that every monitoring module is implicitly (structurally) sensitive to faults affecting the underlying local sensor set and the transmitted sensor information, in a distributed scheme. Based on these properties, we formulated the boolean sensor fault detection decision function and developed the binary sensor fault signature matrix. In this way, the set of analytical redundancy relations is considered to have *uniform sensitivity* to all sensor fault combinations, and for this reason the diagnosis set may contain two or more sensor fault combinations that have possibly occurred. Thus, a sensor fault signature matrix that contains more than binary information would contribute to the reduction of the number of diagnosed sensor fault combinations and the enhancement the isolation

process [Meseguer et al. [2010]]. However, such sensor fault signature matrix may require more information about the sensor fault characteristics (e.g. the time of sensor fault occurrence) than the structured sensor fault sensitivity.

The performance of the proposed sensor fault diagnosis method was analyzed with respect of the sensor fault detectability of the monitoring modules, as well as the multiple sensor fault isolability of the monitoring agents. In the sensor fault detectability analysis, the goal was to analyze the minimum sensor fault effects that are guaranteed to be detected based on the designed adaptive thresholds and under worst case assumptions of bounded modeling uncertainties, i.e. assuming that the modeling uncertainties have reached their limits. The sensor fault detectability analysis is a way to evaluate explicitly the sensitivity of the analytical redundancy relations of residuals and adaptive thresholds of a monitoring module. However, the sensitivity may change over time. Particularly, the local sensor fault effects on the set of ARRs may disappear after some time interval, although some faults in the local sensor sets may be active for ever [Chen and Patton [1999]]. Both the detection and isolation procedure would be improved if we analyze the strong detectability conditions, i.e. the conditions under which the sensor fault effects are guaranteed to be persistent as long as the fault is active [Reppa et al. [2015a]]. Moreover, the results of the performance analysis under worst case assumptions of uncertainties can be exploited for ensuring the dependability of the sensor fault diagnosis architectures, that is the ability to deliver service that can justifiably be trusted [Avižienis et al. [2004]]. Through this analysis, it will be possible to characterize the guaranteed detectable and non-detectable faults, or confirm that a set of known faults (determined based on historical data or vulnerability analysis) are guaranteed detectable or non-detectable, given a set of parameters for the design of the proposed diagnosis method and worst-case bounds of uncertainty.

Along with the performance analysis of the model-based sensor fault diagnosis techniques, we proposed the application of an adaptive approximation methodology that aims to learn the modeling uncertainty that is a function of the system state and input under healthy condi-

tions. Then, the resultant nonlinear approximator is used in the sensor fault diagnosis technique, both in the residual generation and the computation of the adaptive thresholds. In this way, the effects of the modeling uncertainty on the residuals and the thresholds can be reduced, enhancing the fault detectability. The adaptive approximation methodology can contribute significantly to sensor fault diagnosis, when the prior knowledge of the used model is limited. However, under faulty conditions the difference between the approximator and the actual uncertainty can be changed due to a change in the domain of approximation defined by the compact regions of stability of the system state and input. This is commonly observed when the system is operating in closed-loop and the sensor measurements are directly fed back to the controller. A possible solution to avoid the change in the domain of approximation is to intervene in the closed loop operation by either applying adaptive approximation based control [Zhang et al. [2004], Panagi and Polycarpou [2011, 2013]], or observer-based control where the state estimate results from a switching reconfiguration (i.e. use the state estimate with the smaller estimation error) or fusion [Seron et al. [2012], Stoican and Olaru [2013]].

A significant open issue in fault diagnosis is the design of a unified methodology for isolating multiple, sensor and process (plant and actuator) faults in a network of CPS. The challenge in this problem is to distinguish the possibly compensated effects of sensor and process faults on the generated residuals and adaptive thresholds. The proposed methodology offers a framework for incorporating additional agents (modules), which can be designed to be robust and structurally sensitive to subsets of process faults or, combinations of sensor and process faults. In a unified methodology, the heterogeneous design of the agents can be efficient in the sense that some agents dedicated to diagnose some parametric faults can be designed following a system identification-based approach instead of an observer-based approach. Following an heterogeneous design approach may require more than one model of the cyber-physical systems, e.g. input-output model, state space. Then a fault isolation decision logic should be formulated, after analyzing the detectability and isolability of multiple, heterogeneous

faults.

