Actuator Fault Diagnosis With Neural Network-Integral Sliding Mode Based Unknown Input Observers

Nikolas Sacchi * **Gian Paolo Incremona** ** **Antonella Ferrara** *

* *Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, 27100, Pavia, Italy (e-mail: nikolas.sacchi01@universitadipavia.it, antonella.ferrara@unipv.it).*
** *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133, Milan, Italy (e-mail: gianpaolo.incremona@polimi.it)*

Abstract: This paper proposes an integral sliding mode (ISM) based unknown input observer (UIO) which is able to perform fault diagnosis (FD) in condition of lack of knowledge of the plant model. In particular, a two-layer neural network (NN) is employed to estimate online the drift term of the system dynamics needed to compute the so-called integral sliding manifold. The weights of such a NN are updated online using adaptation laws directly derived from theoretical analysis, carried out in this paper. Finally, the proposal has been assessed in simulation relying on a benchmark model of a DC motor.

Keywords: Sliding mode, neural network, fault diagnosis, input observer.

1. INTRODUCTION

Fault diagnosis (FD) has become more and more relevant in the industrial context, where a large number of electromechanical systems are implied in everyday work. The aim of FD is to determine the causes of deviations of the control status from the desired behavior, interpreting information that comes from sensors or from the process model (Isermann, 2005).

In order to compensate the effects of faults and guarantee the continuity of the work processes, *detection*, *isolation* and *identification* must be performed. The detection procedure allows to understand if and when a fault is affecting the system, without prior knowledge on the faulty component. Isolation and identification take care of finding the faulty element and reconstructing the fault signal, respectively. Furthermore, the FD methodologies proposed in literature during the years can be distinguished in two main categories, i.e., *passive* methods, in which the input is fed both into the actual process and its nominal model in order to check differences in the system behaviour (Isermann, 2011), and *active* methods, which rely on the injection of auxiliary signals to improve detectability of faults (Scott et al., 2014; Punčochář and Škach, 2018). Other FD techniques rely instead on NNs (Baimukashev et al., 2021), robust or optimal control approaches, like linear quadratic controllers (LQR) (Zhan and Jiang, 1999) and model predictive control (MPC)(Jarrou et al., 2019).

In this context, also Sliding Mode Control (SMC) has been successfully adopted thanks to its robustness against a wide class of uncertain terms, especially the matched ones acting directly on the input channel (Ferrara et al., 2019a). In the domain of FD, sliding mode based approaches have been studied to control, for instance, robotic systems, as in (Bartolini et al., 1997; Capisani et al., 2009). Other works have proposed sliding mode based observers to design fault tolerant control schemes. In particular, robust estimates of the fault signals can be generated using the so-called equivalent control concept, see e.g., (Capisani et al., 2010; Incremona and Ferrara, 2019; Sacchi et al., 2023). However, such an equivalent control holds only during the sliding phase, while the system is still sensitive to the uncertainties in the reaching phase interval.

In order to improve the robustness of SMC, the concept of ISM has been then introduced in (Utkin and Shi, 1996). Indeed, such an improvement allows to enable a sliding mode, along with the robustness property, since the initial time instant, thus removing the reaching phase. The beneficial effects have been assessed for different applications, as in (Ferrara and Incremona, 2015; Incremona et al., 2017; Ferrara et al., 2019b) among many others. However, one of the main features of the ISM approach is the required knowledge of the nominal system model. If on the one hand it is common to consider the terms multiplying the input known, the same assumption cannot be made on the drift terms. In fact, in many practical implementations, only conservative bounds can be retrieved.

In the last years, also motivated by the growth of the available computational power, the learning paradigm has gained popularity. In particular, powerful function approximators like NNs (Hornik et al., 1989) have been applied

⋆ This is the final version of the accepted paper submitted for inclusion in the Proceedings of the 22nd IFAC World Congress 2023, Yokohama, Japan, July 2023, doi:10.1016/j.ifacol.2023.10.1659. This work has been partially supported by the Italian Ministry for Research in the framework of the 2017 Program for Research Projects of National Interest (PRIN), Grant no. 2017YKXYXJ.

in the design of a variety of control schemes, thus creating several learning-based data-driven control approaches with stability guarantees, like the ones presented in (Bonassi et al., 2022). Moreover, in (Lewis et al., 1999), weight adaptation laws for NNs based on Lyapunov stability analysis have been introduced. Such a technique has been used to estimate part of the system model or to directly approximate the optimal control law, as proposed in (Esfandiari et al., 2022; Cheng et al., 2021). Moreover, several methodologies which combine NNs with SMC have been proposed, see, e.g., (Tai and Ahn, 2010; Fei and Lu, 2017). More recently, in (Sacchi et al., 2022), a two-layer NN which estimates the plant model, while using the aforementioned weight adaptation laws, has been proposed relying on ISM control, giving rise to a novel NN-ISM control method.

Motivated by (Sacchi et al., 2022), and having in mind a FD application, in this paper we propose a novel NN-ISM Unknown Input Observer (UIO) to detect faults affecting the input of a general nonlinear system. In particular, differently from the literature, the main advantage of this work is that we use a two-layer NN to estimate online the a priori unknown drift term of the system dynamics needed to compute the so-called integral sliding manifold. The weights of such a NN are randomly initialized and then updated according to adaptation laws derived by Lyapunov stability analysis. The validity of the proposed NN-ISM UIO is assessed in simulation relying on a benchmark model of a DC motor.

