This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2020.2966111

Peaking Attenuation of High-Gain Observers Using Adaptive Techniques: State Estimation and Feedback Control

Mehran Shakarami, Kasra Esfandiari, Amir Abolfazl Suratgar, and Heidar Ali Talebi

***Abstract*—This paper presents a new state estimation scheme using the second-level adaptation technique and Multiple High-Gain Observers (MHGO) for improving the transient response and attenuating undesired peaks of High-Gain Observers (HGOs). The proposed method considers state estimation as a convex combination of provided information by multiple high-gain observers. In this regard, it is shown that there exist some constant parameters in such combination that result perfect state estimation; then, an adaptive algorithm is employed for estimating those parameters. The stability of the proposed scheme and convergence of state estimation to the state of the plant are guaranteed. In addition, MHGO is proved to be able to provide a state estimation with smaller peaks in comparison to a single HGO. The performance of MHGO in the presence of measurement noise is also investigated. We consider existence of abrupt external disturbances as well. To alleviate the effects of those disturbances and attenuate their resulting peaking, we present a resetting scheme. Moreover, the output feedback control problem is considered, and it is demonstrated that a separation principle is valid for MHGO. Finally, simulation results illustrate that MHGO provides an accurate state estimation, and MHGO-based controller is able to recover the performance of state feedback controller.**

***Index Terms*—High-gain observers, second-level adaptation, output feedback, peaking phenomenon, measurement noise.**

I. Introduction

HIGH-Gain Observers (HGOs) are able to reconstruct system states from output measurements [1], [2]. There exists a vast amount of literature on employing such observers in solving different problems, such as control and state estimation of nonlinear systems [3], [4] and fault detection and isolation [5]. One factor that resulted in wide utilization of HGOs is satisfaction of the separation principle in HGO-based control problems, which was first introduced in [6], [7], and its results were improved in [8]. Despite their advantages, an inherent drawback of HGOs is existence of undesired peaks in their transient response, known as the peaking phenomenon. The interaction of this behaviour with system nonlinearities could result undesired performance of the observer and even instability of the closed-loop system [9]. To address this issue, we propose utilization of multiple HGOs and adaptive techniques.

An approach for recovering the state variables of dynamical systems is utilization of multi observers [10], [11], where multiple observers are employed to obtain multiple estimations of system states. Then, a supervisor selects one of the observers at each time instant based on an appropriate criteria. This method is able to provide preferable estimations; however, it has two drawbacks. Although the structure has multiple observers, only the obtained information from one observer is used at each time instant. Furthermore, it needs to employ c^n observers, where n is the number of states and $c > 1$, to result satisfactory performance [11]. Consequently, the required number of observers grows exponentially by the increase of state variables.

In adaptive systems based on a single adaptive model, the transient response is oscillatory when the system uncertainty is large [12]. One solution to this problem is utilization of multiple models for identification of the plant and a supervisor for selecting the closest model to the plant at any time instant [13]. On the other hand, this method has similar drawbacks to multi observers: a large number of models is required and the available information of all models is not efficiently employed. In [14] a novel scheme, called second-level adaptation technique, is presented for control of *Linear Time Invariant (LTI)* systems in companion form. In this approach, switching between multiple models is eliminated and a convex combination of all information is employed for parameter estimation and controller synthesis. The robustness of this method is investigated in [15]. Moreover, in different research, this concept is applied to systems with *unknown parameters* for solving problems raised in control theory [16], [17]. Besides, the state estimation problem for nonlinear systems is addressed in [18], [19]. In contrast, we provide robustness analysis and address the output feedback control problem.

In this paper, the second-level adaptation technique is employed together with Multiple HGOs (MHGO) to recover the state variables of a special class of nonlinear systems, and the obtained estimation is used in an observer-based controller. The main contributions of this paper are: I) A new methodology using multiple HGOs and the second-level adaptation technique is proposed for state estimation of nonlinear systems. II) By employing the properties of the proposed structure, the state estimation problem is converted to a parameter and state estimation problem, and the parameters

M. Shakarami is with the Engineering and Technology Institute Groningen, University of Groningen, 9747 AG, The Netherlands (e-mail: m.shakarami@rug.nl).

K. Esfandiari is with the Center for Systems Science, Yale University, New Haven, CT, USA (e-mail: kasra.esfandiari@yale.edu).

A. A. Suratgar and H. A. Talebi are with the Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran 15914, Iran (e-mail: a-suratgar@aut.ac.ir; alit@aut.ac.ir).

are estimated using a modification of the Recursive Least Squares (RLS) algorithm. III) It is proved that the obtained state estimation converges to the system state, and the peaking of MHGO observation error can become smaller than a single HGO. IV) In the presence of measurement noise, it is shown that MHGO can provide satisfactory performance by appropriately selection of some design parameters. V) A re-initialization scheme is presented to mitigate peaking due to abrupt output disturbances. VI) The observer-based control problem is considered, and it is proved that a separation principle is valid when MHGO state estimation is employed.

II. Problem Statement and Preliminaries

The considered special class of nonlinear systems and the structure of conventional HGOs are presented in this section.

A. Problem Formulation

Consider the following nonaffine nonlinear system

$$\begin{aligned} \dot{x} &= Ax + Bf(x,u) \\ y &= Cx \end{aligned} \tag{1}$$

where $x \in \mathcal{X} \subseteq \mathcal{R}^n$ is the state vector, $u \in \mathcal{U} \subseteq \mathcal{R}$ and $y \in \mathcal{Y} \subseteq \mathcal{R}$ denote the input and output of the system, respectively, and the $n \times n$ matrix A, the $n \times 1$ vector B, and the $1 \times n$ vector C are defined as follows

$$A = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, C^T = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In addition, $f : \mathcal{X} \times \mathcal{U} \to \mathcal{R}$ is a nonlinear function which is locally Lipschitz in its arguments over the domain of interest, and $f(0,0) = 0$; hence, the origin is an equilibrium point of the system [8].

Similar to [20], we assume that $f(\cdot,\cdot)$ is Lipschitz on $\mathcal{X} \times \mathcal{U}$. For satisfaction of this assumption, the set $\mathcal{X} \times \mathcal{U}$ will be considered to be compact.

B. High-Gain Observers

The structure of a single HGO is as follows [2], [20]

$$\dot{\hat{x}} = A\hat{x} + Bf_0(\hat{x},u) + H(y - C\hat{x}) \tag{2}$$

where $\hat{x}$ is the state estimation vector, $f_0(\cdot)$ is a saturated version of $f(\cdot)$ and agrees with that on $\mathcal{X} \times \mathcal{U}$. Also, $\epsilon \in (0,1]$ and $H = [\frac{\kappa_1}{\epsilon}, \frac{\kappa_2}{\epsilon^2}, \cdots, \frac{\kappa_n}{\epsilon^n}]^T$, where κ_is are chosen such that all the eigenvalues of the matrix $A - HC$ have negative real parts, i.e., $A - HC$ is Hurwitz. For investigating the convergence of $\hat{x}$ to x, let $\tilde{x} = x - \hat{x}$ and subtract (2) from (1) to obtain

$$\dot{\tilde{x}} = A_0\tilde{x} + B[f(x,u) - f_0(\hat{x},u)] \tag{3}$$

where $A_0 = A - HC$. As proved in [2], [20], there exists $\epsilon^* > 0$ such that for every $0 < \epsilon \le \epsilon^*$ and any admissible $x \in \mathcal{X}$ and $u \in \mathcal{U}$, the effect of $f(x,u) - f_0(\hat{x},u)$ on $\tilde{x}$ is rejected, and the state estimation converges to the state of the plant, i.e., $\lim_{t\to\infty} \tilde{x}(t) = 0$.

Throughout the paper, we need the following lemma.

Lemma 1 ([21]): Let $a_1, a_2, \cdots, a_m \in \mathcal{L}$ where $\mathcal{L}$ is a linear space. The intersection of all convex sets in $\mathcal{L}$ containing a_i is called the convex hull $\mathcal{K}$ of $\{a_i(i = 1, 2, \cdots, m)\}$ and any element of which, a', can be expressed as $a' = \sum_{i=1}^m \beta_i a_i$ where $\beta_i \in [0,1]$ is a constant satisfying $\sum_{i=1}^m \beta_i = 1$.

III. The Main Results

In this section, the structure of the proposed observer, MHGO, is presented, and its stability, performance, and robustness to measurement noise are investigated in detail. To counteract peaking resulted from sudden output disturbances, a modification of MHGO with resetting is also introduced. Furthermore, the output feedback control problem is addressed.

A. The Proposed MHGO

The proposed method reconstructs the state of the plant at any given time instant using full knowledge provided by multiple observers. In this regard, N HGOs with the structure of (4) are considered

$$\begin{aligned} \dot{\hat{x}}_i(\Lambda,t) &= A\hat{x}_i(\Lambda,t) + H(y(t) - C\hat{x}_i(\Lambda,t)) \\ &+ Bf_0(\sum_{i=1}^N \alpha_i\hat{x}_i(\Lambda,t), u(t)), \quad i = 1,\cdots,N \end{aligned} \tag{4}$$

where $\hat{x}_i$ denotes the state estimation from the ith observer with $\hat{x}_i(\Lambda,0) = \hat{x}_i(0)$, $\alpha_i \in [0,1]$ is a constant term satisfying $\sum_{i=1}^N \alpha_i = 1$, and $\Lambda = \begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_N\end{bmatrix}^T$. The notation $\hat{x}_i(\Lambda,t)$ is employed to show that $\hat{x}_i$s are functions of α_is. The final estimation is considered as a combination of N estimations as follows

$$\hat{x}_o(t) = \sum_{i=1}^N \alpha_i\hat{x}_i(\Lambda,t) \tag{5}$$

Considering (4), each observer utilizes state estimations provided by the other observers, and the final estimation is a combination of the individual observations (refer to (5)). This approach, employing multiple observations for calculation of the final estimation, assists in having a more accurate estimation. Nevertheless, we need to prove that the proposed observation structure (4) and (5) can estimate plant states, accurately. Toward this end, the following lemma is presented.

Lemma 2: Consider nonlinear system (1), N high-gain observers (4), equation (5), and let $x \in \mathcal{X}$ and $u \in \mathcal{U}$. If the initial conditions of HGOs (4) are selected such that $x(0)$ is in the convex hull $\mathcal{K}$ of $\hat{x}_i(0)$, then there exist some α_i^*s such that for $\alpha_i = \alpha_i^*$, the equality of $x(t) = \hat{x}_o(t)$ holds for all $t \ge 0$.

Proof. At $t = 0$, by employing Lemma 1, it can be seen that since $\hat{x}_i(0)$s are chosen such that $x(0)$ is in the convex hull $\mathcal{K}$ of $\{\hat{x}_i(0)(i = 1, 2, \cdots, N)\}$, some α_i^*s exist such that $x(0) = \sum_{i=1}^N \alpha_i^*\hat{x}_i(0)$, i.e., $\tilde{x}_o(0) = 0$ where $\tilde{x}_o = x - \hat{x}_o$ is the state estimation error. Now to provide the analysis for

$t > 0$, it is required to obtain the observation error dynamics. Subtracting (4) from (1), results in

$$\begin{aligned}\dot{\tilde{x}}_i(\Lambda, t) &= A_0\tilde{x}_i(\Lambda, t) \\ &\quad + B[f(x(t), u(t)) - f_0(\sum_{i=1}^{N} \alpha_i \hat{x}_i(\Lambda, t), u(t))]\end{aligned} \tag{6}$$

where $A_0 = A - HC$ and $\tilde{x}_i = x - \hat{x}_i$ denotes the observation error corresponding to the ith observer. In order to obtain the error dynamics of MHGO, one can use (5) and the fact that $\sum_{i=1}^{N} \alpha_i = 1$ to get $\tilde{x}_o(t) = \sum_{i=1}^{N} \alpha_i \tilde{x}_i(\Lambda, t)$. Therefore, noting that α_is are constant, we use (6) and obtain the time derivative of $\tilde{x}_o(t)$ as follows

$$\dot{\tilde{x}}_o = A_0\tilde{x}_o + B[f(x, u) - f_0(\hat{x}_o, u)] \tag{7}$$

Since f_0 agrees with f on $\mathcal{X} \times \mathcal{U}$, we conclude that $\tilde{x}_o = 0$ is an equilibrium point of the preceding equation. Besides, as stated before, choosing $\alpha_i = \alpha_i^*$ results in $\tilde{x}_o(0) = 0$; and therefore, $\tilde{x}_o(t) = 0$ for all $t \geq 0$. □

The following assumption is a result of Lemma 2.