An interesting research problem in fault diagnosis is the design of a methodology for diagnosing multiple sensor (and maybe process) faults under communication imperfections that may affect the cyber core in a network of CPS [Boem et al. [2013b]]. The challenge in this problem is to ensure that the decision-making process of the cyber agents (modules) is robust against time and transmission delays, packet dropouts, which can be handled as systems uncertainties on one hand, and on the other hand to discriminate between sensor and communication faults such as transmission errors in physical links, network drop-out or even cyber attacks.

The difference between a system and a cyber-physical system is the presence of the cyber part, and the fact that both the physical and the cyber system are treated as one system. When the cyber part is responsible for the control of the physical part, we cannot ignore the closed-loop operation of CPS. The majority of the fault diagnosis methods are passive in the sense that the fault diagnosis agents (modules) are usually monitoring the input and output of the underlying system without intervening in the closed-loop operation. In this work, we exploit some information related to the closed-loop operation of the system for designing the decentralized architecture by using the knowledge of the reference signals (passive way). The closed-loop operation can be further exploited in the fault diagnosis process actively, by designing some input signals that can be injected in the system in order to guarantee the detectability and isolability of the faults [Eshna Ashari et al. [2012], Scott et al. [2014]]. In a network of cyber-physical systems, using a set of input auxiliary signals can help in discriminating the effects of various faults that can be sensor, process or communication faults, which can be local or have been propagated by neighboring subsystems. The active fault diagnosis can be valuable for performing preventive maintenance during the operational life of a CPS.

Acknowledgements

This work is funded by the European Research Council Advanced Grant (ERC-AdG-291508) ‘FAULT-ADAPTIVE: Fault-Adaptive Monitoring and Control of Complex Distributed Dynamical Systems’, and the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n° [626891] ‘FUTuRISM: Multiple sensor FaUlt ToleRant control for management of Interconnected nonlinear SystemeMs’.

Appendices

A

Illustrative Example: Design Parameters

Tables A.1-A.4 summarize the parameters used for the design of the observer-based residuals and adaptive thresholds of the monitoring modules $\mathcal{M}^{(1,1)}$, $\mathcal{M}^{(1,2)}$ and $\mathcal{M}^{(2,1)}$, $\mathcal{M}^{(2,2)}$ of the illustrative example presented in Section 4.5.

		Design Parameter Set		
		1	2	3
$\mathcal{O}^{(1,1)}$	$\hat{x}^{(1,1)}(0)$	$[0, 0]^\top$	$[0, 0]^\top$	$[0, 0]^\top$
	$\text{eig}(A_L^{(1,1)})$	$\{-2, -3\}$	$\{-5, -6\}$	$\{-10, -11\}$
	$L^{(1,1)}$	$[3, 9]^\top$	$[9, 21]^\top$	$[19, 81]^\top$
$\bar{\varepsilon}_{y_1}^{(1,1)}$	$\rho_0^{(1,1)}$	1.6567	1.1196	1.0966
	$\xi_0^{(1,1)}$	1.5	3.75	7.5
	$\rho^{(1,1)}$	0.2579	0.125	0.0704
	$\xi^{(1,1)}$	1	2.5	5
	$\rho_d^{(1,1)}$	3.4	10.8055	19.7374
	$\xi_d^{(1,1)}$	1	4.75	5

Table A.1: Design parameters of the monitoring module $\mathcal{M}^{(1,1)}$.