The paper is structured as follows. In Section 2 the considered nonlinear system, and the concept of ISM observers are presented. In Section 3 the universal approximation property of NNs is introduced. In Section 4, the proposed NN-ISM UIO and the main theoretical results are provided. Section 5 describes the numerical simulation and presents the results, while some conclusions are finally drawn in Section 6.

Notation: Let $x \in \mathbb{R}^m$ be a column vector, then $x^\top \in \mathbb{R}^{1\times m}$ represents its transpose. Given a real matrix $A \in \mathbb{R}^{m\times m}$, then $\overline{\lambda}(A)$ and $\underline{\lambda}(A)$ are the greatest and smallest singular values of A, while $\mathrm{tr}(A)$ is its trace. Given two real matrices $A, B \in \mathbb{R}^{m\times m}$, then $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$, while given $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{m\times n}$, then $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. Given two real column vectors $a, b \in \mathbb{R}^m$, the trace of the outer product is equivalent to the inner product, i.e., $\mathrm{tr}(ba^\top) = a^\top b$. Given a real matrix $A \in \mathbb{R}^{n\times m}$, then $A^\dagger \in \mathbb{R}^{m\times n}$ is its pseudo-inverse, defined as $A^\dagger = (A^\top A)^{-1}A^\top$. Let $\mathbb{1}_{p\times 1} \in \mathbb{R}^p$ be a column vector of p ones.

2. PRELIMINARIES AND PROBLEM STATEMENT

In this section, first the class of considered systems and faults is presented, and then some preliminaries on the design of a ISM based input observer are recalled.

2.1 The considered faulty system

Consider a nonlinear system affected by an actuator fault, expressed in state-space form as

$$\dot{x} = f(x(t)) + B(x(t))\left[u(t) + \Delta u(t)\right], \qquad (1)$$

where $x(t) \in \mathbb{R}^n$ is the system state vector, $f(x(t)) : \mathbb{R}^n \to \mathbb{R}^n$ is the drift dynamics, $B(x(t)) : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the control effectiveness matrix, $u = \kappa(x) \in \mathbb{R}^m$ is any suitable (in whatever appropriate sense in terms of desired response) stabilizing control law for the system without faults, while $\Delta u(t) : \mathbb{R} \to \mathbb{R}^m$ is the actuator fault. Note that, the fault is modelled as an additive disturbance in the input channel, see e.g., (Capisani et al., 2010; Incremona and Ferrara, 2019; Sacchi et al., 2023). Moreover, the following classical assumptions on the fault, the drift term, and the effectiveness matrix need to be introduced.

$\boldsymbol{\mathcal{A}_1}$: There exist known constants $\overline{\delta}, \overline{f}, \overline{b} \in \mathbb{R}_{>0}$, so that the actuator fault $\Delta u(t)$, the drift dynamics $f(x)$ and the effectiveness matrix $B(x)$ are bounded as

$$\sup_{t\in\mathbb{R}_{>0}} \|\Delta u(t)\| \le \overline{\delta}, \quad \sup_{x\in\mathbb{R}^n} \|f(x)\| \le \overline{f}, \quad \sup_{x\in\mathbb{R}^n} \|B(x)\| \le \overline{b}$$

Note that, to recall the concept of ISM UIO, we consider now only the fault Δu unknown but bounded, while the rest of the dynamics is known and bounded.

2.2 ISM based UIO

The goal of this work is to perform a fault diagnosis in presence of actuator fault. More precisely, to estimate the unknown control fault which affects (1), it is possible to design an UIO of the form

$$\dot{\widehat{x}} = f(\widehat{x}) + B(\widehat{x})\left[u + v\right], \quad \widehat{x}(0) = x(0), \qquad (2)$$

where $v \in \mathbb{R}^m$ is the observer input, which is designed, relying on an ISM strategy, as the sum of two components, i.e., $v = v_0 + v_1$. In particular, $v_0 = \kappa_0(e)$ is selected so as to stabilize the nominal dynamics of the observer error $e := x - \widehat{x} \in \mathbb{R}^n$, that is in the case of $\Delta u = 0$. The second term, v_1, has a discontinuous nature and it is selected as

$$v_1 := \rho\frac{s}{\|s\|}, \qquad (3)$$

where $\rho \in \mathbb{R}_{>0}$ is a constant gain to dominate the mismatches with respect to the nominal model, and $s : \mathbb{R}^n \to \mathbb{R}^m$ is the so-called *integral sliding variable*, defined as

$$s := s_0 + z. \qquad (4)$$

In the considered case, it is convenient to select $s_0 \in \mathbb{R}^m$ dependent on the observer error e, i.e., $s_0 := \Lambda e$, with $\Lambda \in \mathbb{R}^{m\times n}$ being a matrix which satisfies the following assumption.

$\boldsymbol{\mathcal{A}_2}$: The matrix $\Lambda \in \mathbb{R}^{m\times n}$ is chosen so that, $\forall x, \widehat{x} \in \mathbb{R}^n$, the matrices $\Lambda B(x) \in \mathbb{R}^{m\times m}$ and $\Lambda B(\widehat{x}) \in \mathbb{R}^{m\times m}$ are positive definite. Moreover, there exists a known constant $\overline{c} \in \mathbb{R}_{>0}$ so that $|\mathbb{1}_{m\times 1}^\top \Lambda \mathbb{1}_{n\times 1}| \le \overline{c}$.