Assumption 1: The initial conditions of HGOs (4), $\hat{x}_i(0)$s, are chosen such that the initial condition of plant (1), $x(0)$, is in their convex hull $\mathcal{K}$.

Remark 1: From convex analysis, we know that since $x(0) \in \mathcal{R}^n$, at least $N = n + 1$ observers are required for satisfaction of Assumption 1. In addition, one can choose $\hat{x}_i(0)$s such that $\mathcal{X} \subset \mathcal{K}$ to guarantee $x(0) \in \mathcal{K}$. It is also worth noting that in comparison to multi observers that need c^n observers with $c > 1$, MHGO requires fewer observers, especially when n is a large number.

From Lemma 2 it can be concluded that there exist some unknown constant α_i^*s that result in perfect state estimation. Therefore, the following equation can be obtained

$$x(t) = \sum_{i=1}^{N} \alpha_i^* \hat{x}_i(\Lambda^*, t), \quad \forall t \geq 0 \tag{8}$$

where $\Lambda^* = \begin{bmatrix}\alpha_1^* & \alpha_2^* & \cdots & \alpha_N^*\end{bmatrix}^T$. Although choosing α_is identical to α_i^*s results in perfect state estimation, such a selection is impossible due to the fact that α_i^*s are unknown. Thus, deriving an appropriate estimation of α_i^* is required. In other words, the understudy state estimation problem is transformed into a combination of state and parameter estimation problem.

In order to estimate α_i^*s, let us use the fact that $\sum_{i=1}^{N} \alpha_i^* = 1$ and rearrange (8) to obtain

$$\sum_{i=1}^{N} \alpha_i^* \tilde{x}_i(\Lambda^*, t) = 0 \tag{9}$$

By considering $\alpha_N^* = 1 - \sum_{i=1}^{N-1} \alpha_i^*$ and defining $\theta^* = \begin{bmatrix}\alpha_1^* & \alpha_2^* & \cdots & \alpha_{N-1}^*\end{bmatrix}^T$, we rewrite (9) as follows

$$\sum_{i=1}^{N-1} \alpha_i^* (\tilde{x}_i(\theta^*, t) - \tilde{x}_N(\theta^*, t)) = -\tilde{x}_N(\theta^*, t) \tag{10}$$

On the left hand side of (10), $\tilde{x}_i(\theta^*, t) - \tilde{x}_N(\theta^*, t)$ is equal to $\hat{x}_N(\theta^*, t) - \hat{x}_i(\theta^*, t)$. Moreover, one can use (4) and get

$$\dot{\hat{x}}_N(\theta^*, t) - \dot{\hat{x}}_i(\theta^*, t) = A_0(\hat{x}_N(\theta^*, t) - \hat{x}_i(\theta^*, t))$$

It can be seen that $\hat{x}_N(\theta^*, t) - \hat{x}_i(\theta^*, t)$ is the state of an LTI system, and in turn, it does not depend on θ^*; thus if a matrix M is defined such that its ith column is $\hat{x}_N(\theta^*, t) - \hat{x}_i(\theta^*, t)$, it is independent of θ^*, and one can write the regression form of (10) as follows

$$M(t)\theta^* = -\tilde{x}_N(\theta^*, t) \tag{11}$$

Although the matrix $M(t)$ is completely known, the right hand side of (11) is composed of x and $\hat{x}_N(\theta^*)$ which are unknown. Therefore, one cannot obtain an estimation of θ^* using conventional adaptive approaches merely based on (11).

To overcome the aforementioned difficulty, let us premultiply (11) by C, and get

$$CM(t)\theta^* = -\tilde{y}_N(\theta^*, t)$$

where $\tilde{y}_N(\theta^*, t) = y(t) - C\hat{x}_N(\theta^*, t)$. The right hand side of the preceding equation is still unknown since θ^* is not available, i.e., $\hat{x}_N(\theta^*, t)$ is unavailable. Therefore, the RLS algorithm which requires $\tilde{y}_N(\theta^*, t)$ for estimating θ^* cannot be utilized. Nonetheless, to find an estimation of θ^*, we propose employing a modified version of the RLS algorithm as follows

$$\begin{aligned}\dot{\hat{\theta}}(t) &= -P(t)M(t)^T C^T(\tilde{y}_N(\hat{\theta}, t) + CM(t)\hat{\theta}(t)) \\ \dot{P}(t) &= -P(t)M(t)^T C^T CM(t)P(t)\end{aligned} \tag{12}$$

where $\hat{\theta}$ represents an approximation of θ^*, $\tilde{y}_N(\hat{\theta}, t) = y(t) - C\hat{x}_N(\hat{\theta}, t)$, $\hat{\theta}(0) = \hat{\theta}_0$, and $P(0) = \gamma I$ with the positive constant γ and the identity matrix I. Note that the modified RLS algorithm uses $\tilde{y}_N(\hat{\theta}, t)$ instead of $\tilde{y}_N(\theta^*, t)$, and it will be shown later that this structure is appropriate for obtaining an accurate state estimation. It is worth noting that since $M(t)$ is independent of θ^*, its ith column is considered as $\hat{x}_N(\hat{\theta}, t) - \hat{x}_i(\hat{\theta}, t)$ that can be easily obtained.

Once $\hat{\theta}$ is calculated, it is employed for state estimation as

$$\begin{aligned}\dot{\hat{x}}_i(\hat{\theta}) &= A\hat{x}_i(\hat{\theta}) + H(y - C\hat{x}_i(\hat{\theta})) + Bf_0(\hat{x}_o, u) \\ \hat{x}_o &= \sum_{i=1}^{N-1} \hat{\alpha}_i \hat{x}_i(\hat{\theta}) + (1 - \sum_{i=1}^{N-1} \hat{\alpha}_i)\hat{x}_N(\hat{\theta})\end{aligned} \tag{13}$$

Therefore, the state estimation is obtained using two interconnected systems (12) and (13).

B. Performance Investigation

This section includes performance investigation of the proposed observation scheme, i.e., the interconnected systems (12) and (13). First, the stability and convergence of MHGO are shown in the following theorem; afterwards, we investigate that whether it can provide a better state estimation in comparison to a single HGO.

Theorem 1: Consider system (1), N high-gain observers (13), and the modified RLS algorithm (12), and let $x \in \mathcal{X}$ and $u \in \mathcal{U}$. Then there exists $\epsilon^* > 0$ such that for $0 < \epsilon \leq \epsilon^*$, $\hat{\theta}$ and P are bounded, $\hat{x}_i$s are uniformly ultimately bounded, and $\hat{x}_o$ converges to x.

Proof. Consider the scaled estimation error $\eta_{(i)} := \frac{x_{(i)} - \hat{x}_{o(i)}}{\epsilon^{n-i}}$ where $x_{(i)}$ and $\hat{x}_{o(i)}$ represent the ith elements of x and $\hat{x}_o$, respectively. As a result, one can obtain

$$D(\epsilon)\eta = x - \hat{x}_o \tag{14}$$

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2020.2966111

where $\eta = [\eta_{(1)}, \cdots, \eta_{(n)}]^T$ and $D(\epsilon) = \text{diag}(\epsilon^{n-1}, \cdots, \epsilon, 1)$. By using $\sum_{i=1}^{N} \hat{\alpha}_i = 1$, the scaled error (14), and the fact that $CD(\epsilon) = \epsilon^{n-1}C$, the dynamics of MHGO, including the modified RLS algorithm (12) and N high-gain observers (13), can be rewritten as follows

$$\eta = \sum_{i=1}^{N-1} \hat{\alpha}_i \eta_i(\hat{\theta}) + (1 - \sum_{i=1}^{N-1} \hat{\alpha}_i)\eta_N(\hat{\theta}) \tag{15}$$

$$\dot{\hat{\theta}} = -\epsilon^{2(n-1)} P M_1^T C^T C(\eta_N(\hat{\theta}) + M_1\hat{\theta}) \tag{16}$$

$$\dot{P} = -\epsilon^{2(n-1)} P M_1^T C^T C M_1 P \tag{17}$$

$$\epsilon\dot{\eta}_i(\hat{\theta}) = A_1\eta_i(\hat{\theta}) + \epsilon B[f(x,u) - f_0(x - D(\epsilon)\eta, u)] \tag{18}$$

where $D(\epsilon)\eta_i = x - \hat{x}_i$ is the scaled state estimation error for the ith observer, the ith column of M_1 is $\eta_i(\hat{\theta}) - \eta_N(\hat{\theta})$, and $A_1 = \epsilon D(\epsilon)^{-1} A_0 D(\epsilon)$. It is worth noting that A_1 is a Hurwitz matrix since κ_is are chosen such that A_0 is Hurwitz.

Since $0 \leq P(t)$ and $\dot{P}(t) \leq 0$, it can be seen that $P(t) \leq P(0) = \gamma I$ and $P(t)$ is bounded. Moreover, using the fact that $P(t)$ is symmetric, we have $\|P(t)\|^2 = \lambda_{\max}(P(t))^2$; thus, $\|P(t)\| \leq \gamma$.

We use the definition of M_1 and rewrite the regression form of η from (15) as follows

$$\eta = M_1\hat{\theta} + \eta_N(\hat{\theta}) \tag{19}$$

Substituting the preceding equation into (16) results in

$$\dot{\hat{\theta}} = -\epsilon^{2(n-1)} P M_1^T C^T C\eta \tag{20}$$

On the other hand, by using (18) and the fact that the ith column of M_1 is $\eta_i(\hat{\theta}) - \eta_N(\hat{\theta})$, we have

$$\epsilon\dot{M}_1 = A_1 M_1 \tag{21}$$

Now by using (18), (20), and (21), the derivative of (19) is

$$\begin{aligned}\dot{\eta} = \frac{1}{\epsilon}A_1\eta - \epsilon^{2(n-1)} M_1 P M_1^T C^T C\eta \\ + B[f(x,u) - f_0(x - D(\epsilon)\eta, u)]\end{aligned} \tag{22}$$

In order to investigate the convergence of $\hat{x}_o(t)$ to $x(t)$, a Lyapunov function candidate $V(\eta) = \eta^T P_1 \eta$ is considered where P_1 is the positive definite matrix satisfying $A_1^T P_1 + P_1 A_1 = -I$. Therefore, the derivate of $V(\eta)$ can be obtained using (22) as follows

$$\begin{aligned}\dot{V}(\eta) = -\frac{1}{\epsilon}\eta^T\eta - 2\epsilon^{2(n-1)}\eta^T P_1 M_1 P M_1^T C^T C\eta \\ + 2\eta^T P_1 B[f(x,u) - f_0(x - D(\epsilon)\eta, u)]\end{aligned} \tag{23}$$

Employing $\|D(\epsilon)\| = 1$, the fact that f is Lipschitz on $\mathcal{X} \times \mathcal{U}$ and f_0 is a saturated version of which and agrees with that on this domain, we conclude that

$$\|f(x,u) - f_0(x - D(\epsilon)\eta, u)\| \leq L_1\|\eta\|, \forall (x,u) \in \mathcal{X} \times \mathcal{U} \tag{24}$$

where $L_1 > 0$ is a Lipschitz constant. One can use (23), (24), and $\|B\| = 1$ to express $\dot{V}(\eta)$ as

$$\begin{aligned}\dot{V}(\eta) \leq -\frac{1}{\epsilon}\|\eta\|^2 + 2\epsilon^{2(n-1)}\|P_1\|\|M_1\|^2\|P\|\|\eta\|^2 \\ + 2L_1\|P_1\|\|\eta\|^2\end{aligned}$$