		Design Parameter Set			
		1	2	3	4
$\mathcal{O}^{(1,2)}$	$\hat{x}^{(1,2)}(0)$	$[0, 0]^\top$	$[0, 0]^\top$	$[0, 0]^\top$	$[0, 0]^\top$
	$\text{eig}(A_L^{(1,2)})$	$\{-2, -3\}$	$\{-10, -11\}$	$\{-14, -15\}$	$\{-20, -21\}$
	$L^{(1,2)}$	$[1.6667, 3]^\top$	$[13.22, 19]^\top$	$[24.3333, 27]^\top$	$[47.6667, 39]^\top$
$\bar{\varepsilon}_{y_q}^{(1,2)}$	$\rho_0^{(1,2)}$	1.2590	4.6383	1.6	1.6
	$\xi_0^{(1,2)}$	1	7	6.3	7
	$\rho^{(1,2)}$	0.1775	0.8333	1.1666	1.6666
	$\xi^{(1,2)}$	1	5	7	10
	$\rho_d^{(1,2)}$	2.0237	19.9720	25.2561	49.2621
	$\xi_d^{(1,2)}$	1.6	7.5	10.5	15
	$\alpha_0^{(1,2)}$	3.5937	1.8409	1.386	1.2216
	$\zeta_0^{(1,2)}$	1.5	7.5	10.5	15
	$\alpha^{(1,2)}$	1	1	1	1
	$\zeta^{(1,2)}$	1	5	7	10
	$\alpha_d^{(1,2)}$	4.8672	19.4554	27.1184	40.5659
	$\zeta_d^{(1,2)}$	1.5	7.5	10.5	15

Table A.2: Design parameters of the monitoring module $\mathcal{M}^{(1,2)}$.

		Design Parameter Set		
		1	2	3
$\mathcal{O}^{(2,1)}$	$\hat{x}^{(2,1)}(0)$	$[0, 0]^\top$	$[0, 0]^\top$	$[0, 0]^\top$
	$\text{eig}(A_L^{(2,1)})$	$\{-5, -6\}$	$\{-14, -15\}$	$\{-20, -21\}$
	$L^{(2,1)}$	$[0, 42]^\top$	$[18, 24]^\top$	$[30, 102]^\top$
$\bar{\varepsilon}_{y_1}^{(2,1)}$	$\rho_0^{(2,1)}$	2.0704	1.6568	1
	$\xi_0^{(2,1)}$	3.75	10.5	10
	$\rho^{(2,1)}$	0.125	0.0667	0.0357
	$\xi^{(2,1)}$	2.5	7	10
	$\rho_d^{(2,1)}$	3.6	21.0897	30.8591
	$\xi_d^{(2,1)}$	1.25	13.3	10

Table A.3: Design parameters of the monitoring module $\mathcal{M}^{(2,1)}$.

Design Parameter Set					
$\mathcal{O}^{(2,2)}$	1	2	3	4	5
	$\hat{x}^{(2,2)}(0)$	$[0, 0]^T$	$[0, 0]^T$	$[0, 0]^T$	$[0, 0]^T$
$\bar{\varepsilon}_{y_q}^{(2,2)}$	$\text{eig}(A_L^{(2,2)})$	$\{-2, -3\}$	$\{-5, -6\}$	$\{-14, -15\}$	$\{-20, -21\}$
	$L^{(2,2)}$	$[06, 1.5]^T$	$[0, 3.5]^T$	$[18.5, 18]^T$	$[30, 36]^T$
	$\rho_0^{(2,2)}$	1.2	1.2	5.0804	1
	$\xi_0^{(2,2)}$	0.5	1.25	10.5	5
	$\rho^{(2,2)}$	0.5	0.5	0.875	2
	$\xi^{(2,2)}$	1.4	3.5	7	14
	$\rho_d^{(2,2)}$	5.5	5.5	21.0155	36
	$\xi_d^{(2,2)}$	1.4	3.5	13.3	15
	$\alpha_0^{(2,2)}$	3	1.4995	1.8	1.3538
	$\zeta_0^{(2,2)}$	0.9	2.5	10.5	15
	$\alpha^{(2,2)}$	1	1	1	1
	$\zeta^{(2,2)}$	1.5	3.75	7	10
	$\alpha_d^{(2,2)}$	7	5.5	18.2777	30
	$\zeta_d^{(2,2)}$	1	2.5	13.3	15

Table A.4: Design parameters of the monitoring module $\mathcal{M}^{(2,2)}$.