As for the term $z(x(t)) : \mathbb{R}^n \to \mathbb{R}^m$, which is the so-called *transient function*, it is defined so that

$$\dot{z} := -\frac{\partial s_0}{\partial e}\dot{e}^0, \quad z(0) = -s_0(0), \qquad (5)$$

with e^0 being the nominal evolution of the observer error dynamics obtained subtracting (2) to (1), with $\Delta u = 0$. Then, as stated in the following theorem, selecting a suitable gain ρ for (3), it is possible to enforce a sliding mode since the initial time instant and, exploiting the concept of *equivalent control*, estimate the control fault $\Delta u(t)$.

Theorem 1. Consider system (1) and the UIO (2), with v_1 defined as in (3). If $\boldsymbol{\mathcal{A}}_\mathbf{1}$ and $\boldsymbol{\mathcal{A}}_\mathbf{2}$ hold, and ρ is selected so that

$$\rho > \frac{\overline{\lambda}(\Lambda B(x))}{\underline{\lambda}(\Lambda B(\widehat{x}))}\overline{\delta}, \tag{6}$$

then a sliding mode $s = 0$ is enforced for any $t \geq 0$. Moreover, letting $\widetilde{v}_1$ be the equivalent control, one has that

$$\widetilde{v}_1 = B(\widehat{x})^\dagger B(x)\Delta u. \tag{7}$$

Proof. Selecting the Lyapunov function

$$V(x) = \frac{1}{2}s^\top s, \tag{8}$$

and computing its time derivative $\dot{V}(x) := s^\top \dot{s}$, one has

$$\begin{aligned}\dot{V}(x) &= s^\top \Lambda B(x)\Delta u - s^\top \Lambda B(\widehat{x})\rho\frac{s}{\|s\|}\\ &\leq \overline{\lambda}(\Lambda B(x))\overline{\delta}\|s\| - \underline{\lambda}(\Lambda B(\widehat{x}))\rho\|s\|\\ &\leq \left[\overline{\lambda}(\Lambda B(x))\overline{\delta} - \underline{\lambda}(\Lambda B(\widehat{x}))\rho\right]\|s\| = -\eta\|s\|,\end{aligned}$$

with $\eta := -\left[\overline{\lambda}(\Lambda B(x))\overline{\delta} - \underline{\lambda}(\Lambda B(\widehat{x}))\rho\right]$. Therefore, if assumptions $\boldsymbol{\mathcal{A}}_\mathbf{1}$ and $\boldsymbol{\mathcal{A}}_\mathbf{2}$, along with condition (6), are satisfied, then $\dot{V}(x) \leq -\eta\|s\| < 0$, which implies that a sliding mode $s = 0$ is enforced. Moreover, since $s(0) = 0$, the sliding mode is enforced since the initial time instant. Finally, computing $\dot{s} = 0$ and solving for $\widetilde{v}_1$ allow to obtain the equivalent control (7), which concludes the proof.

As detailed in (Utkin and Shi, 1996), in practice, $\widetilde{v}_1$ is achieved by using a first-order filter (that is, a low-pass (LP) filter) having in input the discontinuous signal v_1.

Note that the previous result is valid in the case of fully known nominal dynamics. In this work we want to address the case of partial information on the system dynamics and we assume that the drift term is unknown. This condition requires the introduction of a new approach to compensate the lack of knowledge on the system.

3. NN-BASED FUNCTION APPROXIMATION

Motivated by the problem stated at the end of the previous section, in the following a NN-based approach is introduced to estimate the unknown drift term, which is instrumental to design the integral sliding variable.

Indeed, as detailed in (Utkin and Shi, 1996) and evident in (5), in order to design an ISM observer, the nominal dynamics must be known. However, in this paper the drift term $f(x)$ is assumed to be unknown. Exploiting the *universal approximation* property introduced in (Hornik et al., 1989), a NN is therefore adopted to approximate the unknown component of the system.

Let $\Omega \subset \mathbb{R}^n$ be a compact set with $x \in \Omega$, and $\mathbb{S}^n(\Omega)$ the space in which the drift term $f(x)$ is continuous. Then, there exists an ideal two-layer NN so that

$$f(x) = W^\top g(\Phi^\top x) + \varepsilon(x), \tag{9}$$

where $W \in \mathbb{R}^{L\times n}$ and $\Phi \in \mathbb{R}^{n\times L}$, with $L \in \mathbb{N}_{>0}$, are ideal weights, $g(\cdot) : \mathbb{R}^L \to \mathbb{R}^L$ is the ideal activation function vector, while $\varepsilon(x) : \mathbb{R}^n \to \mathbb{R}^n$ is the so-called *function reconstruction error*. The following assumption about the bounds of the ideal NN must be introduced.

