2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

A Distributed Fault Detection Methodology for a Class of Large-scale Uncertain Input-output Discrete-Time Nonlinear Systems

Francesca Boem, Riccardo M. G. Ferrari, Thomas Parisini, and Marios M. Polycarpou

Abstract— **This paper extends very recent results on a distributed fault diagnosis methodology for nonlinear uncertain large-scale discrete-time dynamical systems to the case of partial state measurement. The large scale system being monitored is modeled, following a *divide et impera* approach, as the interconnection of several subsystems that are allowed to overlap sharing some state components. Each subsystem has its own Local Fault Diagnoser: the local detection is based on the knowledge of the local subsystem dynamic model and of an adaptive approximation of the interconnection with neighboring subsystems. A consensus-based estimator is used in order to improve the detectability of faults affecting variables shared among different subsystems. Time-varying threshold functions guaranteeing no false-positive alarms and analytical fault detectability sufficient conditions are presented as well.**

I. Introduction

In recent years, the need to develop autonomous and intelligent systems that operate reliably in the presence of system faults has increased, motivating the research on automated fault diagnosis and accommodation. In dynamical systems, faults are characterized by critical and unpredictable changes in the system dynamics, thus requiring the design of suitable fault diagnosis schemes [1], [2], [3]. Model-based schemes have emerged as prominent approaches to fault diagnosis of continuous and discrete-time systems [4], [5], [3], [6], [7], [2]. This approach is built on a mathematical model of the process that must be monitored, so that *residuals* can be computed by taking the difference between the estimated value of the system output variables and their measured value. The residuals are then compared to suitable *thresholds* by *detection* and *isolation logics* in order to provide a *fault decision* regarding the health of the system. Model-based approaches are well-suited to monitoring centralized systems of moderate dimension, but suffer from scalability and robustness issues when distributed and/or large–scale systems are concerned (see, for instance [8]). This research activity is motivated by several applications, especially in complex large-scale systems, such as traffic networks, environmental systems, communication networks, power grid networks, water distribution networks, etc. The study of controlling spatially distributed systems is not a new problem [9] and there have been many enhancements in the design and analysis of distributed control schemes. On the other hand, one area where there has been much less research activity is the problem of designing fault diagnosis schemes specifically for distributed systems. Due to the complexity of the problem, in practice it is difficult to achieve robust fault diagnosis in large-scale distributed systems within a centralized architecture. While considerable effort was aimed at developing distributed fault diagnosis algorithms suited to discrete event systems (see, among many others, [10]), much less attention was devoted to discrete or continuous–time systems (see [11], [12], [13], [14], [15], [16]). In a previous work [17], a quantitative distributed fault detection and isolation scheme for large-scale, nonlinear and uncertain discrete time systems was developed. In the present paper, the results reported in [17] will be extended to a class of input-output non-linear uncertain discrete-time systems, where the system states are only partially measurable. More specifically, the unstructured uncertainty may affect either the discrete–time state or the output equation. Though several papers dealing with centralized fault diagnosis schemes for input-output systems are present in the literature [18], [19], [20], [21], [22], to the best of the authors knowledge this is the first contribution addressing distributed schemes for input-output large-scale nonlinear systems. The main contributions of this paper are the design of a fault detection scheme in a discrete-time framework with modeling uncertainty and partial state measurement and the derivation of rigorous analytical results for the detectability. The paper is organized as follows. Section II formulates the problem under concern. A distributed fault detection architecture is presented in Section III, followed by the detailed development of the detection analysis in Section IV. Then, in Section V, analytical results are presented regarding the fault detectability. Finally, Section VI reports some concluding remarks.

F. Boem is with the Department of Industrial and Information Engineering, DI3, University of Trieste, Italy. (francesca.boem@phd.units.it)
R. M. G. Ferrari is with Danieli Automation S.p.A., Buttrio, Italy. (r.ferrari@dca.it)
T. Parisini is with the Dept. of Electrical and Electronic Engineering at the Imperial College London, UK, and also with the Dept. of Industrial and Information Engineering at the University of Trieste, Italy (t.parisini@paperplaza.net)
M. M. Polycarpou is with the KIOS Research Center for Intelligent Systems and Networks, Dept. of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus. (mpolycar@ucy.ac.cy)

II. Problem formulation

Let us consider a multi-input multi-output uncertain nonlinear system, referred to as *monolithic system*, described by the following discrete-time dynamic equations:

$$\mathscr{S}: \begin{cases} \boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{f}(\boldsymbol{x}(t),\boldsymbol{u}(t)) + \boldsymbol{\eta}_x(\boldsymbol{x}(t),\boldsymbol{u}(t),t) \\ \qquad\qquad + \beta(t-T_0)\boldsymbol{\phi}(\boldsymbol{x}(t),\boldsymbol{u}(t)) \\ \boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{\eta}_y(\boldsymbol{x}(t),\boldsymbol{u}(t),t), \end{cases} \tag{1}$$

where[1] $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{y} \in \mathbb{R}^p$ are the state, the control input and the measured output vectors respectively, the matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and the vector field $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ represent the nominal healthy dynamics, $\boldsymbol{C} \in \mathbb{R}^{p\times n}$ is the nominal output equation, $\boldsymbol{\eta}_x$ and $\boldsymbol{\eta}_y$ are the uncertainties in the state and in the output equations. The term $\beta(t-T_0)\boldsymbol{\phi}(\boldsymbol{x}(t), \boldsymbol{u}(t))$ denotes the changes in the system dynamics due to the occurrence of a fault. More specifically, the vector $\boldsymbol{\phi}(\boldsymbol{x}(t), \boldsymbol{u}(t))$ represents the functional structure of the deviation in the state equation due to the fault and the function $\beta(t-T_0)$ characterizes the time profile of the fault:

$$\beta(t-T_0) \triangleq \begin{cases} 0 & \text{if} \quad t < T_0 \\ 1 - b^{-(t-T_0)} & \text{if} \quad t \geq T_0 \end{cases}. \tag{2}$$

where T_0 is the unknown fault occurrence time and $b > 1$ denotes the unknown fault-evolution rate, modeling either *incipient* faults characterized by a decaying exponential time-profile, or *abrupt* faults characterized by "step-like" time–profiles, as $b \to \infty$. The following assumptions are needed.