References

- K. Adjallah, D. Maquin, and J. Ragot. Non-linear observer-based fault detection. In *the 3rd IEEE Conference on Control Applications*, pages 1115–1120, 1994.
- C. Alippi, S. Ntalampiras, and M. Roveri. A cognitive fault diagnosis system for distributed sensor networks. *IEEE Transactions on Neural Networks and Learning Systems*, 24:1213–1226, 2013.
- S. Amin, X. Litrico, S. Sastry, and A. M Bayen. Cyber security of water SCADA systems–Part i: analysis and experimentation of stealthy deception attacks. *IEEE Transactions on Control Systems Technology*, 21(5):1963–1970, 2013.
- P. J. Antsaklis, B. Goodwine, V. Gupta, M. J. McCourt, Y. Wang, P. Wu, M. Xia, H. Yu, and F. Zhu. Control of cyberphysical systems using passivity and dissipativity based methods. *European Journal of Control*, 19(5):379 – 388, 2013.
- A. Avižienis, J.-C. Laprie, and B. Randell. Dependability and its threats: a taxonomy. In *Building the Information Society*, pages 91–120. Springer, 2004.
- BEA. Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris. Technical report, Bureau d’Enquêtes et d’Analyses pour la sécurité de l’aviation civile (BEA), 2012.
- G. Betta and A. Pietrosanto. Instrument fault detection and isolation: State of the art and new research trends. *IEEE Transactions on Instrumentation and Measurement*, 49(1):100–107, 2000.

- M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki. *Diagnosis and fault-tolerant control*. Springer-Verlag Berlin Heidelberg, 2016.
- F. Boem, R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou. Distributed fault diagnosis for continuous-time nonlinear systems: The input-output case. *Annual Reviews in Control*, 37(1):163 – 169, 2013a.
- F. Boem, R.M.G. Ferrari, T. Parisini, and M.M. Polycarpou. Distributed fault detection for uncertain nonlinear systems: A network delay compensation strategy. In *American Control Conference (ACC), 2013*, pages 3549–3554, 2013b.
- F. Boem, Y. Xu, C. Fischione, and T. Parisini. Distributed fault detection using sensor networks and pareto estimation. In *the 12th European Control Conference (ECC)*, pages 932–937, 2013c.
- F. Boem, R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou. Optimal topology for distributed fault detection of large-scale systems. pages 60–65, 2015a.
- F. Boem, S. Rivero, G. Ferrari-Trecate, and T. Parisini. A plug-and-play fault diagnosis approach for large-scale systems. pages 601–606, 2015b.
- A. Bregon, G. Biswas, and B. Pulido. A decomposition method for nonlinear parameter estimation in transcend. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Human*, 42(3):751–763, 2012.
- F. Caccavale, F. Pierri, and L. Villani. Adaptive observer for fault diagnosis in nonlinear discrete-time systems. *ASME Journal of Dynamic Systems, Measurement, and Control*, 130(2):021005–1–9, 2008.
- F. Caccavale, P. Cilibrizzi, F. Pierri, and L. Villani. Actuators fault diagnosis for robot manipulators with uncertain model. *Control Engineering Practice*, 17(1):146–157, 2009.
- J. Chen and R. J. Patton. *Robust Model-based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers, 1999.
- W. Chen and J. Li. Decentralized output-feedback neural control for systems with unknown interconnections. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 38(1):258–266, 2008.
- W. Chen and M. Saif. A sliding mode observer-based strategy for fault detection, isolation, and estimation in a class of lipschitz nonlinear systems. *International Journal of Systems Science*, 38(12):943–955, 2007.
- H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, and H. J. Scholl. Understanding smart cities: An integrative framework. In *the 45th Hawaii International Conference on System Science (HICSS)*, pages 2289–2297, 2012.