$\boldsymbol{\mathcal{A}}_\mathbf{3}$: There exist known constants $\overline{W}, \overline{\Phi}, \overline{g}, \overline{\varepsilon} \in \mathbb{R}_{>0}$ such that

$$\sup_{x(t)\in\Omega} \|W\| \leq \overline{W}, \quad \sup_{x(t)\in\Omega} \|\Phi\| \leq \overline{\Phi},$$
$$\sup_{x(t)\in\Omega} \|\varepsilon(x)\| \leq \overline{\varepsilon}, \quad \sup_{x(t)\in\Omega} \|g\| \leq \overline{g}.$$

Since the ideal weights and the ideal activation function vector are not known, an estimation of them is used. The drift term is then rewritten as

$$\widehat{f}(x) := \widehat{W}^\top \widehat{g}(\widehat{\Phi}^\top x), \tag{10}$$

with $\widehat{g}(\cdot) : \mathbb{R}^L \to \mathbb{R}^L$ being a user defined activation function vector, which may differ from the ideal one. For the sake of readability, from now on the terms $g(\Phi^\top x)$ and $\widehat{g}(\widehat{\Phi}^\top x)$ are substituted with their shortcomings g and $\widehat{g}$, respectively. Moreover, the following assumption about $\hat{g}(\cdot)$ is needed.

$\boldsymbol{\mathcal{A}}_\mathbf{4}$: There exists a known constant $\overline{\widehat{g}} \in \mathbb{R}_{>0}$ such that

$$\sup_{x(t)\in\Omega} \|\widehat{g}\| \leq \overline{\widehat{g}}.$$

Finally, the weight estimation errors are computed as

$$\widetilde{W} = W - \widehat{W}, \tag{11a}$$
$$\widetilde{\Phi} = \Phi - \widehat{\Phi}. \tag{11b}$$

We are now in a position to introduce the proposed FD based NN-ISM strategy.

4. THE PROPOSED NN-ISM FAULT DIAGNOSIS SCHEME

The objective of this section is to present the proposed NN-ISM observer, along with the fault detection scheme, illustrated in Fig. 1.

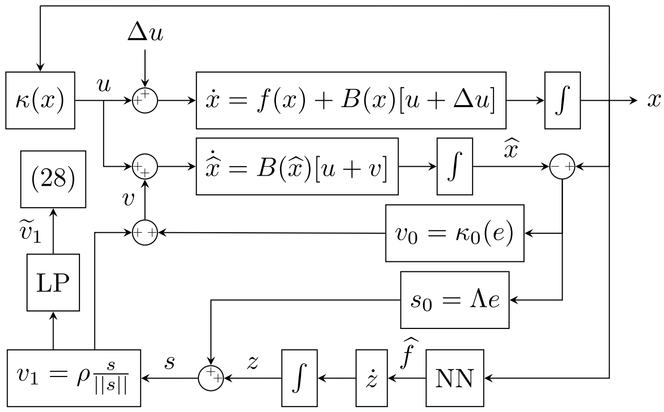


Fig. 1. The proposed scheme with NN-ISM based UIO.

Since the drift term $f(x)$ is considered unknown, the UIO proposed in (2) cannot be implemented. Instead, the proposed UIO is given by

$$\dot{\widehat{x}} = B(\widehat{x})\left[u + v\right], \quad \widehat{x}(0) = x(0), \tag{12}$$

where $u = \kappa(x) \in \mathbb{R}^m$ is the stabilizing control law as defined in §2, while the observer input $v \in \mathbb{R}^m$ is defined as follows

$$v := v_0 + v_1. \tag{13}$$

The first component, i.e., $v_0 = \kappa_0(e)$, is chosen so that it stabilizes the dynamics of the nominal observer error e, which, using (10), can be expressed as

$$\dot{e}^0 = \widehat{W}^\top \widehat{g} + [B(x) - B(\widehat{x})]\, u - B(\widehat{x}) v_0. \tag{14}$$

As for the discontinuous part of the signal, i.e., v_1, it is defined as in (3). In particular, the integral sliding variable is chosen as

$$s := \Lambda(x - \hat{x}) + z, \tag{15}$$

where $\widehat{x}$ is the state of the observer characterized by the dynamics (12), while $\dot{z}$ is defined so that

$$\dot{z} = -\Lambda \left[\widehat{W}^\top \widehat{g} + B(x)u - B(\widehat{x})u - B(\widehat{x}) v_0 \right]. \tag{16}$$

Now, the dynamics of the sliding variable, i.e., $\dot{s} = \Lambda(\dot{x} - \dot{\hat{x}}) - \dot{z}$ can be computed. Specifically, exploiting (1), (9), (12), (16), and the structure of v, it is possible to express $\dot{s}$ as

$$\dot{s} = \Lambda \left[W^\top g - \widehat{W}^\top \widehat{g} + \varepsilon(x) + B(x)\Delta u - B(\widehat{x}) v_1 \right]. \tag{17}$$

The weight adaptation laws, derived from the Lyapunov analysis reported hereafter, are defined as

$$\dot{\widehat{W}} := \Gamma_W \widehat{g} s^\top \Lambda, \tag{18a}$$

$$\dot{\widehat{\Phi}} := \Gamma_\Phi x \left[\dot{\widehat{g}} \widehat{W} (s^\top \Lambda)^\top \right]^\top, \tag{18b}$$

where $\Gamma_W \in \mathbb{R}^{L\times L}$ and $\Gamma_\Phi \in \mathbb{R}^{n\times n}$ are diagonal gain matrices with positive entries, while $\dot{\widehat{g}}$ is the gradient of the activation function vector. If one selects $\widehat{g}$ as a vector of *logistic sigmoid* functions, its gradient can be computed as

$$\dot{\widehat{g}} = \mathrm{diag}\{\hat{g}\}(I_{L\times L} - \mathrm{diag}\{\widehat{g}\}), \tag{19}$$

where $\mathrm{diag}\{\widehat{g}\} \in \mathbb{R}^{L\times L}$ is a diagonal matrix built with the elements of $\widehat{g}$, while $I_{L\times L} \in \mathbb{R}^{L\times L}$ is the identity matrix.