Assumption 1: At time $t = 0$ no faults act on the system. Moreover, the state variables $\boldsymbol{x}(t)$ and control variables $\boldsymbol{u}(t)$ remain bounded before and after the occurrence of a fault, i.e., there exist some bounded regions $\mathcal{R} = \mathcal{R}^x \times \mathcal{R}^u \subset \mathbb{R}^n \times \mathbb{R}^n$, such that $(\boldsymbol{x}(t), \boldsymbol{u}(t)) \in \mathcal{R}^x \times \mathcal{R}^u$, $\forall t \geq 0$.

Assumption 2: The modeling and measuring uncertainty terms represented by the vectors $\boldsymbol{\eta}_x$ and $\boldsymbol{\eta}_y$ are unstructured and possibly unknown nonlinear functions of $\boldsymbol{x}$, $\boldsymbol{u}$, and t, but are bounded by some positive, known and bounded functions $\bar{\boldsymbol{\eta}}_x$ and $\bar{\boldsymbol{\eta}}_y$, i.e., $\left|\boldsymbol{\eta}_x^{(h)}(\boldsymbol{x}(t), \boldsymbol{u}(t), t)\right| \leq \bar{\boldsymbol{\eta}}_x^{(h)}(\boldsymbol{x}(t), \boldsymbol{u}(t), t)$ and $\left|\boldsymbol{\eta}_y^{(k)}(\boldsymbol{x}(t), \boldsymbol{u}(t), t)\right| \leq \bar{\boldsymbol{\eta}}_y^{(k)}(\boldsymbol{x}(t), \boldsymbol{u}(t), t)$, for every h-th and k-th component of the vector, with $h = 1, \ldots, n$, $k = 1, \ldots, p$, for all $(\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{R}^x$ and for all t.

As this paper considers only the fault detection problem and not the fault accommodation one, Assumption 1 is just required for well–posedness. Indeed, Assumption 2 is required for the analysis but, in practical situations, if some a-priori knowledge on healthy and faulty behavior is available, these assumptions do not cause a significant loss of generality.

As in [17], here we consider the decomposition of the monolithic system $\mathscr{S}$ into N subsystems $\mathscr{S}_I, I = 1, \ldots, N$. The decomposition of the monolithic system is based on decomposing its structural graph $\mathcal{G} = \{\mathcal{N}_\mathcal{G}, \mathcal{E}_\mathcal{G}\}$, having the node set $\mathcal{N}_\mathcal{G} \triangleq \left\{x^{(i)} : i \in 1, ..., n\right\} \cup \left\{u^{(i)} : i \in 1, ..., m\right\}$ and the arc set of ordered pairs $\mathcal{E}_\mathcal{G} \triangleq \left\{(x^{(i)}, x^{(j)}) : i, j \in 1, ..., n, \text{“}x^{(i)} \text{ affects } x^{(j)}\text{”}\right\} \cup \left\{(u^{(i)}, x^{(j)}) : i \in 1, ..., m, j \in 1, ..., n, \text{“}u^{(i)} \text{ affects } x^{(j)}\text{”}\right\}$, where the superscripts "(i)" and "(j)" denote the i-th and j-th state or control variables of the monolithic system $\mathscr{S}$, respectively. As in [23], [24], the overlapping of certain states $x^{(s)}$ is allowed. In other terms, certain states may belong to more than one subsystem $\mathscr{S}_I$. It is worth noting that in the present paper, for the sake of generality, we consider the decomposition of the states graph, instead of a decomposition made only with respect to the output variables. More specifically, we are concerned with a scenario in which some subsystems may have common state variables, but may differ in their output variables. For example, consider the case of a subsystem where the position of a rigid mechanical body is estimated by measuring its acceleration, while in another subsystem the same position is estimated by measuring its speed: both subsystems share the body position state variable, but they have no common output. After decomposing the monolithic system (1), the I-th subsystem $\mathscr{S}_I$ dynamics can be described by:

$$\begin{aligned} x_I(t+1) =& A_I x_I(t) + f_I(x_I(t), u_I(t)) + g_I(C_I x_I(t), \\ & u_I(t), z_I(t)) + \beta(t-t_0)\phi_I(x_I(t), z_I(t), u_I(t)) \\ y_I(t) =& C_I x_I(t) + \eta_{y,I}(x_I(t), u_I(t), t), \end{aligned} \tag{3}$$

where $x_I \in \mathbb{R}^{n_I}$, $u_I \in \mathbb{R}^{m_I}$ and $y_I \in \mathbb{R}^{p_I}$ are the local state, the local control input, and the local measured output vectors respectively, $z_I \in \mathbb{R}^{q_I}$ is the vector of the interconnection variables, which are the neighbor subsystems nodes having a connection with the elements of I. The term $g_I : \mathbb{R}^{p_I} \times \mathbb{R}^{m_I} \times \mathbb{R}^{q_I} \mapsto \mathbb{R}^{n_I}$ represents the interconnection function where the effects of the local modeling uncertainty term $\eta_{x,I}$ have been incorporated. The following assumption is needed.

Assumption 3: The decomposition of the monolithic system (1) is such that z_I is made of measured variables only.

In this way, it turns out that all the arguments of the interconnection g_I are known: Assumption 3 is needed in order to allow the learning of the interconnection function. This is a key difference between input–output case and the full–state case. Although this assumption is restrictive, there exist some physical systems that satisfy it: an example may be given by an electric distribution network, where we measure power flows in and out different subsystems.