- C. Commault and J.-M. Dion. Sensor location for diagnosis in linear systems: a structural analysis. *IEEE Transactions on Automatic Control*, 52(2):155–169, 2007.
- M. O. Cordier, P. Dague, F. Lévy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès. Conflicts versus analytical redundancy relations: A comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(5):2163–2177, 2004.
- M. Daigle, X. D. Koutsoukos, and G. Biswas. Distributed diagnosis in formations of mobile robots. *IEEE Transactions on Robotics*, 23(2):353–369, 2007.
- M. Daigle, A. Bregon, G. Biswas, X. Koutsoukos, and B. Pulido. Improving multiple fault diagnosability using possible conflicts. In *the 8th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, pages 144–149, Mexico City, Mexico, 2012.
- M. R. Davoodi, K. Khorasani, H. A. Talebi, and H. R. Momeni. Distributed fault detection and isolation filter design for a network of heterogeneous multiagent systems. *IEEE Transactions on Control Systems Technology*, 22(3):1061–1069, 2014.
- J. De Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial intelligence*, 32(1):97–130, 1987.
- C. De Persis and A. Isidori. A geometric approach to nonlinear fault detection and isolation. *IEEE Transactions on Automatic Control*, 46(6):853–865, 2001.
- S. X. Ding. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer-Verlag London, 2008.
- Y. Ding, E. A. Elsayed, S. Kumara, J. C. Lu, F. Niu, and J. Shi. Distributed sensing for quality and productivity improvements. *IEEE Transactions on Automation Science and Engineering*, 3(4):344–359, 2006.
- M. Du, J. Scott, and P. Mhaskar. Actuator and sensor fault isolation of nonlinear process systems. *Chemical Engineering Science*, 104:294–303, 2013.
- D. Düstegör, E. Frisk, V. Cocquempot, M. Krysander, and M. Staroswiecki. Structural analysis of fault isolability in the damadics benchmark. *Control Engineering Practice*, 14(6):597–608, 2006.

- Daniel Eriksson, Erik Frisk, and Mattias Krysander. A method for quantitative fault diagnosability analysis of stochastic linear descriptor models. *Automatica*, 49(6):1591–1600, 2013.
- A. Eshna Ashari, R. Nikoukhah, and S. L. Campbell. Active robust fault detection in closed-loop systems: quadratic optimization approach. *IEEE Transactions on Automatic Control*, 57(10):2532–2544, 2012.
- J. Farrell and M. M. Polycarpou. *Adaptive approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. Wiley-Interscience, 2006.
- H. Ferdowsi, D. L. Raja, and S. Jagannathan. A decentralized fault detection and prediction scheme for nonlinear interconnected continuous-time systems. In *the 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2012.
- R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou. Distributed fault diagnosis with overlapping decompositions: an adaptive approximation approach. *IEEE Transactions on Automatic Control*, 54:794–799, 2009.
- R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou. Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach. *IEEE Transactions on Automatic Control*, 57(2):275–290, 2012.
- J. Figueras, V. Puig, and J. Quevedo. Multiple fault diagnosis system design using reliability analysis: Application to barcelona rain-gau. In *the 6th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, volume 6, pages 1324–1329, 2006.
- E. Frisk, A. Bregon, J. Aslund, M. Krysander, B. Pulido, and G. Biswas. Diagnosability analysis considering causal interpretations for differential constraints. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 42(5):1216–1229, 2012.
- R. A. García and C. E. D’Attelis. Trajectory tracking in nonlinear systems via nonlinear reduced-order observers. *International Journal of Control*, 62(3):685–715, 1995.
- S. Gentil, J. Montmain, and C. Combastel. Combining FDI and AI approaches within causal-model-based diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(5):2207–2221, 2004.
- J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. CRC, 1998.

- N. Hovakimyan, E. Lavretsky, B. J. Yang, and A. J. Calise. Coordinated decentralized adaptive output feedback control of interconnected systems. *IEEE Transactions on Neural Networks*, 16(1):185–194, 2005.
- S. Indra, E. Chanthery, and L. Travé-Massuyès. Decentralized diagnosis with isolation on request for spacecraft. In *the 8th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, pages 283–288, Mexico City, Mexico, 2012.
- P. A. Ioannou and J. Sun. *Robust Adaptive Control*. Prentice-Hall, 1995.
- R. Isermann. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer Verlag, 2006.
- I. Issury and D. Henry. Multiple and simultaneous fault isolation with minimal fault indicating signals: a case study. In *the 7th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, pages 59–64, Barcelona, Spain, 2009.
- P. Kabore and H. Wang. Design of fault diagnosis filters and fault-tolerant control for a class of nonlinear systems. *IEEE Transactions on Automatic Control*, 46(11):1805–1810, 2001.
- S. Klinkhieo, R. J. Patton, and C. Kambhampati. Robust FDI for FTC coordination in a distributed network system. In *the 16th IFAC World Congress*, pages 13551–13556, 2008.
- J. M. Koscielny, M. Bartys, and M. Syfert. Method of multiple fault isolation in large scale systems. *IEEE Transactions on Control Systems Technology*, 20(5):1302–1310, 2012.
- E. B. Kosmatopoulos, M. M. Polycarpou, M. A. Christodoulou, and P. A. Ioannou. High-order neural network structures for identification of dynamical systems. *IEEE Transactions on Neural Networks*, 6(2):422–431, 1995.
- W. Kröger and E. Zio. *Vulnerable Systems*. Springer Verlag, 2011.
- M. Krysander and E. Frisk. Sensor placement for fault diagnosis. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 38(6):1398–1410, 2008.
- M. Krysander, J. Aslund, and M. Nyberg. An efficient algorithm for finding minimal overconstrained subsystems for model-based diagnosis. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 38(1):197–206, 2008.
- S. Lipschutz and M. Lipson. *Linear algebra*. Schaum’s Series, 2008.