In the following, the main theoretical results relevant to the proposed FD scheme are presented. In particular, the following theorem provides conditions on the gain of the discontinuous part of the observer which, if satisfied, allows to provide an estimate of the actuator fault Δu affecting system (1), with a bounded estimation error.

Theorem 2. Consider the nonlinear system (1), the unknown input observer (12), the integral sliding variable (15) with transient function dynamics (16), and the neural network update laws (18a), (18b). Then, if $\boldsymbol{\mathcal{A}_1}$, $\boldsymbol{\mathcal{A}_2}$, $\boldsymbol{\mathcal{A}_3}$ and $\boldsymbol{\mathcal{A}_4}$ hold and the condition

$$\rho > \frac{\overline{c}\Big(\overline{W}(\overline{g} + \overline{\dot{g}}) + \overline{\varepsilon}\Big) + \overline{\lambda}(\Lambda B(x))\overline{\delta}}{\underline{\lambda}(\Lambda B(\widehat{x}))} \tag{20}$$

is satisfied, a sliding mode is enforced and it yields

$$\begin{aligned}\widetilde{v}_1 = {}& B(\widehat{x})^\dagger \left[W^\top (g - \widehat{g}) + \widetilde{W}^\top \widehat{g} + \varepsilon(x) \right] + \\ & + B(\widehat{x})^\dagger B(x) \Delta u,\end{aligned} \tag{21}$$

with $\widetilde{v}_1$ being the equivalent control.

Proof. Consider the Lyapunov-like candidate function

$$V(x) = \frac{1}{2} s^\top s + \frac{1}{2}\mathrm{tr}\{\widetilde{W}^\top \Gamma_W^{-1} \widetilde{W}\}. \tag{22}$$

Then, its derivative with respect to time can be computed as

$$\dot{V}(x) = s^\top \dot{s} + \mathrm{tr}\{\widetilde{W}^\top \Gamma_W^{-1} \dot{\widetilde{W}}\}. \tag{23}$$

Substituting (17) and exploiting (11a) to write $\dot{\widetilde{W}} = -\dot{\widehat{W}}$, (23) can be then rewritten as

$$\begin{aligned}\dot{V}(x) = {}& s^\top \Lambda \left[W^\top g - \widehat{W}^\top \widehat{g} + \varepsilon(x) + B(x)\Delta u - B(\widehat{x}) v_1 \right] + \\ & - \mathrm{tr}\{\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\} \\ = {}& s^\top \Lambda \Big[W^\top (g - \widehat{g}) + \widetilde{W}^\top \widehat{g} + \varepsilon(x) + B(x)\Delta u + \\ & - B(\widehat{x}) v_1 \Big] - \mathrm{tr}\{\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\}.\end{aligned} \tag{24}$$

Using the definition of v_1 and rearranging the terms, one can write the above equation as

$$\begin{aligned}\dot{V}(x) = {}& s^\top \Lambda \Big[W^\top (g - \widehat{g}) + \varepsilon(x) \Big] + s^\top \Lambda B(x) \Delta u + \\ & - \rho s^\top \Lambda B(\widehat{x}) \frac{s}{\|s\|} + s^\top \Lambda \widetilde{W}^\top \widehat{g} - \mathrm{tr}\{\widetilde{W}^\top \Gamma_W^{-1} \dot{\widehat{W}}\}.\end{aligned} \tag{25}$$

Then, substituting the weight adaptation law (18a) and exploiting the trace property,

$$\begin{aligned}\dot{V}(x) = {}& s^\top \Lambda \Big[W^\top (g - \widehat{g}) + \varepsilon(x) \Big] + s^\top \Lambda B(x) \Delta u + \\ & - \rho s^\top \Lambda B(\widehat{x}) \frac{s}{\|s\|} + s^\top \Lambda \widetilde{W}^\top \widehat{g} - s^\top \Lambda \widetilde{W}^\top \widehat{g}.\end{aligned} \tag{26}$$

If assumptions $\boldsymbol{\mathcal{A}_1}$, $\boldsymbol{\mathcal{A}_2}$, $\boldsymbol{\mathcal{A}_3}$ and $\boldsymbol{\mathcal{A}_4}$ hold, the above equation can be upper bounded as

$$\begin{aligned}\dot{V}(x) \le {}& \mathbb{1}_{m\times 1}^\top \Lambda \mathbb{1}_{n\times 1} \Big[\overline{W}(\overline{g} + \overline{\dot{g}}) + \overline{\varepsilon} \Big] \|s\| + \\ & \overline{\lambda}(\Lambda B(x)) \overline{\delta} \|s\| - \rho \underline{\lambda}(\Lambda B(\widehat{x})) \|s\| \\ \le {}& \overline{c} \Big[\overline{W}(\overline{g} + \overline{\dot{g}}) + \overline{\varepsilon} \Big] \|s\| + \overline{\lambda}(\Lambda B(x)) \overline{\delta} \|s\| + \\ & - \rho \underline{\lambda}(\Lambda B(\widehat{x})) \|s\| = -\eta \|s\|,\end{aligned} \tag{27}$$

with

$$\eta = -\left[\overline{c}(\overline{W}(\overline{g} + \overline{\dot{g}}) + \overline{\varepsilon}) + \overline{\lambda}(\Lambda B(x)) \overline{\delta} \right] + \rho \underline{\lambda}(\Lambda B(\widehat{x})).$$

Hence, if condition (20) is satisfied, then $\dot{V}(x) \le -\eta \|s\| < 0$, $\forall x \in \Omega$, implying that a sliding mode $s = 0$ is enforced. When this happens, the fault can be estimated relying on the equivalent control signal (21), obtained by posing (17) to zero, which concludes the proof.