The matrix $A_I \in \mathbb{R}^{n_I \times n_I}$ and the vector field $f_I : \mathbb{R}^{n_I} \times \mathbb{R}^{m_I} \mapsto \mathbb{R}^{n_I}$ represent the local nominal healthy dynamics, $C_I \in \mathbb{R}^{p_I \times n_I}$ is the nominal local output matrix. $\eta_{y,I}$ is the uncertainty function in the local output equation and includes the measurement error; $\phi_I : \mathbb{R}^{p_I} \times \mathbb{R}^{m_I} \times \mathbb{R}^{q_I} \mapsto \mathbb{R}^{n_I}$ is the local fault function. Finally, the following further assumptions are in place.

Assumption 4: (A_I, C_I) is an observable pair.

Assumption 5: The interconnection function g_I is an unstructured and uncertain nonlinear function, whose k-th component is bounded by some known and bounded function, i.e., $\left|g_I^{(k)}(C_I x_I(t), z_I(t), u_I(t))\right| \leq \bar{g}_I^{(k)}(C_I x_I(t), z_I(t), u_I(t))$, for all $I = 1, \ldots, N$ and for all $(\boldsymbol{x}(t), \boldsymbol{u}(t)) \in \mathcal{R}^x \times \mathcal{R}^u$.

III. Distributed FD Architecture

The proposed Distributed Fault Detection (FD) architecture is made of two layers: the physical system $\mathscr{S}$, decomposed into N subsystems $\mathscr{S}_I$, and the detection architecture, that is decomposed as well into N entities $\mathscr{L}_I$, the Local Fault Diagnosers (LFD) (see Fig. 1). Each LFD is devoted to monitor exactly one subsystem: by taking local measurements and by communicating only with neighboring LFDs,

[1]Here and in the following the use of boldface letters will indicate that a given quantity is related to the monolithic system.

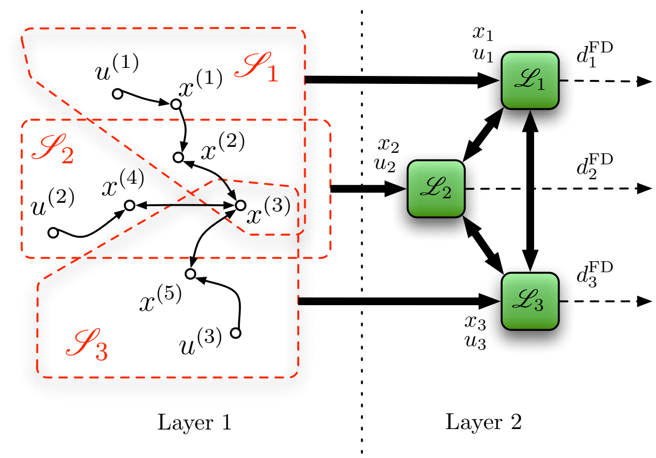

Fig. 1. Scheme of the proposed DFD architecture, applied to a system decomposed into $\mathscr{S}_1$, $\mathscr{S}_2$ and $\mathscr{S}_3$, represented in the first layer. In the second layer each local fault diagnoser $\mathscr{L}_I$ is rendered as a square. Thin arrows symbolize physical interactions in the first layer, thick arrows stand for consistent information flows between layer one and two and between parts of layer two, and then dashed arrows correspond to the LFDs fault decision, which are sporadic.

it produces a fault decision d_I^{FD} regarding its subsystem health (healthy or faulty). For detection purposes, each LFD is equipped with a non-linear adaptive estimator of the local state x_I and of the local output y_I, with $I = 1, ..., N$. The local estimator, called *Fault Detection Approximation Estimator* (FDAE), is based on the nominal model and is used for fault detection. The difference between the estimated output $\hat{y}_I$ and the measurements y_I is the output estimation error $\epsilon_{y,I}(t) \triangleq y_I(t) - \hat{y}_I(t)$, which plays the role of a residual and will be compared, component by component, to a suitable detection threshold $\bar{\epsilon}_{y,I}(t) \in \mathbb{R}^{p_I}$. The following

$$\left|\epsilon_{y,I}^{(k)}(t)\right| \leq \bar{\epsilon}_{y,I}^{(k)}(t), \forall k = 1, ..., p_I \tag{4}$$

is a necessary (but generally not sufficient) condition for the fault-free hypothesis $\mathcal{H}_I$: "The system $\mathscr{S}_I$ is healthy". If the condition is violated at some time instant t, then the hypothesis $\mathcal{H}_I$ is falsified.

Definition 3.1: The local fault detection time is defined as $T_{d,I} = \min\left\{t : \exists k, k \in 1, ..., p_I, \left|\epsilon_{y,I}^{(k)}(t)\right| > \bar{\epsilon}_{y,I}^{(k)}(t)\right\}$.

The local FDAE estimation, in the case of non-shared state variables, can be computed as:

$$\begin{cases} \hat{x}_I(t+1) = A_I\hat{x}_I(t) + f_I(\hat{x}_I(t), u_I(t)) + \\ \qquad \hat{g}_I(y_I(t), u_I(t), v_I(t), \hat{\vartheta}_I) + L_I(y_I(t) - \hat{y}_I(t)) \;, \\ \hat{y}_I(t) = C_I\hat{x}_I(t) \end{cases} \tag{5}$$

where $\hat{g}_I$ is the output of an adaptive approximator designed to learn the unknown interconnection function g_I and $\hat{\vartheta}_I \in \hat{\Theta}_I$ denotes its adjustable parameters vector. Due to the uncertain output measurements, it follows that, instead of receiving the actual interconnection vector z_I, each LFD receives from its neighbors the vector $v_I(t) = z_I(t) + \varsigma_I(t)$, where $\varsigma_I(t)$ is made with the components of $\eta_{y,J}$ that affect the relevant components of the neighboring subsystems measurements y_J. In the case of variables $x^{(s)}$ shared among more than one LFDs, we take advantage of the redundancy obtained by means of the overlap. We propose a deterministic consensus protocol defined on a generic communication graph $\mathcal{G}_s \triangleq (\mathcal{O}_s, \mathcal{E}_s)$, whose nodes are the LFDs in the overlap set $\mathcal{O}_s$ of $x^{(s)}$ (see [17]):