By defining $\epsilon^* := 1/(4L_1\|P_1\|)$ and choosing $0 < \epsilon \leq \epsilon^*$, we have

$$\dot{V}(\eta) \leq -\frac{1}{2\epsilon}\|\eta\|^2 + 2\epsilon^{2(n-1)}\|P_1\|\|M_1\|^2\|P\|\|\eta\|^2 \tag{25}$$

As it was shown before, $\|P(t)\| \leq \gamma$. Moreover, from (21), we know that $M_1(t) = e^{\frac{1}{\epsilon}A_1 t}M_1(0)$. Since A_1 is a Hurwitz matrix, a Lyapunov function candidate $W = \text{Tr}[e^{\frac{1}{\epsilon}A_1^T t} P_1 e^{\frac{1}{\epsilon}A_1 t}]$ can be considered to obtain $\dot{W} = -\frac{1}{\epsilon}\text{Tr}[e^{\frac{1}{\epsilon}A_1^T t} e^{\frac{1}{\epsilon}A_1 t}]$. Furthermore, we have

$$\lambda_{\min}(P_1)\,\text{Tr}[e^{\frac{1}{\epsilon}A_1^T t} e^{\frac{1}{\epsilon}A_1 t}] \leq W \leq \lambda_{\max}(P_1)\,\text{Tr}[e^{\frac{1}{\epsilon}A_1^T t} e^{\frac{1}{\epsilon}A_1 t}] \tag{26}$$

where $\lambda_{\min}(P_1)$ and $\lambda_{\max}(P_1)$ are the smallest and largest eigenvalues of P_1, respectively. By using the above equation, one has $\dot{W} \leq -\frac{1}{\epsilon\lambda_{\max}(P_1)}W$, and in turn,

$$W(t) \leq e^{-\frac{1}{\epsilon\lambda_{\max}(P_1)}t}W(0) \tag{27}$$

Now, one can consider the following inequalities

$$\|e^{\frac{1}{\epsilon}A_1 t}\|^2 \leq \text{Tr}[e^{\frac{1}{\epsilon}A_1^T t} e^{\frac{1}{\epsilon}A_1 t}] \leq n\|e^{\frac{1}{\epsilon}A_1 t}\|^2$$

Thus, (26), (27), and the above equation can be used to get

$$\|e^{\frac{1}{\epsilon}A_1 t}\| \leq k e^{-\frac{\lambda}{\epsilon}t} \tag{28}$$

with $k = \sqrt{n\lambda_{\max}(P_1)/\lambda_{\min}(P_1)}$ and $\lambda = 1/(2\lambda_{\max}(P_1))$. Using (28), we have $\|M_1(t)\| \leq k\|M_1(0)\|e^{-\frac{\lambda}{\epsilon}t}$. Then, the obtained upper bounds of $\|P(t)\|$ and $\|M_1(t)\|$ and (25) can be utilized to get $\dot{V}(\eta) \leq -\frac{1}{2\epsilon}\|\eta\|^2 + \rho e^{-2\frac{\lambda}{\epsilon}t}\|\eta\|^2$ with $\rho = 2k^2\gamma\epsilon^{2(n-1)}\|P_1\|\|M_1(0)\|^2$. Note that the following inequality always holds.

$$\lambda_{\min}(P_1)\|\eta\|^2 \leq V(\eta) \leq \lambda_{\max}(P_1)\|\eta\|^2 \tag{29}$$

Thus, (29) can be used to obtain

$$\dot{V}(\eta) \leq (-\frac{1}{2\epsilon\lambda_{\max}(P_1)} + \frac{1}{\lambda_{\min}(P_1)}\rho e^{-2\frac{\lambda}{\epsilon}t})V(\eta)$$

By solving the preceding equation, we have

$$V(t) \leq e^{-\frac{1}{2\epsilon\lambda_{\max}(P_1)}t} e^{\frac{\epsilon}{2\lambda\lambda_{\min}(P_1)}\rho(1 - e^{-2\frac{\lambda}{\epsilon}t})}V(0)$$

Considering $1 - e^{-2\frac{\lambda}{\epsilon}t} \leq 1$, the preceding equation is rewritten as follows

$$V(t) \leq k_1 e^{-\frac{1}{2\epsilon\lambda_{\max}(P_1)}t}V(0) \tag{30}$$

with $k_1 = e^{\frac{\epsilon}{2\lambda\lambda_{\min}(P_1)}\rho}$. As a result, $\lim_{t\to\infty}\eta(t) = 0$, and $\hat{x}_o(t)$ converges to $x(t)$.

To show that $\hat{\theta}$ is bounded, (20) is used to get

$$\hat{\theta}(t) = \hat{\theta}_0 - \epsilon^{2(n-1)}\int_0^t P(\tau)M_1^T(\tau)C^T C\eta(\tau)d\tau \tag{31}$$

Furthermore, one can get the following inequality by employing (29) and (30)

$$\|\eta(t)\| \leq k_2 e^{-\frac{1}{4\epsilon\lambda_{\max}(P_1)}t}\|\eta(0)\| \tag{32}$$

where $k_2 = \sqrt{k_1\lambda_{\max}(P_1)/\lambda_{\min}(P_1)}$. By utilizing the preceding equation, the upper bounds of $\|P(t)\|$ and $\|M_1(t)\|$, and (31), one can get $\|\hat{\theta}(t)\| \leq \|\hat{\theta}_0\| + k_3\int_0^t e^{-(\frac{\lambda}{\epsilon} + \frac{1}{4\epsilon\lambda_{\max}(P_1)})\tau}d\tau$ with $k_3 = kk_2\gamma\epsilon^{2(n-1)}\|M_1(0)\|\|\eta(0)\|$. We conclude from $\|\hat{\theta}(t)\| \leq \|\hat{\theta}_0\| + k_3\frac{4\epsilon\lambda_{\max}(P_1)}{4\lambda\lambda_{\max}(P_1)+1}$ that $\hat{\theta}$ is bounded.

To show that $\hat{x}_i$s are bounded, using the boundedness of x, we need to show that η_is are bounded. For that, a Lyapunov function candidate $V_i(\eta_i) = \eta_i^T P_1 \eta_i$ and (18) are used to get $\dot{V}_i(\eta_i) \leq -\frac{1}{\epsilon}\|\eta_i\|^2 + 2L_1\|P_1\|\|\eta_i\|\|\eta\|$. By using (32), we concluded that $\dot{V}_i(\eta_i) < 0$ for $\|\eta_i\| > 2\epsilon L_1 k_2 \|P_1\|\|\eta(0)\|$, and in turn, η_is are uniformly ultimately bounded. □

From Theorem 1, we see that the proposed observer is stable and its estimation converges to system state. On the other hand, the main purpose of the proposed observer is obtaining a better estimation in comparison to a single HGO. In order to investigate that, the analysis of estimation error of MHGO and a single HGO needs to be performed. Toward this end, it is required to consider the following assumption.

Assumption 2: The initial conditions of HGOs (4) are chosen such that the matrix $M(0)M(0)^T$ has full rank.

Remark 2: Since $M(0)$ is $n \times (N-1)$ and $N-1 \geq n$, Assumption 2 is not restrictive.

The performance analysis of MHGO and its comparison to a single HGO are performed in the following lemma.

Lemma 3: Let conditions of Theorem 1 be satisfied and Assumptions 1 and 2 hold. If $\gamma := \xi/\epsilon^{2(n-1)}$ with the positive constant ξ, then,

(i) there exist $\xi_1^* > 0$ and $\epsilon_1^* > 0$ such that by choosing $\xi > \xi_1^*$ and $0 < \epsilon < \epsilon_1^*$, the state estimation errors of MHGO and a single HGO can peak to $\mathcal{O}(\|\tilde{x}_o(0)\|/(\xi\epsilon^{n-1}))$ and $\mathcal{O}(\|\tilde{x}(0)\|/\epsilon^{n-1})$, respectively, where $\tilde{x}_o(0) = \tilde{x}(0)$ are the initial estimation errors.

(ii) if $N = n+1$, there exist $\xi_2^* > 0$ and $\epsilon_2^* > 0$ such that by choosing $\xi > \xi_2^*$ and $0 < \epsilon < \epsilon_2^*$, the norm of parameter estimation error $\tilde{\theta} = \hat{\theta} - \theta^*$ is less or equal to $\mathcal{O}(\|\tilde{\theta}_0\|/\xi)$ where $\tilde{\theta}_0$ is the initial parameter estimation error.

Proof. See the Appendix. □

It can be seen from Lemma 3 that by choosing a large enough ξ, which is equivalent to taking γ large, the peak of MHGO state estimation error can become arbitrary smaller than the peak of a single HGO estimation error. As a result, the proposed MHGO may provide state estimations with more preferable transient response in comparison to a single HGO. Other approaches for attenuation of peaking are using hybrid modifications and projection [22] or introducing saturation into a low-power modification of HGO [23]. In contrast to [22], the considered plant is more general and is not limited to interconnected second order plants. Also, peaking in [23] can grow to $\mathcal{O}(1/\epsilon)$; however, for the MHGO, peaking can become arbitrary small. It is also worth noting that since $\alpha_i^* \in [0,1]$, one could employ projection to compel $\hat{\alpha}_i \in [0,1]$, and in turn, would need to appropriately modify the preceding analysis.

Remark 3: One can see from Theorem 1 and Lemma 3 that Assumptions 1 and 2 are not required for the convergence of state estimation. However, if the initial conditions of HGOs are selected such that they are held, by selecting a large enough γ, the proposed observation scheme can provide better state estimations than a single HGO.

Remark 4: It can be seen from the presented analysis that $x(0) \in \mathcal{K}$ is required for the existence of θ^* which satisfies $x(0) = M(0)\theta^* + \hat{x}_N(0)$. Therefore, Assumption 1 can be relaxed into selection of $\hat{x}_i(0)$ such that $M(0)$ is full row rank, i.e., $x(0) - \hat{x}_N(0) \in \text{Range}(M(0))$.

C. The Effect of Measurement Noise

In the theory of HGOs, it is well-known that there exists a trade-off between the speed of state estimation and sensitivity to noise. In [24], it is shown that when there exists measurement noise, the state estimation error of HGO converges to an ultimate bound $\mathcal{O}(\mu/\epsilon^{n-1})$ where μ is an upper bound of the norm of noise. Consequently, even though choosing smaller ϵ results in faster convergence of state estimation, it produces bigger transient peaks and bigger ultimate estimation error bound. In this section, we assume that the output measurement is contaminated by noise as follows

$$y = Cx + \nu \tag{33}$$

where ν is the bounded measurement noise such that $\|\nu\| \leq \mu$. The goal is to investigate the performance of MHGO in the presence of noise; hence, the following lemma is presented.

Lemma 4: Let conditions of Theorem 1 be satisfied and the output of plant be contaminated by noise as (33). Then, there exist positive constants ϵ^*, k_1, k_2, k_3, and λ such that by choosing $0 < \epsilon \leq \epsilon^*$, we have

$$\begin{aligned}\|\tilde{x}_o(t)\| \leq{}& \frac{k_1}{\epsilon^{n-1}} e^{-\frac{\lambda}{2\epsilon}t}\|\tilde{x}_o(0)\| + k_2\gamma\epsilon^n(e^{-\frac{\lambda}{2\epsilon}t} - e^{-2\frac{\lambda}{\epsilon}t})\mu \\ &+ \frac{k_3}{\epsilon^{n-1}}(1 - e^{-\frac{\lambda}{2\epsilon}t})\mu \end{aligned}\tag{34}$$

where $\tilde{x}_o$ is the state estimation error of MHGO and $\|\nu\| \leq \mu$.

Proof. See the Appendix. □

From (34), it is obvious that the ultimate estimation error bound of MHGO is $\mathcal{O}(\mu/\epsilon^{n-1})$, which is the same as a single HGO, obtained in [24]. On the other hand, the measurement noise and initial estimation error have linear effects on the right hand side of (34). As a result, even though choosing a large γ reduces the effect of initial estimation error and the peaking phenomenon (as it was shown in Lemma 3), it will also increase the transient effect of noise. In other words, there is a trade-off for choosing γ. To reduce the effect of $\tilde{x}_o(0)$, we need to choose a large γ, and for mitigating the effect of noise, γ should be set small enough. One method for selecting the design parameters ϵ and γ is to choose a large ϵ to reduce the ultimate estimation error bound and set γ large enough to improve the transient response.