- R. Mattone and A. De Luca. Nonlinear fault detection and isolation in a three-tank heating system. *IEEE Transactions on Control Systems Technology*, 14(6):1158–1166, 2006.
- J. Meseguer, V. Puig, and T. Escobet. Fault diagnosis using a timed discrete-event approach based on interval observers: application to sewer networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 40:900–916, 2010.
- M. Milanese and C. Novara. Set membership identification of nonlinear systems. *Automatica*, 40:957–975, 2004.
- P. Mohajerin Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. Cyber attack in a two-area power system: Impact identification using reachability. In *the 2010 American Control Conference (ACC)*, pages 962–967, 2010.
- A. Monteriu, P. Asthan, K. Valavanis, and S. Longhi. Model-based sensor fault detection and isolation system for unmanned ground vehicles: theoretical aspects (Part I). In *the 2007 IEEE International Conference on Robotics and Automation*, pages 2736–2743, 2007.
- S. Narasimhan, P. Vachhani, and R. Rengaswamy. New nonlinear residual feedback observer for fault diagnosis in nonlinear systems. *Automatica*, 44(9):2222–2229, 2008.
- M. Nyberg. A fault isolation algorithm for the case of multiple faults and multiple fault types. In *the 6th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, Beijing, China, 2006.
- 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*, 83:2622–2640, 2010.
- P. Panagi and M. M. Polycarpou. Distributed fault accommodation for a class of interconnected nonlinear systems with partial communication. *IEEE Transactions on Automatic Control*, 56(12):2962–2967, 2011.
- P. Panagi and M. M. Polycarpou. A coordinated communication scheme for distributed fault tolerant control. *IEEE Transactions on Industrial Informatics*, 9(1):386–393, 2013.
- M. M. Polycarpou and A. J. Helmicki. Automated fault detection and accommodation: a learning systems approach. *IEEE Transactions on Systems, Man and Cybernetics*, 25(11):1447–1458, 1995.

- M. M. Polycarpou and A. B. Trunov. Learning approach to nonlinear fault diagnosis: detectability analysis. *IEEE Transactions on Automatic Control*, 45(4):806–812, 2000.
- V. Puig, A. Stancu, and J. Quevedo. Robust fault isolation using nonlinear interval observers: the DAMADICS benchmark case study. In *the 16th IFAC World Congress*, pages 1850–1855, 2005.
- V. Puig, J. Gertler, J. Figueras, and J. Quevedo. Design of structured residuals using interval models: Application to multiple sequential fault isolation. In *the 6th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, pages 914–919, Beijing, China, 2006.
- B. Pulido and C. A. González. Possible conflicts: a compilation technique for consistency-based diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(5):2192–2206, 2004.
- T. Raïssi, Gaétan Videau, and A. Zolghadri. Interval observer design for consistency checks of nonlinear continuous-time systems. *Automatica*, 46(3):518–527, 2010.
- R. Rajamani. Observers for lipschitz nonlinear systems. *IEEE Transactions on Automatic Control*, 43(3):397–401, 1998.
- R. Rajamani and A. Ganguli. Sensor fault diagnostics for a class of non-linear systems using linear matrix inequalities. *International Journal of Control*, 77(10):920–930, 2004.
- S. Rajaraman, J. Hahn, and M. S. Mannan. A methodology for fault detection, isolation, and identification for nonlinear processes with parametric uncertainties. *Industrial & Engineering Chemistry Research*, 43(21):6774–6786, 2004.
- S. Rajaraman, J. Hahn, and M. S. Mannan. Sensor fault diagnosis for non-linear processes with parametric uncertainties. *Journal of Hazardous Materials*, 130(1-2):1–8, 2006.
- S. Reece, S. Roberts, C. Claxton, and D. Nicholson. Multi-sensor fault recovery in the presence of known and unknown faults. In *the 12th International Conference on Information Fusion*, pages 1695–1703, 2009.
- V. Reppa, M. M. Polycarpou, and C. G. Panayiotou. A distributed architecture for sensor fault detection and isolation using adaptive approximation. In *the IEEE World Congress on Computational Intelligence, 2012*, pages 2340–2347, Brisbane, Australia, 2012.