Therefore, by virtue of Theorem 2, the equivalent control signal in (21) can be used to implement a fault detection strategy which raises a flag $\phi_i \in \{0, 1\}$ if a fault is present on the ith control variable, i.e.,

$$\phi_i := \begin{cases} 1 & \text{if } \ |\widetilde{v}_{1,i}| \ge \tau_i, \\ 0 & \text{if } \ |\widetilde{v}_{1,i}| < \tau_i, \end{cases} \tag{28}$$

where $\widetilde{v}_{1,i}$ denote the ith component of $\widetilde{v}_1$, while τ_i is the ith element of the threshold vector $\tau(\widehat{x}) \in \mathbb{R}^m$. The latter is computed as

$$\tau(\widehat{x}) = B(\widehat{x})^\dagger \mathbb{1}_{n\times 1} \left(\overline{W}\overline{g} - \min(\widehat{W}_{ij}) \overline{\dot{g}} + \overline{\varepsilon} \right), \tag{29}$$

where $\overline{W}$, $\overline{g}$, $\overline{\dot{g}}$ and $\overline{\varepsilon}$ are the known bounds introduced in assumptions $\boldsymbol{\mathcal{A}_3}$ and $\boldsymbol{\mathcal{A}_4}$, while $\widehat{W}_{ij} \in \mathbb{R}$ is the entry at ith row and jth column of matrix $\widehat{W}$. Moreover, $\widetilde{v}_1$ is achieved using a first-order filter having in input the discontinuous signal v_1, see (Utkin and Shi, 1996) for further details. As for the term $\min(\widehat{W}_{ij})$, it does not affect the applicability of the fault detection mechanism proposed in (28). In fact, by virtue of Theorem 2, if condition (20) is satisfied, the weight estimation error $\widetilde{W}$ is always bounded, implying the boundedness of $\widehat{W}$.

5. NUMERICAL EXAMPLE

In this section, the proposal is assessed in simulation relying on a benchmark example, that is a DC motor (Raimondo et al., 2016), whose state space representation is given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -R_a L^{-1} x_1 - K_e^{-1} x_2 \\ K_t J_1^{-1} x_1 - F_r J_1^{-1} x_2 \end{bmatrix} + \begin{bmatrix} L^{-1} \\ 0 \end{bmatrix} (u + \Delta u), \quad (30)$$

where $x_1 \in \mathbb{R}$ is the electrical current, $x_2 \in \mathbb{R}$ is the angular speed, $u \in \mathbb{R}$ is the input voltage, while R_a, L, K_e, K_t, J_1, $F_r \in \mathbb{R}$ are the motor resistance, inductance, torque constant, back EMF constant, motor inertia and friction coefficient, respectively. In the simulations, $R_a = 1.203\ \Omega$, $L = 5.584 \times 10^{-3}$ H, $K_e = 8.574 \times 10^{-2}$V rad/s, $K_t = 1.0005\ K_e$, $J_1 = 1.3528 \times 10^{-4}$ Nms^2/rad and $F_r = 2.3396 \times 10^{-4}$ Nms/rad. For physical reasons all the quantities are bounded, as required by $\boldsymbol{\mathcal{A}_1}$.

The simulation window is 15 seconds, starting with initial condition $x(0) = [0\ 0]^\top$. The motor is controlled in order to reach a desired angular velocity $\omega^\star = 50$ rad/s using a PI controller, i.e., $u(t) = 0.1(\omega^\star - x_2(t)) + 0.4 \int_0^t (\omega^\star - x_2(y))\mathrm{d}y$, while being subject to a fault defined as $\Delta u(t) = 0.8\sin(2\pi(t-5))$ if $5 \le t \le 8$, $\Delta u(t) = 0.7\sin(2\pi(t-9)) + 0.2\cos(\frac{3}{2}\pi(t-9))$ if $9 \le t \le 14$, and $\Delta u(t) = 0$ for any other value of t.