$$\begin{aligned} \hat{x}_I^{(s_I)}(t+1) = &\sum_{J \in \mathcal{O}_s} W_s^{(I,J)} \Big[A_J^{(s_J)}\hat{x}_J(t) \\ &+ f_J^{(s_J)}(\hat{x}_J(t), u_J(t)) + \hat{g}_J^{(s_J)}(y_J(t), u_J(t), v_J(t), \hat{\vartheta}_J) \\ &+ L_J^{(s_J)}(y_J(t) - \hat{y}_J(t))\Big] \end{aligned} \tag{6}$$

where the terms $W_s^{(I,J)}$ are the components of a doubly stochastic weighted adjacency matrix, as for instance the Metropolis matrix [25], [26]:

$$W_s^{(I,J)} \triangleq \begin{cases} 0 & (I,J) \notin \mathcal{E}_s \\ \frac{1}{1+\max\left\{d_s^{(I)}, d_s^{(J)}\right\}} & (I,J) \in \mathcal{E}_s, I \neq J \\ 1 - \sum_{K \neq I} W_s^{(I,K)} & I = J \end{cases} \tag{7}$$

with $d_s^{(I)}$ being the degree of the I-th node in the communication graph $\mathcal{G}_s$. It is important to note that, in order to implement (6), the I-th LFD does not need the information about the expressions of $A_J^{(s_J)}$, $f_J^{(s_J)}$, $\hat{g}_J^{(s_J)}$ and of $L_J^{(s_J)}$; instead, it is sufficient that each LFD computes locally the term $A_J^{(s_J)}\hat{x}_J(t) + f_J^{(s_J)}(t) + \hat{g}_J^{(s_J)}(t) + L_J^{(s_J)}(y_J(t) - \hat{y}_J(t))$ and communicates it to other LFDs according to the communication graph $\mathcal{G}_s$.

IV. Healthy Modes of Behavior: Analysis

We now analyze the dynamics of the FDAE estimation errors before the occurrence of a fault. In the non-shared case, the i-th state estimation error component is:

$$\begin{aligned} \epsilon_{x,I}^{(i)}(t+1) = &A_I^{(i)}x_I(t) + f_I^{(i)}(x_I(t), u_I(t)) \\ &+ g_I^{(i)}(C_Ix_I(t), u_I(t), z_I(t)) - A_I^{(i)}\hat{x}_I(t) \\ &- f_I^{(i)}(\hat{x}_I(t), u_I(t)) - \hat{g}_I^{(i)}(y_I(t), u_I(t), v_I(t), \hat{\vartheta}_I) \\ &- L_I^{(i)}(y_I(t) - \hat{y}_I(t)) \\ =&A_{0,I}^{(i)}\epsilon_{x,I}(t) + \Delta f_I^{(i)}(t) + \Delta g_I^{(i)}(t) - L_I^{(i)}\eta_{y,I}(t), \end{aligned} \tag{8}$$

where $A_{0,I} \triangleq A_I - L_IC_I$ is a stable matrix (thanks to Assumption 4), $\Delta f_I^{(i)}(t) \triangleq f_I^{(i)}(x_I(t), u_I(t)) - f_I^{(i)}(\hat{x}_I(t), u_I(t))$ and $\Delta g_I^{(i)}(t) \triangleq g_I^{(i)}(C_Ix_I(t), u_I(t), z_I(t)) - \hat{g}_I^{(i)}(y_I(t), u_I(t), v_I(t), \hat{\vartheta}_I)$. We denote with $A^{(i)}$ the i-th row of the matrix A.

In the case of shared variables, the dynamics of the LFD state estimation error component can be written as:

$$\begin{aligned} \epsilon_{x,I}^{(s_I)}(t+1) = &x_I^{(s_I)}(t+1) - \hat{x}_I^{(s_I)}(t+1) = A_I^{(s_I)}x_I(t) \\ &+ f_I^{(s_I)}(x_I(t), u_I(t)) + g_I^{(s_I)}(C_Ix_I(t), u_I(t), z_I(t)) \\ &- \sum_{J \in \mathcal{O}_s} W_s^{(I,J)} \Big[A_J^{(s_J)}\hat{x}_J(t) + f_J^{(s_J)}(\hat{x}_J(t), u_J(t)) \\ &+ \hat{g}_J^{(s_J)}(y_J(t), u_J(t), v_J(t), \hat{\vartheta}_J) + L_J^{(s_J)}(y_J(t) - \hat{y}_J(t))\Big] \;. \end{aligned}$$

Because of the way the model decomposition was obtained [17], the following holds for shared variables, $\forall J \in \mathcal{O}_s$:

$$\begin{aligned} &A_I^{(s_I)} x_I + f_I^{(s_I)}(x_I, u_I) + g_I^{(s_I)}(C_I x_I, u_I, z_I) \\ &= A_J^{(s_J)} x_J + f_J^{(s_J)}(x_J, u_J) + g_J^{(s_J)}(C_J x_J, u_J, z_J) \\ &= A^{(s)} x + f^{(s)}(x, u) + \eta_x^{(s)}(x, u, t); \end{aligned}$$

Moreover, by assumption it holds $\sum_{J\in\mathcal{O}_s} W_s^{(I,J)} = 1$. Then, the state estimation error component can be rewritten as:

$$\begin{aligned} \epsilon_{x,I}^{(s_I)}(t+1) = \sum_{J\in\mathcal{O}_s} W_s^{(I,J)} \Big[& A_{0,J}^{(s_J)} \epsilon_{x,J}(t) \\ & + f_J^{(s_J)}(x_J(t), u_J(t)) - f_J^{(s_J)}(\hat{x}_J(t), u_J(t)) \\ & + g_J^{(s_J)}(C_J x_J(t), u_J(t), z_J(t)) \\ & - \hat{g}_J^{(s_J)}(y_J(t), u_J(t), v_J(t), \hat{\vartheta}_J) - L_J^{(s_J)} \eta_{y,J}(t) \Big] . \end{aligned} \quad (9)$$