- V. Reppa, P. Papadopoulos, M. M. Polycarpou, and C. G. Panayiotou. A distributed virtual sensor scheme for smart buildings based on adaptive approximation. In *the 2014 International Joint Conference on Neural Networks (IJCNN)*, pages 99–106, 2014a.
- V. Reppa, M. M. Polycarpou, and C. G. Panayiotou. Adaptive approximation for multiple sensor fault detection and isolation of nonlinear uncertain systems. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):137–153, 2014b.
- 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)*, pages 1485–1490, 2015a.
- V. Reppa, M. M. Polycarpou, and C. G. Panayiotou. Decentralized isolation of multiple sensor faults in large-scale interconnected nonlinear systems. *IEEE Transactions on Automatic Control*, 60(6):1582 – 1596, 2015b.
- V. Reppa, M. M. Polycarpou, and C. G. Panayiotou. Distributed sensor fault diagnosis for a network of interconnected cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 2:11–23, 2015c.
- K. Salahshoor, M. Mosallaei, and M. Bayat. Centralized and decentralized process and sensor fault monitoring using data fusion based on adaptive extended kalman filter algorithm. *Measurement*, 41(10):1059–1076, 2008.
- I. Samy, I. Postlethwaite, and D. W. Gu. Survey and application of sensor fault detection and isolation schemes. *Control Engineering Practice*, 19: 658–674, 2011.
- J. K. Scott, R. Findeisen, R. D. Braatz, and D. M. Raimondo. Input design for guaranteed fault diagnosis using zonotopes. *Automatica*, 50(6):1580 – 1589, 2014.
- M. M. Seron, J. A. De Doná, and S. Olaru. Fault tolerant control allowing sensor healthy-to-faulty and faulty-to-healthy transitions. *IEEE Transactions on Automatic Control*, 57:1657–1669, 2012.
- M. M. Seron, J. A. De Doná, and J. H. Richter. Integrated sensor and actuator fault-tolerant control. *International Journal of Control*, 86(4):689–708, 2013.
- I. Shames, A. M. H. Teixeira, H. Sandberg, and K. H. Johansson. Distributed fault detection for interconnected second-order systems. *Automatica*, 47 (12):2757–2764, 2011.