D. Re-initialization Against Abrupt Disturbances

In the proposed observation scheme, all of the individual observers converge to each other in the long run, i.e., $M_1(t) \to 0$ as $t \to \infty$. This means that after the transient response, $\hat{x}_o$ is equal to a single HGO's estimation. From a practical point of view, abrupt disturbances may occur at the output of the plant after the transient response. In this case, as shown in the previous section, MHGO is still stable; however, it is not able to improve the transient response related to the disturbances since $M(t)$ has already become small. For addressing this issue, we propose a re-initialization scheme to enhance transient response resulting from abrupt disturbances.

To achieve the mentioned goals, we reset the state estimation of the ith observer to its initial value, i.e., $\hat{x}_i(t_k) = \hat{x}_i(0)$ where t_k is the time instant of the kth re-initialization; thus,

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2020.2966111

we have $M(t_k) = M(0)$. For the RLS algorithm, $P(t_k) = \gamma I$ can be selected; however, if we choose $\hat{\theta}(t_k) = \hat{\theta}_0$, the state estimation $\hat{x}_o$ will reset to the initial value $\hat{x}_o(0)$ and this will cause discontinuities in $\hat{x}_o$. To avoid that, $\hat{\theta}(t_k)$ should be chosen such that $\hat{x}_o(t_k) = \hat{x}_o(t_k^-)$ where t_k^- is the time instant before the kth resetting, i.e., $M(0)\hat{\theta}(t_k) + \hat{x}_N(0) = \hat{x}_o(t_k^-)$. Considering this discussion and Assumption 2, the following re-initialization scheme is presented

$$\begin{aligned} &\hat{x}_i(t_k) = \hat{x}_i(0), i = 1, \cdots, N \\ &P(t_k) = \gamma I \\ &\hat{\theta}(t_k) = M(0)^T (M(0)M(0)^T)^{-1} (\hat{x}_o(t_k^-) - \hat{x}_N(0)) \end{aligned} \tag{35}$$

The following lemma analyzes the performance of MHGO under the presented re-initialization scheme.

Lemma 5: Let conditions of Theorem 1 be satisfied and Assumption 2 holds. Also, there exists abrupt output disturbance ν given by (33) which is bounded as $\|\nu\| \leq \mu$. Then, there exist positive constants ϵ^*, k_1, k_2, k_3, and λ such that by choosing $0 < \epsilon \leq \epsilon^*$, the state estimation error of MHGO with resetting (35) satisfies

$$\begin{aligned} &\|\tilde{x}_o(t)\| \leq \frac{k_1}{\epsilon^{n-1}} e^{-\frac{\lambda}{2\epsilon}t} \|\tilde{x}_o(0)\| \\ &+ \sum_{j=1}^{k-1} e^{-\frac{\lambda}{2\epsilon}(t-t_j)} \Big[k_2 \gamma \epsilon^n (e^{-\frac{\lambda}{2\epsilon}(t_j - t_{j-1})} - e^{-2\frac{\lambda}{\epsilon}(t_j - t_{j-1})}) \\ &+ \frac{k_3}{\epsilon^{n-1}} (1 - e^{-\frac{\lambda}{2\epsilon}(t_j - t_{j-1})}) \Big] \mu \\ &+ k_2 \gamma \epsilon^n (e^{-\frac{\lambda}{2\epsilon}(t - t_{k-1})} - e^{-2\frac{\lambda}{\epsilon}(t - t_{k-1})}) \mu \\ &+ \frac{k_3}{\epsilon^{n-1}} (1 - e^{-\frac{\lambda}{2\epsilon}(t - t_{k-1})}) \mu \end{aligned} \tag{36}$$

where $t \in [t_{k-1}, t_k)$, t_j is the time instant of the jth resetting for $j = 0, 1, \cdots, k-1$, and $t_0 = 0$.

Proof. See the Appendix. □

Since the time instant of disturbance occurrence is unknown, it is required to consider a scheme for determining the resetting time. In this regard, the $e^{-\frac{\delta}{\epsilon}t}$-weighed $\mathcal{L}_2$ norm of the output estimation error can be utilized as a monitoring signal, i.e., $\vartheta(t) = \int_0^t e^{-\frac{\delta}{\epsilon}(t-\tau)} (y(\tau) - C\hat{x}_o(\tau))^2 d\tau$ with the design parameter $\delta > 0$ [25]. This signal can be implemented using a linear filter as follows

$$\dot{\vartheta} = -\frac{\delta}{\epsilon}\vartheta + (y - C\hat{x}_o)^2, \vartheta(0) = 0 \tag{37}$$

As a result, t_k is defined as the time at which $\vartheta(t) \geq \bar{\vartheta}$ where $\bar{\vartheta} > 0$ is a design parameter, and we choose $\vartheta(t_k) = 0$ to re-initialize the filter at the resetting time. It is worth noting that the design parameters δ and $\bar{\vartheta}$ should be selected proportionate to the application to prevent false resetting signals (due to very small $\bar{\vartheta}$ or very large δ) or missing out disturbance occurrences (for very large or small $\bar{\vartheta}$ and δ, respectively).

Remark 5: We need to assure that the summation in (36) is bounded even if the number of resettings tends to infinity. Let the minimum switching time interval be $T = \min\{t_j - t_{j-1}\}$ for $j = 0, 1, \cdots, k-1$, then $t \in [t_{k-1}, t_k)$ can be used to get $t - t_j \geq (k - 1 - j)T$. By noting $e^{-\frac{\lambda}{2\epsilon}(t_j - t_{j-1})} - e^{-2\frac{\lambda}{\epsilon}(t_j - t_{j-1})} \leq 0.5$ and $1 - e^{-\frac{\lambda}{2\epsilon}(t_j - t_{j-1})} \leq 1$, it is only required to investigate $\lim_{k\to\infty} \sum_{j=1}^{k-1} e^{-\frac{\lambda}{2\epsilon}(t-t_j)}$, which is

$$\begin{aligned} \lim_{k\to\infty} \sum_{j=1}^{k-1} e^{-\frac{\lambda}{2\epsilon}(t-t_j)} &\leq \lim_{k\to\infty} \sum_{j=1}^{k-1} e^{-\frac{\lambda}{2\epsilon}(k-1-j)T} \\ &= \lim_{k\to\infty} \sum_{j=0}^{k-2} \left(e^{-\frac{\lambda}{2\epsilon}T}\right)^j \end{aligned}$$

The above summation converges to a bounded value for all $T > 0$, which is always satisfied based on its definition. It is worth noting that the switching excludes the Zeno behavior, which means that a subsequence of switching times t_{j_k} such that $\lim_{k\to\infty}(t_{j_k} - t_{j_{k-1}}) = 0$ does not occur. To show this, considering the structure of filter (37), it suffices the input signal $y - C\hat{x}_o$ to be bounded in finite time. This is guaranteed as x belongs to the compact set $\mathcal{X}$, the disturbance ν is bounded, and $\hat{x}_o$ is continuous.

E. Control

This section considers the output feedback control problem using the state estimation of MHGO. The proposed control strategy is inspired by the following theorem which addresses the output feedback control problem using a single HGO.

Theorem 2 ([2]): Let $u = \omega(x)$ with the locally Lipschitz function $\omega(x)$, which is globally bounded and $\omega(0) = 0$, to be a state feedback controller that asymptotically stabilizes the origin of (1) with the region of attraction Ω. Now, consider the output feedback controller $u = \omega(\hat{x})$ where $\hat{x}$ is generated by the single HGO (2). Let $\mathcal{S}$ be any compact set in the interior of Ω and $\mathcal{Q}$ be any compact subset of $\mathcal{R}^n$. Then,

(i) there exists $\epsilon_1^* > 0$ such that, for every $0 < \epsilon \leq \epsilon_1^*$, the solution $(x(t), \hat{x}(t))$ of the closed-loop system, starting in $\mathcal{S} \times \mathcal{Q}$, is bounded for all $t \geq 0$.

(ii) given any $\mu > 0$, there exist $\epsilon_2^* > 0$ and $T_2 > 0$, both dependent on μ, such that, for every $0 < \epsilon \leq \epsilon_2^*$, the solutions of the closed-loop system, starting in $\mathcal{S} \times \mathcal{Q}$, satisfy $\|x(t)\| \leq \mu$ and $\|\hat{x}(t)\| \leq \mu$ for all $t \geq T_2$.

(iii) given any $\mu > 0$, there exists $\epsilon_3^* > 0$, dependent on μ, such that, for every $0 < \epsilon \leq \epsilon_3^*$, the solutions of the closed-loop system, starting in $\mathcal{S} \times \mathcal{Q}$, satisfy $\|x(t) - x_r(t)\| \leq \mu$ for all $t \geq 0$, where $x_r(t)$ is the solution of system (1) under $u = \omega(x)$, starting at $x(0)$.

(iv) if the origin of system (1) under $u = \omega(x)$ is exponentially stable and $f(x, u)$ is continuously differentiable in some neighborhood of $x = 0$, then there exists $\epsilon_4^* > 0$ such that, for every $0 < \epsilon \leq \epsilon_4^*$, the origin of the closed-loop system is exponentially stable and $\mathcal{S} \times \mathcal{Q}$ is a subset of its region of attraction.

Theorem 2 posits that by choosing a sufficiently small ϵ, the output feedback controller that employs $\hat{x}$ from (2), recovers the performance of the state feedback controller. In the sequel, a theorem corresponding to the performance of the observer-based control scheme that utilizes state estimation obtained from the proposed MHGO, is provided.

Theorem 3: Let conditions of Theorem 2 be satisfied, and consider the proceeding MHGO

$$\dot{\hat{x}}_i(\hat{\theta}) = A\hat{x}_i(\hat{\theta}) + H(y - C\hat{x}_i(\hat{\theta})) + Bf_0(\hat{x}_o, \omega(\hat{x}_o))$$
$$\hat{x}_o = \sum_{i=1}^{N-1} \hat{\alpha}_i \hat{x}_i(\hat{\theta}) + (1 - \sum_{i=1}^{N-1} \hat{\alpha}_i)\hat{x}_N(\hat{\theta})$$

where $i = 1, \cdots, N$ and $\hat{x}_o$ denotes the reconstructed system state; moreover, $\hat{\theta} = \begin{bmatrix}\hat{\alpha}_1 & \cdots & \hat{\alpha}_{N-1}\end{bmatrix}^T$ is obtained from the modified RLS algorithm (12). Then, the output feedback controller $u = \omega(\hat{x}_o)$ recovers the performance of the state feedback controller, in the sense of Theorem 2. Moreover, $\hat{\theta}$ and P are bounded and $\hat{x}_i$s are uniformly ultimately bounded.

Proof. By using the scaled estimation error as defined in (14) and the fact that $CD(\epsilon) = \epsilon^{n-1}C$, one can rewrite the dynamics of the overall closed-loop system as follows

$$\dot{x} = Ax + Bf(x, \omega(x - D(\epsilon)\eta)) \quad (38)$$
$$\eta = M_1\hat{\theta} + \eta_N(\hat{\theta}) \quad (39)$$
$$\dot{\hat{\theta}} = -\epsilon^{2(n-1)} P M_1^T C^T C\eta \quad (40)$$
$$\dot{P} = -\epsilon^{2(n-1)} P M_1^T C^T C M_1 P \quad (41)$$
$$\epsilon\dot{\eta}_i = A_1\eta_i + \epsilon B\Delta(x, D(\epsilon)\eta) \quad (42)$$

where $A_1 = \epsilon D(\epsilon)^{-1}(A - HC)D(\epsilon)$, the ith column of M_1 is $\eta_i(\hat{\theta}) - \eta_N(\hat{\theta})$, and $\Delta(x, D(\epsilon)\eta) = f(x, \omega(x - D(\epsilon)\eta)) - f_0(x - D(\epsilon)\eta, \omega(x - D(\epsilon)\eta))$.