Summing up, the dynamics of a component of the state estimation error can be written as in (8) in the case of non-shared state variables, while, for shared state variables:

$$\begin{aligned} \epsilon_{x,I}^{(s_I)}(t+1) = \sum_{J\in\mathcal{O}_s} W_s^{(I,J)} \Big[& A_{0,J}^{(s_J)} \epsilon_{x,J}(t) + \Delta f_J^{(s_J)}(t) \\ & + \Delta g_J^{(s_J)}(t) - L_J^{(s_J)} \eta_{y,J}(t) \Big] . \end{aligned} \quad (10)$$

We now introduce a general formulation of the state error equation for analysis purpose. To this end we define the extended state estimation error vector $\epsilon_{x,E} \in \mathbb{R}^{n_E \times 1}$, with $n_E = \sum_{J=1}^{N} n_J$, that is a column vector collecting the state estimation error vectors of the N sub-systems: $\epsilon_{x,E}(t) \triangleq \text{col}\,(\epsilon_{x,J}(t) \,:\, J = 1, ..., N)$. The dynamics of $\epsilon_{x,E}(t)$ are:

$$\begin{aligned} \epsilon_{x,E}(t+1) = W \,[& A_{0,E} \epsilon_{x,E}(t) + \Delta f_E(t) + \Delta g_E(t) \\ & - L_E \eta_{yE}(t)] \end{aligned} \quad (11)$$

where W is a $N \times N$ block matrix

$$W \triangleq \begin{bmatrix} W_{1,1} & \dots & W_{1,N} \\ \dots & \dots & \dots \\ W_{N,1} & \dots & W_{N,N} \end{bmatrix} ,$$

such that each block $W_{I,J}$, with $J = 1, ..., N$ and $I = 1, ..., N$ collects the consensus weights of the subsystem I with regard to the subsystem J. The diagonal blocks $W_{I,I}$ are square diagonal matrices in $\mathbb{R}^{n_I \times n_I}$, whose s_I–th diagonal element, with $s_I = 1, ..., n_I$, is equal to the weight $W_s^{(I,I)}$ defined in Eq. (7) if $x_I^{(s_I)}$ is a shared variable, and is equal to 1 otherwise. The matrices $W_{I,J} \in \mathbb{R}^{n_I \times n_J}$, with $J \neq I$, have non-null elements only in positions (s_I, s_J) corresponding to shared variables x_s, and here they take the value of the consensus weight $W_s^{(I,J)}$. This results in W being a symmetrical, sparse and doubly–stochastic $n_E \times n_E$ matrix. $A_{0,E}$ is a $N \times N$ diagonal block matrix:

$$A_{0,E} \triangleq \begin{bmatrix} A_{0,1} & 0 & 0 & 0 \\ 0 & A_{0,2} & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & A_{0,N} \end{bmatrix} ,$$

where the generic block is $A_{0,J} = A_J - L_J C_J \in \mathbb{R}^{n_J \times n_J}$, for $J = 1, .., N$, resulting in $A_{0,E}$ being a sparse $n_E \times n_E$ matrix. $\Delta f_E(t)$ is a $n_E \times 1$ matrix, collecting the values $\Delta f_J^{(s_J)}(t)$, for each $s_J = 1, ..., n_J$ and for every $J = 1, ..., N$. $\Delta g_E(t)$ is defined in an analogous way as $\Delta f_E(t)$. Furthermore, $L_E \triangleq \text{blkdiag}(L_J \,:\, J = 1, .., N)$ is a $N \times N$ diagonal block matrix with dimension $n_E \times p_E$, where $p_E \triangleq \sum_{J=1}^{N} p_J$, while $\eta_{y,E}(t)$ is a $p_E \times 1$ column vector collecting the uncertainty terms of the N subsystems: $\eta_{y,E} \triangleq \text{col}\,(\eta_{y,I} \,:\, J = 1, ..., N)$. The state estimation error solution can be written as:

$$\begin{aligned} \epsilon_{x,E}(t) = \sum_{h=0}^{t-1} (W A_{0,E})^{t-1-h} \,[& W \Delta f_E(h) + W \Delta g_E(h) \\ & - W L_E \eta_{y,E}(h)] + (W A_{0,E})^t \epsilon_{x,E}(0) \end{aligned} \quad (12)$$

The extended output estimation error is then defined as:

$$\epsilon_{y,E}(t) \triangleq C_E \epsilon_{x,E}(t) + \eta_{y,E}(t) \quad (13)$$

where $C_E \triangleq \text{blkdiag}(C_J \,:\, J = 1, .., N)$ is a $N \times N$ diagonal block matrix, with dimension $p_E \times n_E$. From (11), (13) and the definition of C_E, the following learning law for the adjustable parameter vector $\hat{\vartheta}_I$ of the adaptive approximator $\hat{g}_I$, $I \in 1, \dots, N$ can be derived:

$$\hat{\vartheta}_I(t+1) = P_{\hat{\Theta}_I} \left[\hat{\vartheta}_I(t) + \gamma_I(t) H_I^\top(t) W_{I,I}^\top C_I^\top \epsilon_{y,I}(t+1) \right] \quad (14)$$

$$H_I^\top(t) = \partial \hat{g}_I(t) / \partial \hat{\vartheta}_I$$

$$\gamma_I(t) = \frac{\mu_I}{\varepsilon_I + \left\| H_I^\top(t) W_{I,I}^\top C_I^\top \right\|_F^2} ,$$