The drift term used for computing the dynamics of the transient function z is estimated using a NN with 2 inputs, $L = 32$ hidden layers, and 2 outputs. The parameters of the network, i.e., $\widehat{W}$ and $\widehat{\Phi}$, are initialized with small random numbers and then updated by using the adaptation laws (18), setting $\Gamma_W = 1.5 \cdot I_{32\times 32}$ and $\Gamma_\Phi = 5 \cdot I_{2\times 2}$. The stabilizing part of the observer input is chosen again as a PI law $v_0(t) = k_\mathrm{P} e(t) + k_\mathrm{I} \int_0^t e(y)\mathrm{d}y$, with $k_\mathrm{P} = [0.2\ 0.2]$

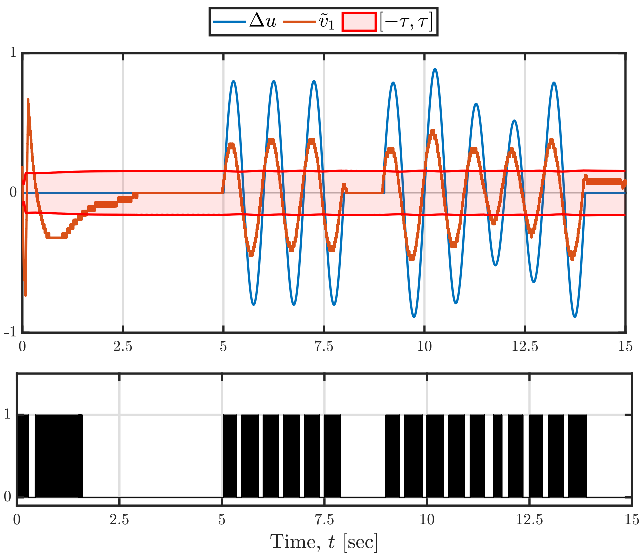


Fig. 2. Time evolution of the fault Δu (blue line), of the equivalent control $\widetilde{v}_1$ (orange line) and detection layer $[-\tau, \tau]$ (red line).

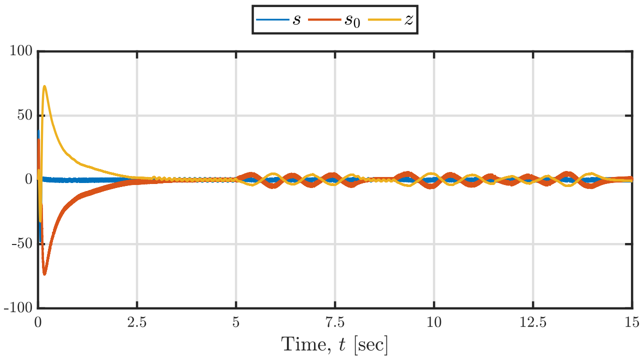
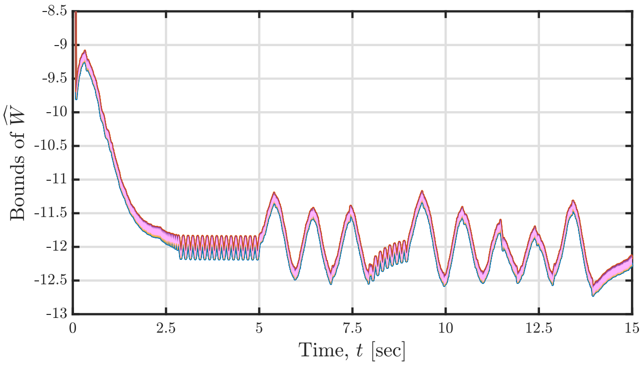
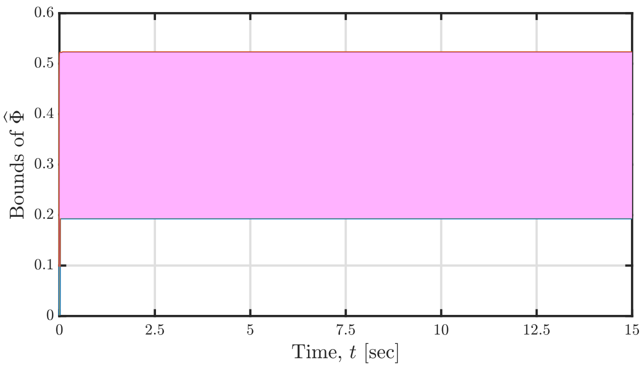


Fig. 3. Time evolution of the integral sliding variable s (blue line) and its components s_0 (orange line) and z (yellow line).



(a) outer layer weights, $\widehat{W}$



(b) inner layer weights, $\widehat{\Phi}$

Fig. 4. Time evolution of the bounds for the NN outer weights (a) and inner weights (b).

and $k_\mathrm{I} = [0.5\ 0.5]$, while the sliding variable parameter vector is chosen as $\Lambda = [5\ 5]$. Moreover, selecting the bounds $\overline{W} = 15$, $\overline{g} = \overline{\widehat{g}} = 1$, $\overline{\varepsilon} = 3$, $\overline{\delta} = 1$ and $\overline{c} = 10$ implies $\rho = 1.6$, which satisfies (20).

The results of the fault identification and detection using the settings specified above are presented in Fig. 2. In particular, as highlighted in (21), the proposal is able to estimate a quantity which includes the fault plus a term which depends on the NN estimation error. Nevertheless, the identification is accurate enough to successfully perform fault detection. This performance is further highlighted in Fig. 2 on the bottom, where the flags ϕ_i are reported: apart from a false positive detected during the first transient (2 seconds) in which weights were not adjusted, the detection strategy (28) is able to correctly detect the occurrences of fault. As for the integral sliding variable, Fig. 3 shows that, apart from the very first time instants, even without the exact knowledge of the drift term, a sliding mode is always enforced. Finally, one of the results derived by Theorem 2 is the boundedness of the weight estimation error (11a), that is the boundedness of the estimated weights $\widehat{W}$, as shown in Fig. 4.