Similar to the proof of Theorem 1, since $\dot{P} \le 0$ and $P \ge 0$, the matrix P is bounded and $\|P\| \le \gamma$. In addition, one can consider (39) and use (21), (40), and (42) to obtain

$$\epsilon\dot{\eta} = A_1\eta - \epsilon^{2n-1} M_1 P M_1^T C^T C\eta + \epsilon B\Delta(x, D(\epsilon)\eta) \quad (43)$$

It can be seen that (38) and (43) construct a singular perturbation model. Therefore, according to the perturbation theory, the analysis is divided into two stages: finding the reduced model and finding the boundary-layer model. To obtain the reduced system, it is required to set $\epsilon = 0$ in (38) and (43). Since A_1 is a Hurwitz matrix, it has full rank. Thus by setting $\epsilon = 0$ and performing some basic manipulations, one can get $\dot{x} = Ax + Bf(x, \omega(x))$. It is obvious that the reduced system is the closed-loop system under the state feedback. Thus it is asymptotically stable with the region of attraction Ω. According to the converse Lyapanouv's theorem [2], there exists a Lyapanouv function $V(x)$ and a positive definite function $U(x)$, defined for $x \in \Omega$, such that

$$\begin{aligned} &V(x) \to \infty \text{ as } x \to \partial\Omega \\ &\frac{\partial V}{\partial x}[Ax + Bf(x, \omega(x))] \le -U(x), \forall x \in \Omega \end{aligned} \quad (44)$$

and for any $c > 0$, $\Omega_c = \{V(x) \le c\}$ is a compact subset of Ω. Define S as a compact set in the interior of Ω, then $S \subset \Omega_c \subset \Omega$.

In the next step, to obtain the boundary-layer model, the change of time variable $\tau = t/\epsilon$ is employed. Then, by setting $\epsilon = 0$, the model is derived as $\frac{d\eta}{d\tau} = A_1\eta$. Since all the eigenvalues of A_1 have negative real parts, there exists a Lyapunov function $W(\eta) = \eta^T P_1 \eta$ such that $\frac{\partial W}{\partial \eta} A_1\eta = -\|\eta\|^2$; where the matrix P_1 satisfies $A_1{}^T P_1 + P_1 A_1 = -I$.

Let $\Sigma = \{W(\eta) \le \rho^2\}$ and $\Pi = \Omega_c \times \Sigma$. The proof of (i) is divided in two steps. In the first step, it is shown that there exist $\epsilon_1 > 0$ such that by selecting $0 < \epsilon \le \epsilon_1$, the set Π is a positively invariant set. On the other hand, it is not possible to guarantee that $\eta(0) \in \Sigma$. To address this challenge, in the second step, it is shown that there exists $\epsilon_2 > 0$ such that for $0 < \epsilon \le \epsilon_2$ and $(x(0), \hat{x}_o(0)) \in \mathcal{S} \times \mathcal{Q}$, the trajectory $(x(t), \eta(t))$ enters Π in finite time.

For the first step, we can use (38) and consider the derivative of V as follows

$$\dot{V} = \frac{\partial V}{\partial x}[Ax + Bf(x, \omega(x - D\eta))] \quad (45)$$

Since f and ω are locally Lipschitz functions, we can consider the following inequality for small enough ϵ

$$\|f(x, \omega(x - D\eta)) - f(x, \omega(x))\| \le k_1\|\eta\|, \forall (x, \eta) \in \Pi \quad (46)$$

where k_1 is a Lipschitz constant. Moreover, over Ω_c, we have $\|\frac{\partial V}{\partial x}\| \le k_2$, and in turn, by adding and subtracting $\frac{\partial V}{\partial x} Bf(x, \omega(x))$ to the right hand side of (45) and using (44) and (46), one has

$$\dot{V} \le -U(x) + k_1 k_2 \|\eta\| \quad (47)$$

In the set Π we have $W(\eta) \le \rho^2$, and consequently $\|\eta\| \le \rho/\sqrt{\lambda_{\min}(P_1)}$. Therefore, one can rewrite (47) as $\dot{V} \le -U(x) + k_1 k_2 \frac{\rho}{\sqrt{\lambda_{\min}(P_1)}}$. By selecting $\rho = \beta/(k_1 k_2 \frac{1}{\sqrt{\lambda_{\min}(P_1)}})$ where $\beta = \min_{x \in \partial\Omega_c} U(x)$, it can be shown that $\dot{V} \le 0$.

For $W(\eta)$, by using (43) we have

$$\begin{aligned} \dot{W} = &-\frac{1}{\epsilon}\eta^T\eta - 2\epsilon^{2(n-1)}\eta^T P_1 M_1 P M_1^T C^T C\eta \\ &+ 2\eta^T P_1 B\Delta(x, D(\epsilon)\eta) \end{aligned}$$

Now $\|P(t)\| \le \gamma$, $\|M_1(t)\| \le k\|M_1(0)\|$, and the fact that $\|\Delta(x, D(\epsilon)\eta)\| \le L_1\|\eta\|$ for $(x, \eta) \in \Pi$ and small enough ϵ can be employed to get

$$\dot{W} \le (-\frac{1}{\epsilon} + \epsilon^{2(n-1)} k_3 + k_4)\|\eta\|^2$$

where $k_3 = 2k^2\gamma\|P_1\|\|M_1(0)\|^2$ and $k_4 = 2L_1\|P_1\|$. Consequently, there exists $\epsilon_1 > 0$ small enough such that for any $0 < \epsilon \le \epsilon_1$ one has $\epsilon^{2(n-1)}k_3 + k_4 \le \frac{1}{2\epsilon}$, and in turn, we get

$$\dot{W} \le -\frac{1}{2\epsilon}\|\eta\|^2 < 0 \quad (48)$$

Therefore, for all $(x, \eta) \in \Pi$, we have $\dot{V} \le 0$ and $\dot{W} < 0$, i.e., Π is a positively invariant set.

Now, in the second step, it is considered that $(x(0), \hat{x}_o(0)) \in \mathcal{S} \times \mathcal{Q}$. Since $f(x, \omega(x - D\eta))$ is locally Lipschitz and $\omega(x - D\eta)$ is a globally bounded function of $x - D\eta$, for all $x \in \Omega_c$ we have

$$\|Ax + Bf(x, \omega((x - D\eta)))\| \le k_5 \quad (49)$$

where k_5 is non-negative and independent of ϵ. Therefore, by using (38), (49), and the fact that $x(0)$ is in the interior of Ω_c, we have $\|\int_0^t \dot{x}(\tau)d\tau\| \le \int_0^t \|\dot{x}(\tau)\|d\tau \le k_5 t$, and in turn, it can be seen that for $x(t) \in \Omega_c$, it is valid to say

$$\|x(t) - x(0)\| \le k_5 t \quad (50)$$

This means that there exists $T_0 > 0$, independent of ϵ, such that $x(t)$ is in the interior of Ω_c for all $t \in [0, T_0]$. During this time interval, by choosing $0 < \epsilon \leq \epsilon_1$, the equation (48) is satisfied. Thus, it can be shown that $\dot{W} \leq -\frac{1}{2\epsilon\lambda_{\max}(P_1)} W$; and consequently, we have

$$W(t) \leq e^{-\frac{1}{2\epsilon\lambda_{\max}(P_1)} t} W(0) \tag{51}$$

Now by defining $T(\epsilon) = 2\epsilon\lambda_{\max}(P_1)\ln(\frac{W(0)}{\rho^2})$, we see that there exists $\epsilon_2 > 0$ such that the inequality $T(\epsilon) \leq \frac{1}{2}T_0$ is satisfied for any $0 < \epsilon \leq \epsilon_2$. It should be noted that such a selection exists since if ϵ tends to zero, $T(\epsilon)$ tends to zero too. Therefore, by selecting $0 < \epsilon \leq \epsilon_2$ it can be guaranteed that $W(T(\epsilon)) \leq \rho^2$. In other words, $\eta(t)$ enters Σ before $x(t)$ leaves Ω_c. Now by considering $\epsilon_1^* = \min\{\epsilon_1, \epsilon_2\}$ and selecting $0 < \epsilon \leq \epsilon_1^*$, it is valid to say that in the time interval $[0, T(\epsilon)]$, $(x(t), \eta(t))$ enters Π and stays in the interior of Π for all $t \geq T(\epsilon)$. Thus, the trajectory $(x(t), \eta(t))$ is bounded for all $t \geq T(\epsilon)$. We see from (50) and (51) that $(x(t), \eta(t))$ is also bounded for $t \in [0, T(\epsilon)]$, and this concludes the proof of (i).

As it was shown, by selecting $0 < \epsilon \leq \epsilon_1^*$, (51) is valid. Therefore, similar to the proof of Theorem 1, one can get $\|\hat{\theta}(t)\| \leq \|\hat{\theta}_0\| + k_6 \frac{\epsilon}{\lambda + \frac{1}{4\lambda_{\max}(P_1)}}$ with $k_6 = k k_7 \gamma \epsilon^{2(n-1)} \|M_1(0)\| \|\eta(0)\|$ and $k_7 = \sqrt{\lambda_{\max}(P_1)/\lambda_{\min}(P_1)}$. In addition, by considering a Lyapunov function candidate $W_i(\eta_i) = \eta_i^T P_1 \eta_i$, it can be obtained that $\dot{W}_i(\eta_i) < 0$ for $\|\eta_i\| > 2\epsilon L_1 k_7 \|P_1\| \|\eta(0)\|$. Hence, $\hat{\theta}$ is bounded and η_is are uniformly ultimately bounded. On the other hand, since the boundedness of $x(t)$ is guaranteed, $\hat{x}_i$s are also uniformly ultimately bounded.

The proof of (ii)-(iv) is similar to Theorem 2 and can be found in [2]. □

Remark 6: The theorem states that the separation principle proved for nonlinear systems and a single HGO is still valid for the proposed MHGO. As a result, a globally bounded state feedback controller and the proposed MHGO can be employed to stabilize nonlinear systems. On the other hand, as shown before, MHGO can provide better state estimations than a single HGO. This means that MHGO can provide more accurate state estimations for the controller which can result in a more preferable performance.

IV. Simulation Results

In this section, simulations are performed on two practical systems, the Van der Pol oscillator and a robotics system. The first example discusses the state estimation problem, and in the second example, the output feedback control problem is addressed.

A. Example 1

This simulation considers the state estimation problem of the Van der Pol oscillator which is $\ddot{x} = -\alpha^2 x + \beta(1 - x^2)\dot{x}$, $y = x$ with $\alpha = 1$, $\beta = 0.5$, and the initial conditions $x(0) = 3$ and $\dot{x}(0) = 2$ [2]. In order to represent the system in the form of (1), one can choose $x_{(1)} = x$, $x_{(2)} = \dot{x}$, and $f(x) = -\alpha^2 x_{(1)} + \beta(1 - x_{(1)}^2)x_{(2)}$. To estimate the system states, a saturated version of $f(x)$ is considered as

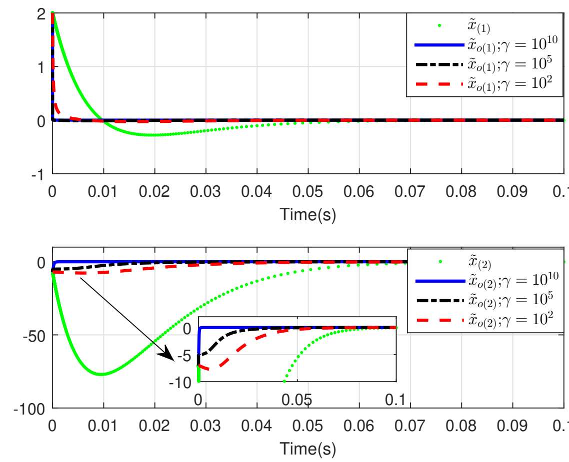

Fig. 1: State estimation error of Van der Pol oscillator using MHGO and a single HGO.