where $P_{\hat{\Theta}_I}$ is a projection operator restricting $\hat{\vartheta}_I$ within $\hat{\Theta}_I$ [27], $\|\cdot\|_F$ denotes the Frobenius norm and $\varepsilon_I > 0$, $0 < \mu_I < 2$ are design constants that guarantee the stability of the learning law [28], [29], [30], [27], [31], [32]. The component-wise output estimation error can be written as: $\epsilon_{y,E}^{(k)}(t) = C_E^{(k)} \epsilon_{x,E}(t) + \eta_{y,E}^{(k)}(t)$, for all $k = 1, ..., p_E$. Since each row of C_E, because of the way the matrix was defined, presents non-null values only in correspondence to the state components of a single subsystem, it is possible to write: $\epsilon_{y,I}^{(k)}(t) = C_I^{(k)} \epsilon_{x,I}(t) + \eta_{y,I}^{(k)}(t)$, for all $k = 1, ..., p_I$, and for each subsystem $\mathscr{S}_I$, $I \in 1, \dots, N$. In the general form, the component-wise output estimation error can be bounded by the following threshold, that can be computed in a distributed way:

$$\begin{aligned} & \left| \epsilon_{y,E}^{(k)}(t) \right| \leq \left| C_E^{(k)} \epsilon_{x,E}(t) \right| + \left| \eta_{y,E}^{(k)}(t) \right| \\ & \leq \left| C_E^{(k)} \Big\{ \sum_{h=0}^{t-1} (W A_{0,E})^{t-1-h} \Big[W \Delta f_E(h) + W \Delta g_E(h) \right. \\ & \qquad \left. - W L_E \eta_{y,E}(h) \Big] + (W A_{0,E})^t \epsilon_{x,E}(0) \Big\} \right| + \bar{\eta}_{y,E}^{(k)}(t) \end{aligned}$$

$$\leq \left|C_E^{(k)}\right| \left\{ \sum_{h=0}^{t-1} \left\|(WA_{0,E})^{t-1-h}\right\| \left| W\left[\Delta f_E(h) + \Delta g_E(h) - L_E \eta_{y,E}(h)\right]\right| + \left\|(WA_{0,E})^t\right\| \left|\epsilon_{x,E}(0)\right| \right\} \Big| + \bar{\eta}_{y,E}^{(k)}(t)$$

$$\leq \left|C_E^{(k)}\right| \left\{ \sum_{h=0}^{t-1} \alpha\delta^{t-1-h} W\left[\bar{\Delta} f_E(h) + \bar{\Delta} g_E(h) + |L_E|\, \bar{\eta}_{y,E}(h)\right] + \alpha\delta^t \bar{\epsilon}_{x,E}(0) \right\} + \bar{\eta}_{y,E}^{(k)}(t) \triangleq \bar{\epsilon}_{y,E}^{(k)}(t) \quad (15)$$

where we denote with $|A|$ the element by element absolute value of the matrix A; α and δ, analogously to [33], are two constants such that $\|(WA_{0,E})^t\| \leq \alpha\delta^t \leq \|WA_{0,E}\|^t$, $\alpha > 0$, $0 < \delta \leq 1$. Furthermore,

$$\bar{\Delta} f_E^{(s)}(t) = \max_{x^{(s)} \in R^{x^{(s)}}} \left\{ \left|\Delta f_E^{(s)}(t)\right| \right\},$$

$$\bar{\epsilon}_{x,E}^{(s)}(0) = \max_{x^{(s)} \in R^{x^{(s)}}} \left\{ \left|x^{(s)} - \hat{x}^{(s)}(0)\right| \right\},$$

for every $s = 1, ..., n_E$. As far as the $\bar{\Delta} g_E(t)$ term is concerned, some considerations are necessary. Although the aim of the adaptive approximator $\hat{g}_I$ is to learn the uncertain function g_I, generally it cannot be expected to match the actual term g_I even if the weights of the adaptive approximator could be optimally selected. This may be formalized by introducing an *optimal weight vector* [34] $\hat{\vartheta}_I^*$:

$$\hat{\vartheta}_I^* \triangleq \arg \min_{\hat{\vartheta}_I \in \Theta_I} \sup_{x_I, z_I, u_I} \| g_I(C_I x_I, z_I, u_I) - \hat{g}_I(C_I x_I, z_I, u_I, \hat{\vartheta}_I \|, \quad (16)$$

with x_I, z_I, u_I taking values in their respective domains. This leads to the definition of the *Minimum Functional Approximation Error* (MFAE) ν_I, which describes the least possible approximation error that can be achieved at time t if $\hat{\vartheta}_I$ were optimally chosen: $\nu_I(t) \triangleq g_I(C_I x_I, z_I, u_I) - \hat{g}_I(C_I x_I, z_I, u_I, \hat{\vartheta}_I^*)$. By defining the parameter estimation error $\tilde{\vartheta}_I \triangleq \hat{\vartheta}_I^* - \hat{\vartheta}_I$ and the function $\Delta\hat{g}_I \triangleq \hat{g}_I(C_I x_I, z_I, u_I, \hat{\vartheta}_I) - \hat{g}_I(y_I, v_I, u_I, \hat{\vartheta}_I)$, it turns out that $\Delta g_I(t) = H_I \tilde{\vartheta}_I + \nu_I(t) + \Delta\hat{g}_I(t)$.

We can upper bound this term with: $\bar{\Delta} g_I(t) \triangleq \|H_I\| \,\kappa_I(\hat{\vartheta}_I) + \bar{\nu}_I(t) + \max_{\eta_{yI}} \max_{\varsigma_I} |\Delta\hat{g}_I(t)|$, with the function κ_I being such that $\kappa_I(\hat{\vartheta}_I) \geq \left\|\tilde{\vartheta}_I\right\|$. The extended upper bound $\bar{\Delta} g_E(t)$ simply collects the upper bounds of the N subsystems. The threshold in Eq. (15) guarantees that no false-positive alarms will be issued until T_0 because of the uncertainties. This, of course and in rough terms, comes at the cost of the impossibility of detecting faults "hidden by the uncertainties in the system dynamics". This is formalized in the following section in which a distributed detectability sufficient condition will be devised.