- L. Sherry and R. Mauro. Controlled flight into stall (cfis): Functional complexity failures and automation surprises. In *the Integrated Communications, Navigation and Surveillance Conference (ICNS), 2014*, pages D1–11, 2014.
- M. Staroswiecki and G. Comtet-Varga. Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica*, 37: 687–699, 2001.
- Florin Stoican and Sorin Olaru. *Fault Detection and Isolation in Multisensor Systems*. Wiley & Sons, Inc., 2013.
- H. A. Talebi, K. Khorasani, and S. Tafazoli. A recurrent neural-network-based sensor and actuator fault detection and isolation for nonlinear systems with application to the satellite’s attitude control subsystem. *IEEE Transactions on Neural Networks*, 20(1):45–60, 2009.
- B. T. Thumati and S. Jagannathan. A model-based fault-detection and prediction scheme for nonlinear multivariable discrete-time systems with asymptotic stability guarantees. *IEEE Transactions on Neural Networks*, 21(3): 404–423, 2010.
- L. Travé-Massuyès. Bringing technologies for diagnosis. In *the 8th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, pages 361–372, Mexico City, Mexico, 2012.
- L. Travé-Massuyès. Bridging control and artificial intelligence theories for diagnosis: A survey. *Engineering Applications of Artificial Intelligence*, 27: 1–16, 2014.
- L. Travé-Massuyès, T. Escobet, and X. Olive. Diagnosability analysis based on component-supported analytical redundancy relations. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 36(6): 1146–1160, 2006.
- A. B. Trunov and M. M. Polycarpou. Automated fault diagnosis in nonlinear multivariable systems using a learning methodology. *IEEE Transactions on Neural Networks*, 11(1):91–101, 2000.
- Z. Uykan, C. Guzelis, M. E. Celebi, and H. N. Koivo. Analysis of input-output clustering for determining centers of RBFN. *IEEE Transactions on Neural Networks*, 11(4):851–858, 2000.
- H. Veda and V. Venkatasubramanian. Signed digraph based multiple fault diagnosis. *Computers & Chemical Engineering*, 21:S655–S660, 1997.
- A.T. Vemuri and M.M. Polycarpou. Robust nonlinear fault diagnosis in input-output systems. *International Journal of Control*, 68(2):343–360, 1997.

- V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin. A review of process fault detection and diagnosis Part I: Quantitative model-based methods. *Computers & chemical engineering*, 27(3):327–346, 2003a.
- V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin. A review of process fault detection and diagnosis Part II: Qualitative models and search strategies. *Computers & chemical engineering*, 27(3):327–346, 2003b.
- A. Xu and Q. Zhang. Nonlinear system fault diagnosis based on adaptive estimation. *Automatica*, 40(7):1181–1193, 2004a.
- A. Xu and Q. Zhang. Residual generation for fault diagnosis in linear time-varying systems. *IEEE Transactions on Automatic Control*, 49(5):767–772, 2004b.
- X. G. Yan and C. Edwards. Sensor fault detection and isolation for nonlinear systems based on a sliding mode observer. *International Journal of Adaptive Control and Signal Processing*, 21(8-9):657–673, 2007.
- X. G. Yan and C. Edwards. Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer. *International Journal of Control*, 81(4):591–606, 2008.
- W. Yu and J. de Jesús Rubio. Recurrent neural networks training with stable bounding ellipsoid algorithm. *IEEE Transactions on Neural Networks*, 20(6):983–991, 2009.
- Q. Zhang and X. Zhang. Distributed sensor fault diagnosis in a class of interconnected nonlinear uncertain systems. *Annual Reviews in Control*, 37:170 – 179, 2013.
- Q. Zhang, X. Zhang, M. M. Polycarpou, and T. Parisini. Distributed sensor fault detection and isolation for multimachine power systems. *International Journal of Robust and Nonlinear Control*, 24(8-9):1403–1430, 2014.
- X. Zhang. Sensor bias fault detection and isolation in a class of nonlinear uncertain systems using adaptive estimation. *IEEE Transactions on Automatic Control*, 56(5):1220 –1226, 2011.
- X. Zhang and Q. Zhang. Distributed fault detection and isolation in a class of large-scale nonlinear uncertain systems. In *the 18th IFAC World Congress*, pages 4302–4307, Milan, Italy, 2011.
- X. Zhang, M. M. Polycarpou, and T. Parisini. A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. *IEEE Transactions on Automatic Control*, 47(4):576–593, 2002.
- X. Zhang, T. Parisini, and M. M. Polycarpou. Adaptive fault-tolerant control of nonlinear uncertain systems: an information-based diagnostic approach. *IEEE Transactions on Automatic Control*, 49(8):1259–1274, 2004.

- X. Zhang, T. Parisini, and M. M. Polycarpou. Sensor bias fault isolation in a class of nonlinear systems. *IEEE Transactions on Automatic Control*, 50(3):370–376, 2005.
- 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*, 32(1):107–121, 2008.
- Y. Zhang and X. R. Li. Detection and diagnosis of sensor and actuator failures using imm estimator. *IEEE Transactions on Aerospace and Electronic Systems*, 34(4):1293–1313, 1998.
- F. Zhu and Z. Han. A note on observers for lipschitz nonlinear systems. *IEEE Transactions on Automatic Control*, 47(10):1751–1754, 2002.