6. CONCLUSIONS

In this paper, a novel NN-ISM based UIO for the detection of faults affecting the control input of nonlinear systems has been proposed. In general, to update the transient function of the integral sliding variable, the drift term should be known, which is not true in the considered case. Therefore, the sliding manifold has been designed relying on the approximation provided by a two-layer NN, whose weights are adapted following laws derived directly from theoretical analysis. Such an analysis also provides conditions for the enforcement of a sliding mode, hints about the shape of the estimation of the fault, and a formula to design the detection threshold. The performances of the proposal have been finally assessed in simulation with satisfactory results.

REFERENCES

Baimukashev, D., Rakhim, B., Rubagotti, M., and Varol, H.A. (2021). End-to-end deep fault tolerant control. *IEEE/ASME Transactions on Mechatronics*.

Bartolini, G., Caputo, W., Cecchi, M., Ferrara, A., and Fridman, L. (1997). Vibration damping in elastic robotic structures via sliding modes. *Journal of Robotic Systems*, 14(9), 675–696.

Bonassi, F., Farina, M., Xie, J., and Scattolini, R. (2022). On recurrent neural networks for learning-based control: recent results and ideas for future developments. *Journal of Process Control*, 114, 92–104.

Capisani, L.M., Ferrara, A., and Magnani, L. (2009). Design and experimental validation of a second-order sliding-mode motion controller for robot manipulators. *International Journal of Control*, 82(2), 365–377.

Capisani, L.M., Ferrara, A., and Pisu, P. (2010). Sliding mode observers for vision-based fault detection, isolation and identification in robot manipulators. In *Proceedings of the American Control Conference*, 4540–4545.

Cheng, L., Wang, Z., Jiang, F., and Li, J. (2021). Adaptive neural network control of nonlinear systems with unknown dynamics. *Advances in Space Research*, 67(3), 1114–1123.

Esfandiari, K., Abdollahi, F., and Talebi, H.A. (2022). *Neural Network-Based Adaptive Control of Uncertain Nonlinear Systems*. Springer.

Fei, J. and Lu, C. (2017). Adaptive sliding mode control of dynamic systems using double loop recurrent neural network structure. *IEEE Transactions on Neural Networks and Learning Systems*, 29(4), 1275–1286.

Ferrara, A. and Incremona, G.P. (2015). Design of an integral suboptimal second-order sliding mode controller for the robust motion control of robot manipulators. *IEEE Transactions on Control Systems Technology*, 23(6), 2316–2325.

Ferrara, A., Incremona, G.P., and Cucuzzella, M. (2019a). *Advanced and optimization based sliding mode control: Theory and applications*. SIAM.

Ferrara, A., Incremona, G.P., and Sangiovanni, B. (2019b). Tracking control via switched integral sliding mode with application to robot manipulators. *Control Engineering Practice*, 90, 257–266.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359–366.

Incremona, G.P. and Ferrara, A. (2019). Fault diagnosis for robot manipulators via vision servoing based suboptimal second order sliding mode. In *Proceedings of the 18th European Control Conference*, 3090–3095.

Incremona, G.P., Ferrara, A., and Magni, L. (2017). MPC for robot manipulators with integral sliding modes generation. *IEEE/ASME Transactions on Mechatronics*, 22(3), 1299–1307.

Isermann, R. (2005). *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media.

Isermann, R. (2011). *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*. Springer Science & Business Media.

Jarrou, A., Sauter, D., and Alami, K. (2019). Fault diagnosis and fault tolerant control based on model predictive control for nearly zero energy buildings. In *Proceedings of the 4th Conference on Control and Fault Tolerant Systems*, 219–225.

Lewis, F., Jagannathan, S., and Yesildirak, A. (1999). *Neural network control of robot manipulators and non-linear systems*. Taylor & Francis.

Punčochář, I. and Škach, J. (2018). A survey of active fault diagnosis methods. *IFAC-PapersOnLine*, 51(24), 1091–1098.

Raimondo, D.M., Marseglia, G.R., Braatz, R.D., and Scott, J.K. (2016). Closed-loop input design for guaranteed fault diagnosis using set-valued observers. *Automatica*, 74, 107–117.

Sacchi, N., Incremona, G.P., and Ferrara, A. (2022). Neural network-based practical/ideal integral sliding mode control. *IEEE Control Systems Letters*, 6, 3140–3145.

Sacchi, N., Incremona, G.P., and Ferrara, A. (2023). Sliding mode based fault diagnosis with deep reinforcement learning add-ons for intrinsically redundant manipulators. *International Journal of Robust and Nonlinear Control*, -(-), 1–19.

Scott, J.K., Findeisen, R., Braatz, R.D., and Raimondo, D.M. (2014). Input design for guaranteed fault diagnosis using zonotopes. *Automatica*, 50(6), 1580–1589.

Tai, N.T. and Ahn, K.K. (2010). A RBF neural network sliding mode controller for sma actuator. *International Journal of Control, Automation and Systems*, 8(6), 1296–1305.

Utkin, V. and Shi, J. (1996). Integral sliding mode in systems operating under uncertainty conditions. In *Proceedings of the 35th IEEE conference on decision and control*, volume 4, 4591–4596.

Zhan, Y. and Jiang, J. (1999). An interacting multiple-model based fault detection, diagnosis and fault-tolerant control approach. In *Proceedings of the 38th IEEE Conference on Decision and Control*, volume 4, 3593–3598.