V. Faulty Modes of Behavior: Analysis and Detectability Sufficient Conditions

Let us assume that at time $t = T_0$ a fault $\boldsymbol{\phi}$ occurs in the monolithic system. ϕ_E denotes the extended fault function vector collecting the N subsystems fault functions. After the occurrence of the fault, for $t > T_0$, the state estimation error dynamics becomes

$$\epsilon_{x,E}(t+1) = W\left[A_{0,E}\epsilon_{x,E}(t) + \Delta f_E(t) + \Delta g_E(t) - L_E \eta_{y,E}(t)\right] + (1 - b^{-(t-T_0)})\phi_E(t) \quad (17)$$

and the output estimation error equation for the k-th component is:

$$\epsilon_{y,E}^{(k)}(t) = C_E^{(k)} \epsilon_{x,E}(t) + \eta_{y,E}^{(k)}(t) = C_E^{(k)} \left\{ \sum_{h=0}^{t-1} (WA_{0,E})^{t-1-h} \Big[W\Delta f_E(h) + W\Delta g_E(h) - WL_E \eta_{y,E}(h) + (1 - b^{-(h-T_0)})\phi_E(h) \Big] + (WA_{0,E})^t \epsilon_{x,E}(0) \right\} + \eta_{y,E}^{(k)}(t) \quad (18)$$

Now, we are able to state and prove a sufficient condition for distributed fault detectability.

Theorem 5.1 (Fault Detectability): If there exists a time instant $t_1 > T_0$ such that the fault ϕ_E satisfies the inequality

$$\left| \sum_{h=T_0}^{t_1-1} C_E^{(k)} (WA_{0,E})^{t_1-1-h} (1 - b^{-(h-T_0)})\phi_E(h) \right| > 2\bar{\epsilon}_{y,E}^{(k)}(t_1) \quad (19)$$

for at least one component $k \in \{1, ..., p_E\}$, then the fault will be detected at time t_1, that is $\left|\epsilon_{y,E}^{(k)}(t_1)\right| > \bar{\epsilon}_{y,E}^{(k)}(t_1)$.

Proof: At time instant $t_1 > T_0$, the output estimation error can be written as:

$$\epsilon_{y,E}^{(k)}(t_1) = \sum_{h=0}^{t_1-1} C_E^{(k)} (WA_{0,E})^{t_1-1-h} \left[W\Delta f_E(h) + W\Delta g_E(h) - WL_E \eta_{y,E}(h)\right] + C_E^{(k)} (WA_{0,E})^{t_1} \epsilon_{x,E}(0) + \eta_{y,E}^{(k)}(t_1) + \sum_{h=T_0}^{t_1-1} C_E^{(k)} (WA_{0,E})^{t_1-1-h} (1 - b^{-(h-T_0)})\phi_E(h)$$

Using the triangle inequality we obtain:

$$\left|\epsilon_{y,E}^{(k)}(t_1)\right| \geq - \left| \sum_{h=0}^{t_1-1} C_E^{(k)} (WA_{0,E})^{t_1-1-h} \left[W\Delta f_E(h) + W\Delta g_E(h) - WL_E \eta_{y,E}(h)\right] + C_E^{(k)} (WA_{0,E})^{t_1} \epsilon_{x,E}(0) + \eta_{y,E}^{(k)}(t_1) \right| + \left| \sum_{h=T_0}^{t_1-1} C_E^{(k)} (WA_{0,E})^{t_1-1-h} (1 - b^{-(h-T_0)})\phi_E(h) \right|$$

By recalling how the threshold was defined (Eq. 15), it is easy to see that the following inequality is implied:

$$\left|\epsilon_{y,E}^{(k)}(t_1)\right| \geq -\bar{\epsilon}_{y,E}^{(k)}(t_1) + \left| \sum_{h=T_0}^{t_1-1} C_E^{(k)} (WA_{0,E})^{t_1-1-h} (1 - b^{-(h-T_0)})\phi_E(h) \right| .$$

In this way the fault detection condition $\left|\epsilon_{y,E}^{(k)}(t_1)\right| > \bar{\epsilon}_{y,E}^{(k)}(t_1)$ is implied by the theorem hypothesis. ■

Theorem 5.1 represents a sufficient condition for the off-line characterization, in a non-closed form, of a class of faults that can be detected by the proposed FD methodology.

VI. Concluding Remarks

In this paper, a distributed fault detection scheme for a class of large-scale input-output non-linear discrete-time uncertain systems was proposed, that relies on nonlinear adaptive estimators based on a nominal model of the healthy system dynamics. Each subsystem is monitored by a local fault detection unit, which is able to detect the presence of faults affecting the corresponding subsystem based on its own measurements and on communication with neighboring subsystems. An adaptive approximation scheme is developed in order to learn the functional uncertainty in the interconnection between neighboring subsystems, before any fault is detected. Both abrupt and incipient kinds of faults were addressed and theoretical results characterizing the ability of the FD scheme to detect a fault were derived. Future research efforts will be devoted, first of all, to show the effectiveness of the proposed technique by extensive simulation trials on models of large-scale systems of practical interest. Moreover, a thorough analysis of the conservativeness of the detectability conditions will be carried out. Besides, ongoing research aims at weakening some of the assumptions made in the paper like, for instance, Assumption 3, restricting the decomposition to be such that the interconnection variables are measurable.

References

[1] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer-Verlag, 2006.

[2] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault Tolerant Control*. Berlin: Springer, 2003.

[3] R. Patton, P. Frank, and D. Clark, *Fault Diagnosis in Dynamic Systems: Theory and Application*. Upper Saddle River, NJ, USA: Prentice Hall, 1989.

[4] R. Isermann, "Process fault detection based on modeling and estimation methods. a survey." *Automatica—J. IFAC*, vol. 20, no. 4, pp. 387–404, 1984.