$f_0(\hat{x}_o) = 200\tanh(f(\hat{x}_o)/200)$. Also, since $N \geq n + 1$ observers are required, $N = 3$ observers with the initial conditions of $\hat{x}_1(0) = \begin{bmatrix}+5 & +5\end{bmatrix}^T, \hat{x}_2(0) = \begin{bmatrix}-5 & +5\end{bmatrix}^T, \hat{x}_3(0) = \begin{bmatrix}+5 & -5\end{bmatrix}^T$, and the design parameters $\kappa_1 = 2$, $\kappa_2 = 1$, and $\epsilon = 0.01$ are utilized. Thus $x(0)$ is in the convex hull $\mathcal{K}$ of $\hat{x}_i(0)$s, and $M(0)M(0)^T$ has full rank. Moreover, the initial values for estimating θ^* are chosen as $\hat{\theta}(0) = 0$ and $P(0) = \gamma I$. As it was shown earlier, by choosing a large enough γ, MHGO is able to provide better estimations. To demonstrate that, $\gamma = 10^2$, $\gamma = 10^5$, and $\gamma = 10^{10}$ are employed for simulation. To investigate the performance of MHGO, we compare its results to a single HGO. To provide a reasonable comparison, f_0 and the design parameters of the HGO are selected the same as the MHGO, i.e., κ_is and ϵ. Since different initial conditions result in different peaking values, we set the initial condition of HGO equal to the MHGO as $\hat{x}(0) = \sum_{i=1}^{3} \hat{\alpha}_i(0)\hat{x}_i(0) = \begin{bmatrix}+5 & -5\end{bmatrix}^T$. The obtained estimation errors using the single HGO and MHGO are presented in Fig. 1. From this figure, we see that by choosing a large enough γ, the proposed methodology yields a faster convergence rate and smaller peaks in transient response. To demonstrate the performance of MHGO in estimating θ^*, the evolution of $\hat{\theta}$ needs to be compared to θ^*. In this regard, one can calculate θ^* by assuming $x(0)$ is known and using (8) at $t = 0$. The result of MHGO parameter estimation is depicted in Fig. 2, and it can be seen than $\hat{\theta}$ converges to a small vicinity of θ^* when γ is selected large enough, which is in accordance with Lemma 3.

For investigating the robustness of MHGO to measurement noise, we assume that y is obtained from (33), where the high-frequency measurement noise is $\nu = 10^{-2}\sin(10^3 t)$ [20]. The observation errors of MHGO and a single HGO are depicted in Fig. 3. As it was shown in Lemma 4, since there exists noise in the output measurement, observation error converges to an ultimate bound dependent on the noise, which is the same for both MHGO and a single HGO. In addition, even though choosing a large γ reduces the effect of initial observation error, it can increase the effect of noise on transient response. This trade-off can be easily seen in Fig. 3 where choosing

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2020.2966111

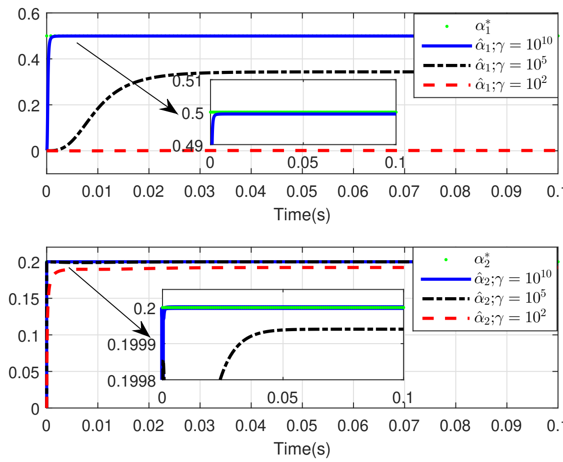

Fig. 2: Performance of MHGO parameter estimation for Van der Pol oscillator.

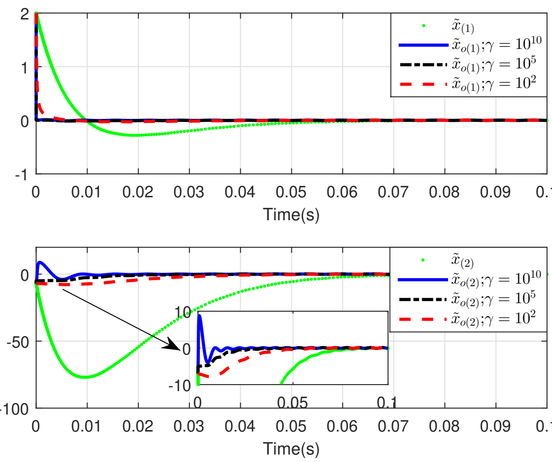

Fig. 3: State estimation error of Van der Pol oscillator using MHGO and a single HGO in the presence of noise.

$\gamma = 10^5$ made the obtained estimation more preferable in comparison to a larger value, i.e., $\gamma = 10^{10}$. Finally, we present a simulation to illustrate the performance of MHGO with resetting when abrupt output disturbances occur. The disturbance ν is considered as follows

$$\nu(t) = \begin{cases} 1 & \text{if } 0.1 \leq t \leq 0.2 \\ 0 & \text{otherwise} \end{cases}$$

For determination of resetting time instants, the monitoring signal ϑ with $\delta = 1$ and $\bar{\vartheta} = 10^{-3}$ is used, and Fig.4 shows the simulation results. It can be seen that when the MHGO is reset (based on the monitoring signal), the estimation error $\tilde{x}_o(t_k^-)$ is an initial estimation error for $t \geq t_k^-$, and the transient response is improved by choosing a large γ.

B. Example 2

This section evaluates the performance of output feedback controllers using MHGO. In this regard, a single-link flexible

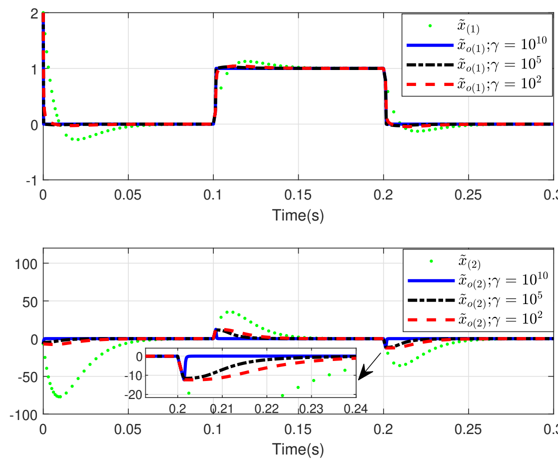

Fig. 4: State estimation error of Van der Pol oscillator using MHGO with resetting and a single HGO in the presence of abrupt disturbance.

joint manipulator with the following dynamic equation is considered as a case study [26],

$$\begin{aligned} \dot{x}_{(1)} &= x_{(2)} \\ \dot{x}_{(2)} &= -\frac{MgL}{I_i}\sin x_{(1)} - \frac{k}{I_i}(x_{(1)} - x_{(3)}) \\ \dot{x}_{(3)} &= x_{(4)} \\ \dot{x}_{(4)} &= \frac{k}{J}(x_{(1)} - x_{(3)}) + \frac{1}{J}u \\ y &= x_{(1)} \end{aligned} \tag{52}$$

where M =2kg is the Link mass, k =100N/m is the Joint elastic constant, L =1m is the distance between rotation axis and the link center of mass, g =9.8m/s^2 is the gravitational acceleration, I_i =0.5kg.m^2 is the Link inertia moment, and J =0.5kg.m^2 is the Rotor inertia moment. The system can be transformed into the canonical form $\dot{z} = Az + Bf(z,u), y = Cz$ [2], where $z_{(1)} = x_{(1)}$, $f(z,u) = \phi(z) + ku/(I_iJ)$, and

$$\begin{aligned} \phi(z) = &\frac{MgL}{I_i}\sin z_{(1)}\left(z_{(2)}^2 + \frac{MgL}{I_i}\cos z_{(1)} + \frac{k}{I_i}\right) \\ &- \left(z_{(3)} + \frac{MgL}{I_i}\sin z_{(1)}\right)\left(\frac{k}{I_i} + \frac{k}{J} + \frac{MgL}{J}\cos z_{(1)}\right). \end{aligned}$$

The initial condition is $z(0) = \begin{bmatrix} -0.1 & 0 & 0.2 & 0 \end{bmatrix}^T$. Assuming full state measurement, the following controller stabilizes the system and forces its output to track the desired trajectory

$$\begin{aligned} u(z) = 100\tanh(\frac{I_iJ}{100k}(&-\phi(z) - 10z_{(1)} - 19z_{(2)} \\ &- 13z_{(3)} - 5z_{(4)} + 10\sin t)) \end{aligned} \tag{53}$$

To estimate the state variables, we consider five HGOs ($N = 5$) with $f_0(\hat{z}_o) = 200\tanh(\frac{1}{200}(\phi(\hat{z}_o) + \frac{k}{I_iJ}u(\hat{z}_o)))$ and the initial conditions $\hat{z}_1(0) = [-1,-1,+1,-1]^T$, $\hat{z}_2(0) = [-1,+1,-1,+1]^T$, $\hat{z}_3(0) = [+1,+1,+1,-1]^T$, $\hat{z}_4(0) = [+1,-1,-1,+1]^T$, and $\hat{z}_5(0) = [+1,-1,+1,+1]^T$. Therefore, $z(0)$ is in the convex hull $\mathcal{K}$ of $\hat{z}_i(0)$s , and $M(0)M(0)^T$ has full rank. Furthermore, the design parameters of the HGOs are selected as $\kappa_1 = 4$, $\kappa_2 = 6$, $\kappa_3 = 4$, $\kappa_4 = 1$,

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TAC.2020.2966111

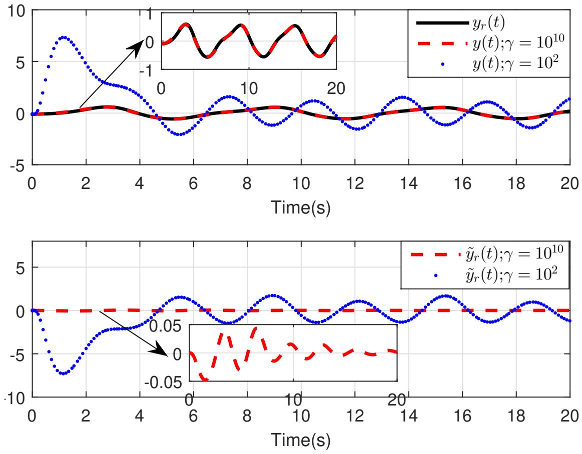

Fig. 5: Output and output error of the manipulator under MHGO-based controller.

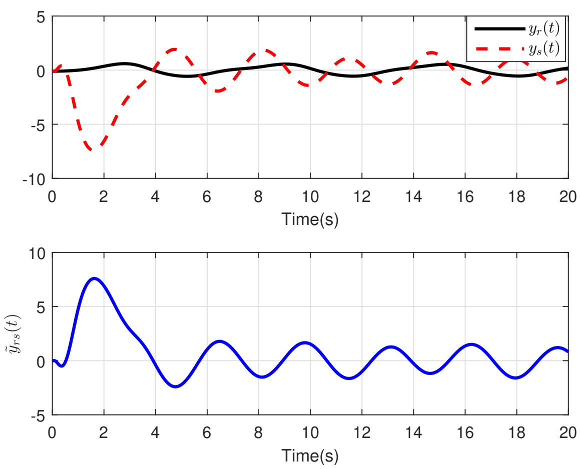

Fig. 6: Output and output error of the manipulator under single HGO-based controller.

and $\epsilon = 0.01$. For obtaining an estimation of θ^*, the initial values of $\hat{\theta}(0) = 0$ and $P(0) = \gamma I$ are employed.

The plant output for $\gamma = 10^2$ and $\gamma = 10^{10}$ $(y(t))$, the plant output under the state feedback controller $(y_r(t))$, and the error between $y_r(t)$ and $y(t)$ $(\tilde{y}_r(t) = y_r(t) - y(t))$ are illustrated in Fig. 5. We see that when γ is large, the output feedback controller $u(\hat{z}_o)$ recovers the performance of state feedback controller. That is because by choosing a large γ, MHGO provides more accurate state estimations for the controller.