[5] J. Gertler, "Survey of model-based failure detection and isolation in complex plants," *IEEE Control Syst. Mag.*, vol. 8, no. 6, pp. 3–11, 1988.

[6] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge–based redundancy – a survey and some new results," *Automatica—J. IFAC*, vol. 26, no. 3, pp. 459–474, 1990.

[7] 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 & Chem. Eng.*, vol. 27, pp. 293–311, 2003.

[8] C. Langbort, R. Chandra, and R. D'Andrea, "Distributed control design for systems interconnected over an arbitrary graph," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1502–1519, 2004.

[9] N. Sandell, P. Varaiya, M. Athans, and M. Safonov, "Survey of decentralized control methods for large scale systems," *IEEE Trans. Autom. Control*, vol. 23, no. 2, pp. 108–128, 1978.

[10] P. Baroni, G. Lamperti, P. Pogliano, and M. Zanella, "Diagnosis of large active systems," *Artificial Intelligence*, vol. 110, no. 1, pp. 135–189, 1999.

[11] M. Staroswiecki, G. Hoblos, and A. Aitouche, "Sensor network design for fault tolerant estimation," *Int. J. Adapt. Control Signal Process.*, vol. 18, no. 1, pp. 55–72, 2004.

[12] R. J. Patton, C. Kambhampati, A. Casavola, P. Zhang, S. Ding, and D. Sauter, "A generic strategy for fault-tolerance in control systems distributed over a network," *European Journal of Control*, vol. 13, no. 2–3, pp. 280–296, 2007.

[13] S. Klinkhieo and R. J. Patton, "A two-level approach to fault-tolerant control of distributed systems based on the sliding mode," in *Preprints of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. Barcelona, Spain*, 2009, pp. 1043–1048.

[14] W. Li, W. Gui, Y. Xie, and S. Ding, "Decentralized fault detection system design for large-scale interconnected systems," in *Preprints of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. Barcelona, Spain*, 2009, pp. 816–821.

[15] T. P. X. Zhang, M. M. Polycarpou, "Decentralized fault detection in a class of large-scale nonlinear uncertain system," in *Proc. Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, China*, 2009, pp. 6988–6993.

[16] N. Meskin, K. Khorasani, and C. A. Rabbath, "Fault consensus in a network of unmanned vehicles," in *Preprints of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. Barcelona, Spain*, 2009, pp. 1001–1006.

[17] R. M. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault diagnosis of large-scale discrete-time nonlinear systems: New results on the isolation problem," *Proc. 49th IEEE Conf. on Decision and Control (CDC '10)*, pp. 1619–1626, 2010.

[18] A. Vemuri and M. Polycarpou, "Robust nonlinear fault diagnosis in input-output systems," *International Journal of Control*, vol. 68, no. 2, pp. 343–360, 1997.

[19] Q. Zhang, M. Basseville, and A. Benveniste, "Fault detection and isolation in nonlinear dynamic systems: A combined input–output and local approach," *Automatica*, vol. 34, no. 11, pp. 1359–1373, 1998.

[20] X. Zhang, M. M. Polycarpou, and T. Parisini, "Robust fault isolation for a class of non–linear input–output systems," *Internat. J. Control*, vol. 74, no. 13, pp. 1295–1310, 2001.

[21] K. Subbarao and A. Vemuri, "Fault isolation using extrinsic curvature for multi-input-multi-output systems with nonlinear fault models," *Proc. American Control Conf., 2007 (ACC '07)*, 2007.

[22] Z. Zhang and I. Jaimoukha, "Fault detection and isolation for linear discrete-time systems using input/output measurement analysis," in *Proc. 48th IEEE Conf. on Decision and Control, held jointly with the 28th Chinese Control Conference.*, 2009, pp. 4908–4913.

[23] R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault diagnosis with overlapping decompositions: an adaptive approximation approach," *IEEE Trans. Autom. Control*, vol. 54, no. 4, pp. 794–799, 2009.

[24] R. M. Ferrari, "Fault diagnosis of distributed large-scale discrete-time nonlinear systems," *University of Trieste Technical report, http://control.units.it/ferrari*, 2010.

[25] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," *Proc. 4th Int. Symp. on Information Process. in Sensor Netw. (IPSN '05)*, pp. 63–70, 2005.

[26] L. Xiao, S. Boyd, and S. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33–46, 2007.

[27] M. M. Polycarpou, "On–line approximators for nonlinear system identification: a unified approach," in *Control and Dynamic Systems: Neural Network Systems Techniques and Applications*, X. Lenodes, Ed. New York: Academic, 1998, vol. 7, pp. 191–230.

[28] G. Goodwin and Sin, *Adaptive filtering prediction and control*. Englewood Cliffs, NJ: Prentice Hall, 1984.

[29] C. R. Johnson, *Lectures on adaptive parameter estimation*. Upper Saddle River, NJ, USA: Prentice Hall, 1988.

[30] J. Astrom and B. Wittenmark, *Adaptive Control*, 2nd ed. Prentice Hall, 1994.

[31] F. Lewis, Yeşildirek, and S. Jagannathan, *Neural Network Control of Robot Manipulators and Non-Linear Systems*. CRC Press, 1998.

[32] J. Farrell and M. M. Polycarpou, *Adaptive Approximation Based Control: Unifying Neural, Fuzzy, and Traditional Adaptive Approximation Approaches*. Hoboken, NJ: Wiley-Interscience, 2006.

[33] R. M. Ferrari, T. Parisini, and M. M. Polycarpou, "A robust fault detection and isolation scheme for a class of uncertain input-output discrete-time nonlinear systems," *Proc. American Control Conf., 2008 (ACC '08)*, p. 6, 2008.

[34] A. Vemuri and M. M. Polycarpou, "On-line approximation methods for robust fault detection," *Proc. 13th IFAC World Congress*, vol. K, pp. 319–324, 1996.