As a comparison, the plant is also controlled using a single HGO with the same f_0, κ_is, ϵ, and initial condition, i.e. $\hat{z}(0) = \begin{bmatrix} +1 & -1 & +1 & +1 \end{bmatrix}^T$. The simulation result $(y_s(t))$ along with the error between $y_r(t)$ and $y_s(t)$ $(\tilde{y}_{rs}(t) = y_r(t) - y_s(t))$ are depicted in Fig. 6. By comparing Fig. 5 to Fig. 6, it can be clearly seen that the proposed MHGO-based control strategy can reconstruct the behaviour of full state-based control scheme more rapidly and accurately.

V. Conclusion

This paper deals with state estimation and control of a class of nonaffine nonlinear systems. To address the peaking phenomenon of HGOs, MHGO was presented that considers state estimation as a convex combination of multiple HGOs. It was shown that MHGO is stable and provides an accurate state estimation with smaller peaks; also, its robustness to measurement noise was investigated. Furthermore, a resetting scheme was presented to attenuate peaking from sudden output disturbances. The output feedback control problem was also considered and it was shown that the separation principle is valid for MHGO. As a result, one can employ MHGO along with state feedback controller to stabilize the nonlinear system.

Acknowledgment

The authors would like to thank Hamed Rezaee for helpful discussions on this research.

References

[1] H. K. Khalil and L. Praly, "High-gain observers in nonlinear feedback control," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 6, pp. 993–1015, 2014.

[2] H. K. Khalil, *Nonlinear Systems*, ser. Pearson Education. Prentice Hall New Jersey, 2002.

[3] A. A. Prasov and H. K. Khalil, "A nonlinear high-gain observer for systems with measurement noise in a feedback control framework," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 569–580, 2013.

[4] J. Lei and H. K. Khalil, "Feedback linearization for nonlinear systems with time-varying input and output delays by using high-gain predictors," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2262–2268, 2016.

[5] L. Torres, C. Verde, G. Besançon, and O. González, "High-gain observers for leak location in subterranean pipelines of liquefied petroleum gas," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 6, pp. 1127–1141, 2014.

[6] A. Teel and L. Praly, "Global stabilizability and observability imply semi-global stabilizability by output feedback," *Systems & Control Letters*, vol. 22, no. 5, pp. 313–325, 1994.

[7] ——, "Tools for semiglobal stabilization by partial state and output feedback," *SIAM Journal on Control and Optimization*, vol. 33, no. 5, pp. 1443–1488, 1995.

[8] A. N. Atassi and H. K. Khalil, "A separation principle for the stabilization of a class of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 44, no. 9, pp. 1672–1687, 1999.

[9] F. Esfandiari and H. K. Khalil, "Output feedback stabilization of fully linearizable systems," *International Journal of control*, vol. 56, no. 5, pp. 1007–1037, 1992.

[10] M. S. Chong, D. Nešić, R. Postoyan, and L. Kuhlmann, "Parameter and state estimation of nonlinear systems using a multi-observer under the supervisory framework," *IEEE Transactions on Automatic Control*, vol. 60, no. 9, pp. 2336–2349, 2015.

[11] R. Postoyan, M. H. Hamid, and J. Daafouz, "A multi-observer approach for the state estimation of nonlinear systems," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 1793–1798.

[12] B. D. O. Anderson and A. Dehghani, "Challenges of adaptive control–past, permanent and future," *Annual reviews in control*, vol. 32, no. 2, pp. 123–135, 2008.

[13] K. S. Narendra and J. Balakrishnan, "Adaptive control using multiple models," *IEEE Transactions on Automatic Control*, vol. 42, no. 2, pp. 171–187, 1997.

[14] Z. Han and K. S. Narendra, "New concepts in adaptive control using multiple models," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 78–89, 2012.

[15] K. S. Narendra, Y. Wang, and W. Chen, "Stability, robustness, and performance issues in second level adaptation," in *American Control Conference (ACC)*. IEEE, 2014, pp. 2377–2382.

[16] K. S. Narendra and K. Esfandiari, "Adaptive control of linear periodic systems using multiple models," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 589–594.

[17] V. K. Pandey, I. Kar, and C. Mahanta, "Control of twin-rotor MIMO system using multiple models with second level adaptation," *IFAC-PapersOnLine*, vol. 49, no. 1, pp. 676–681, 2016.

[18] M. Shakarami, K. Esfandiari, A. A. Suratgar, and H. A. Talebi, "On the peaking attenuation and transient response improvement of high-gain observers," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 577–582.
[19] J. Bernat and S. Stepien, "Multi-modelling as new estimation schema for high-gain observers," *International Journal of Control*, vol. 88, no. 6, pp. 1209–1222, 2015.
[20] D. Astolfi and L. Marconi, "A high-gain nonlinear observer with limited gain power," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3059–3064, 2015.
[21] I. J. Bakelman, *Convex analysis and nonlinear geometric elliptic equations*. Springer Science & Business Media, 2012.
[22] V. Andrieu, C. Prieur, S. Tarbouriech, and L. Zaccarian, "A hybrid scheme for reducing peaking in high-gain observers for a class of nonlinear systems," *Automatica*, vol. 72, pp. 138–146, 2016.
[23] D. Astolfi, L. Marconi, and A. Teel, "Low-power peaking-free high-gain observers for nonlinear systems," in *European Control Conference (ECC)*. IEEE, 2016, pp. 1424–1429.
[24] H. K. Khalil, "High-gain observers in nonlinear feedback control," in *International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2008, pp. xlvii–lvii.
[25] L. Vu and D. Liberzon, "Supervisory control of uncertain linear time-varying systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 1, pp. 27–42, 2011.
[26] M. Vidyasagar, *Nonlinear systems analysis*. SIAM, 2002.
[27] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *Siam Review*, vol. 23, no. 1, pp. 53–60, 1981.
[28] W. Cheney, *Analysis for applied mathematics*. Springer Science & Business Media, 2013, vol. 208.

APPENDIX

Proof of Lemma 3. (i) Since Assumption 1 holds, (11) can be used to obtain $M_1(t)\theta^* + \eta_N(\theta^*, t) = 0$. Thus, from (19), we have

$$\eta = M_1\tilde{\theta} + \sigma(\hat{\theta}) \tag{54}$$

where $\tilde{\theta} = \hat{\theta} - \theta^*$ and $\sigma(\hat{\theta}) = \eta_N(\hat{\theta}) - \eta_N(\theta^*)$. To prove (i), we need to find an upper bound for the norm of η. Therefore, it is required to find $\tilde{\theta}$ and in this regard, (20) and (54) are utilized to obtain

$$\dot{\tilde{\theta}} = -\epsilon^{2(n-1)} P M_1^T C^T C(M_1\tilde{\theta} + \sigma(\hat{\theta})) \tag{55}$$

Now by considering $\frac{d(P^{-1}\tilde{\theta})}{dt} = -P^{-1}\dot{P}P^{-1}\tilde{\theta} + P^{-1}\dot{\tilde{\theta}}$, (17), and (55), we get $\frac{d(P^{-1}\tilde{\theta})}{dt} = -\epsilon^{2(n-1)} M_1^T C^T C\sigma(\hat{\theta})$. By taking the integral of this equation and premultiplying by $P(t)$, $\tilde{\theta}(t)$ can be obtained as follows

$$\begin{aligned}\tilde{\theta}(t) &= P(t)P(0)^{-1}\tilde{\theta}_0 \\ &\quad - \epsilon^{2(n-1)} P(t) \int_0^t M_1(\tau)^T C^T C\sigma(\hat{\theta}(\tau), \tau) d\tau\end{aligned} \tag{56}$$

In the preceding equation, it is required to obtain $P(t)$. Toward this end, $\frac{dP^{-1}}{dt} = -P^{-1}\dot{P}P^{-1}$ and (17) can be utilized to get $\frac{dP^{-1}}{dt} = \epsilon^{2(n-1)} M_1^T C^T C M_1$. Taking the integral of this equation and employing $M_1(t) = e^{\frac{1}{\epsilon}A_1 t} M_1(0)$ results in

$$P(t) = \left[P(0)^{-1} + \epsilon^{2(n-1)} M_1(0)^T \Gamma_o(t) M_1(0)\right]^{-1}$$

where $\Gamma_o(t) = \int_0^t e^{\frac{1}{\epsilon}A_1^T\tau} C^T C e^{\frac{1}{\epsilon}A_1\tau} d\tau$ is the observability Gramian corresponding to the observable pair (A_1, C). Hence $\Gamma_o(t)$ is always positive definite. Now, by considering (56), it can be seen that one needs to calculate $P(t)P(0)^{-1}$ and $\epsilon^{2(n-1)}P(t)$. Toward this end, the preceding equation and $P(0) = \gamma I$ is used to obtain $P(t)P(0)^{-1} = \left[I + \gamma\epsilon^{2(n-1)} M_1(0)^T \Gamma_o(t) M_1(0)\right]^{-1}$. By employing the matrix inversion lemma [27], we have

$$\begin{aligned}P(t)P(0)^{-1} &= I - M_1(0)^T \\ &\quad \times \big(\frac{1}{\gamma\epsilon^{2(n-1)}}\Gamma_o(t)^{-1} + M_1(0)M_1(0)^T\big)^{-1} M_1(0)\end{aligned} \tag{57}$$

Given Assumption 2, $M_1(0)M_1(0)^T$ is a full rank matrix. Thus, we can get

$$\begin{aligned}\big(\frac{1}{\gamma\epsilon^{2(n-1)}}\Gamma_o(t)^{-1} + M_1(0)M_1(0)^T\big)^{-1} &= \big(M_1(0)M_1(0)^T\big)^{-1} \\ &\quad \times \big(I + \frac{1}{\gamma\epsilon^{2(n-1)}}\Gamma_o(t)^{-1}(M_1(0)M_1(0)^T)^{-1}\big)^{-1}\end{aligned} \tag{58}$$

Using the Neumann series [28], we can rewrite the above equation and obtain the effect of $\gamma\epsilon^{2(n-1)}$. Toward this end, one has to show that there exists $\gamma^* > 0$ such that for every $\gamma > \gamma^*$, the following equation is satisfied

$$\frac{1}{\gamma\epsilon^{2(n-1)}} \|\Gamma_o(t)^{-1}(M_1(0)M_1(0)^T)^{-1}\| < 1 \tag{59}$$

In this regard, let $\gamma = \xi/\epsilon^{2(n-1)}$ with the positive constant ξ and $\|(M_1(0)M_1(0)^T)^{-1}\| = k_4$. Moreover, since $\Gamma_o(t)$ is symmetric and positive definite, there exists a positive constant k_5 such that $0 < k_5 \leq \lambda_{\min}(\Gamma_o(t))$; hence $k_5 I \leq \Gamma_o(t)$ which results in $\Gamma_o(t)^{-1} \leq \frac{1}{k_5}I$. In addition, since $\Gamma_o(t)$ is a symmetric matrix, similar to $P(t)$ in the proof of Theorem 1, it can be shown that $\|\Gamma_o(t)^{-1}\| \leq \frac{1}{k_5}$; consequently, we have

$$\frac{1}{\gamma\epsilon^{2(n-1)}} \|\Gamma_o(t)^{-1}(M_1(0)M_1(0)^T)^{-1}\| \leq \frac{1}{\xi}\frac{k_4}{k_5}$$

As a result, by defining $\xi^* := \frac{k_4}{k_5}$ and choosing $\xi > \xi^*$, condition (59) will be satisfied. Now the Neumann series can be employed to obtain

$$\begin{aligned}\big(I + \frac{1}{\gamma\epsilon^{2(n-1)}}\Gamma_o(t)^{-1}(M_1(0)M_1(0)^T)^{-1}\big)^{-1} &= \sum_{k=0}^{\infty}(-1)^k \\ &\quad \times (\frac{1}{\xi})^k \big(\Gamma_o(t)^{-1}(M_1(0)M_1(0)^T)^{-1}\big)^k\end{aligned}$$

We use the above equation and (58) to rewrite (57) as

$$\begin{aligned}P(t)P(0)^{-1} &= I - M_1(0)^T(M_1(0)M_1(0)^T)^{-1}M_1(0) \\ &\quad + M_1(0)^T(M_1(0)M_1(0)^T)^{-1}G_1(t)M_1(0)\end{aligned} \tag{60}$$

with $G_1 = \sum_{k=1}^{\infty}(-1)^{k+1}(\frac{1}{\xi})^k(\Gamma_o(t)^{-1}(M_1(0)M_1(0)^T)^{-1})^k$. As it can be seen from (54) and (56), it is required to calculate $M_1(t)P(t)P(0)^{-1}$ and $\epsilon^{2(n-1)}M_1(t)P(t)$. In this regard, we employ (60) and $M_1(t) = e^{\frac{1}{\epsilon}A_1 t}M_1(0)$ to obtain

$$M_1(t)P(t)P(0)^{-1} = e^{\frac{1}{\epsilon}A_1 t} G_1(t) M_1(0) \tag{61}$$

Because $\epsilon^{2(n-1)}P(t) = \gamma\epsilon^{2(n-1)}P(t)P(0)^{-1}$, one can use $\gamma\epsilon^{2(n-1)} = \xi$ and (60) to get

$$\begin{aligned}\epsilon^{2(n-1)}M_1(t)P(t) &= e^{\frac{1}{\epsilon}A_1 t}\Gamma_o(t)^{-1}G_2(t) \\ &\quad \times (M_1(0)M_1(0)^T)^{-1}M_1(0)\end{aligned} \tag{62}$$

with $G_2 = \sum_{k=0}^{\infty}(-1)^k(\frac{1}{\xi})^k((M_1(0)M_1(0)^T)^{-1}\Gamma_o(t)^{-1})^k$. It is worth noting that

$$\begin{aligned}\|G_1(t)\| &\leq \sum_{k=1}^{\infty}(\frac{\xi^*}{\xi})^k = \frac{\xi^*}{\xi-\xi^*}\\ \|G_2(t)\| &\leq \sum_{k=0}^{\infty}(\frac{\xi^*}{\xi})^k = \frac{\xi}{\xi-\xi^*}\end{aligned} \tag{63}$$

Employing (56), (61), and (62), we obtain

$$\begin{aligned}M_1(t)\tilde{\theta}(t) = &e^{\frac{1}{\epsilon}A_1 t}G_1(t)M_1(0)\tilde{\theta}_0\\ &- e^{\frac{1}{\epsilon}A_1 t}\Gamma_o(t)^{-1}G_2(t)\int_0^t e^{\frac{1}{\epsilon}A_1^T\tau}C^TC\sigma(\hat{\theta}(\tau),\tau)d\tau\end{aligned}$$

Then, one can use (28) and (63) to get

$$\begin{aligned}\|M_1(t)\tilde{\theta}(t)\| \leq &k\frac{\xi^*}{\xi-\xi^*}\|M_1(0)\tilde{\theta}_0\|e^{-\frac{\lambda}{\epsilon}t}\\ &+\frac{k^2}{k_5}\frac{\xi}{\xi-\xi^*}\sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}(\tau),\tau)\|e^{-\frac{\lambda}{\epsilon}t}\int_0^t e^{-\frac{\lambda}{\epsilon}\tau}d\tau\end{aligned}$$

Since $e^{-\frac{\lambda}{\epsilon}t} \leq 1$ and $e^{-\frac{\lambda}{\epsilon}t} - e^{-2\frac{\lambda}{\epsilon}t} \leq 1/4$, we can say

$$\begin{aligned}\|M_1(t)\tilde{\theta}(t)\| \leq &k\frac{\xi^*}{\xi-\xi^*}\|M_1(0)\tilde{\theta}_0\|\\ &+\frac{k^2}{k_5}\frac{\epsilon}{4\lambda}\frac{\xi}{\xi-\xi^*}\sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}(\tau),\tau)\|\end{aligned} \tag{64}$$

For analyzing the above equation, we need to obtain the supremum of $\|\sigma(\hat{\theta})\|$. Toward this end, we use Lemma 2 to conclude that $\epsilon\dot{\eta}_N(\theta^*) = A_1\eta_N(\theta^*)$. As a result, one can use the definition of $\sigma(\hat{\theta}) = \eta_N(\hat{\theta}) - \eta_N(\theta^*)$ and get $\dot{\sigma}(\hat{\theta}) = \frac{1}{\epsilon}A_1\sigma(\hat{\theta}) + B\left[f(x,u) - f_0(x - D\eta, u)\right]$. Thus we have

$$\begin{aligned}\sigma(\hat{\theta}(t),t) =&e^{\frac{1}{\epsilon}A_1 t}\sigma(\hat{\theta}(0),0) + \int_{\tau=0}^t e^{\frac{1}{\epsilon}A_1(t-\tau)}B\\ &\times\left[f(x(\tau),u(\tau)) - f_0(x(\tau)-D\eta(\tau),u(\tau))\right]d\tau\end{aligned}$$

Moreover, from the definition of σ it can be seen that $\sigma(\theta(\hat{0}),0) = 0$. Therefore, (24) and (28) are utilized to obtain

$$\begin{aligned}\|\sigma(\hat{\theta}(t),t)\| &\leq kL_1\sup_{0\leq\tau\leq t}\|\eta(\tau)\|\int_{\tau=0}^t e^{-\frac{\lambda}{\epsilon}(t-\tau)}d\tau\\ &\leq kL_1\frac{\epsilon}{\lambda}\sup_{0\leq\tau\leq t}\|\eta(\tau)\|\end{aligned}$$

Since the right hand side of the preceding equation is non-decreasing, we can employ the fact that the supremum of a function is its least upper bound and derive

$$\sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}(\tau),\tau)\| \leq kL_1\frac{\epsilon}{\lambda}\sup_{0\leq\tau\leq t}\|\eta(\tau)\| \tag{65}$$

Therefore, (54), (64), (65), and $\eta(0) = M_1(0)\tilde{\theta}_0$ can be used to get

$$\begin{aligned}\|\eta(t)\| \leq &k\frac{\xi^*}{\xi-\xi^*}\|\eta(0)\|\\ &+(\frac{k^2}{k_5}\frac{\epsilon}{4\lambda}\frac{\xi}{\xi-\xi^*}+1)kL_1\frac{\epsilon}{\lambda}\sup_{0\leq\tau\leq t}\|\eta(\tau)\|\end{aligned}$$

It can be seen that the right hand side of the preceding equation is nondecreasing, therefore, it is also greater than or equal to $\sup\|\eta(\tau)\|$ for $0\leq\tau\leq t$. Moreover, there exist some $\xi_1^* > \xi^*$ and $\epsilon_1^* > 0$ such that for $\xi > \xi_1^*$ and $0 < \epsilon < \epsilon_1^*$, we have $1 - (\frac{k^2}{k_5}\frac{\epsilon}{4\lambda}\frac{\xi}{\xi-\xi^*}+1)kL_1\frac{\epsilon}{\lambda} > 0$. Thus, one can write

$$\sup_{0\leq\tau\leq t}\|\eta(\tau)\| \leq \frac{k\frac{\xi^*}{\xi-\xi^*}\|\eta(0)\|}{1-(\frac{k^2}{k_5}\frac{\epsilon}{4\lambda}\frac{\xi}{\xi-\xi^*}+1)kL_1\frac{\epsilon}{\lambda}} \tag{66}$$

On the other hand, from (14) it is obtained

$$\epsilon^{2(n-1)}\|\eta\|^2 \leq \|\tilde{x}_o\|^2 \leq \|\eta\|^2 \tag{67}$$

Finally, the above equation and (66) result in

$$\sup_{0\leq\tau\leq t}\|\tilde{x}_o(\tau)\| \leq \frac{k\frac{1}{\epsilon^{n-1}}\frac{\xi^*}{\xi-\xi^*}\|\tilde{x}_o(0)\|}{1-(\frac{k^2}{k_5}\frac{\epsilon}{4\lambda}\frac{\xi}{\xi-\xi^*}+1)kL_1\frac{\epsilon}{\lambda}} \tag{68}$$

For comparison, we need to employ a similar approach for a single HGO. For that, by considering (3) and (7), we see that state estimation obtained from (4) and (5) with fixed α_is is equal to the estimation of a single HGO. Therefore, a single HGO with the initial condition of $\hat{x}(0) = \sum_{i=1}^N\hat{\alpha}_i(0)\hat{x}_i(0)$ performs like an MHGO with the same initial condition and fixed $\hat{\alpha}_i$s, i.e., $\hat{x}(t) = \sum_{i=1}^N\hat{\alpha}_i(0)\hat{x}_i(\Lambda,t)$. Hence, the scaled state estimation error of a single HGO, $\eta_s = D(\epsilon)^{-1}\tilde{x}$, is

$$\eta_s(t) = M_1(t)\tilde{\theta}_0 + \sigma(\hat{\theta}_0,t)$$

where $\sigma(\hat{\theta}_0,t) = \eta_N(\hat{\theta}_0,t) - \eta_N(\theta^*,t)$. Note that in the preceding equation $\hat{\theta}$ is fixed, i.e., $\tilde{\theta}(t) = \tilde{\theta}_0$. By using a similar approach to MHGO and $\eta_s(0) = M_1(0)\tilde{\theta}_0$, we get

$$\begin{aligned}\|M_1(t)\tilde{\theta}(0)\| &\leq k\|\eta_s(0)\|\\ \sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}_0,\tau)\| &\leq kL_1\frac{\epsilon}{\lambda}\sup_{0\leq\tau\leq t}\|\eta_s(\tau)\|\end{aligned}$$

Therefore, by choosing ϵ similar to MHGO, one can obtain the following equation for the estimation error of a single HGO

$$\sup_{0\leq\tau\leq t}\|\tilde{x}(\tau)\| \leq \frac{k\frac{1}{\epsilon^{n-1}}\|\tilde{x}(0)\|}{1-kL_1\frac{\epsilon}{\lambda}} \tag{69}$$

By comparing (68) and (69) and using $\tilde{x}_o(0) = \tilde{x}(0)$, it can be seen that by choosing ξ large enough, when ϵ tends to zero, the transient estimation errors obtained from MHGO and a single HGO can peak to $\mathcal{O}(\|\tilde{x}_o(0)\|/(\xi\epsilon^{n-1}))$ and $\mathcal{O}(\|\tilde{x}(0)\|/\epsilon^{n-1})$, respectively, and this completes the proof of (i).

(ii) Since $N = n+1$ and $M_1(0)M_1(0)^T$ has full rank, the matrix $M_1(0) \in \mathcal{R}^{n\times(N-1)}$ is invertible. Therefore, (60) becomes

$$P(t)P(0)^{-1} = M_1(0)^TG(t)M_1(0)^{-T} \tag{70}$$

with $G = \sum_{k=1}^{\infty}(-1)^{k+1}(\frac{1}{\xi})^k\left((M_1(0)M_1(0)^T)^{-1}\Gamma_o(t)^{-1}\right)^k$. Also, using $\epsilon^{2(n-1)}P(t) = \xi P(t)P(0)^{-1}$ and (70), one has

$$\epsilon^{2(n-1)}P(t)M_1(0)^T = \xi M_1(0)^TG(t) \tag{71}$$

Note that similar to (63), we have $\|G(t)\| \leq \xi^*/(\xi-\xi^*)$. Thus, (56), (28), (70), and (71) can be utilized to get

$$\begin{aligned}\|\tilde{\theta}(t)\| \leq &\frac{\xi^*}{\xi-\xi^*}\|\tilde{\theta}_0\|\\ &+k\frac{\epsilon}{\lambda}\|M_1(0)\|\frac{\xi\xi^*}{\xi-\xi^*}\sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}(\tau),\tau)\|\end{aligned} \tag{72}$$

On the other hand, we use (65), (54), and (28) to get

$$\sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}(\tau),\tau)\| \leq k^2L_1\frac{\epsilon}{\lambda}\|M_1(0)\|\sup_{0\leq\tau\leq t}\|\tilde{\theta}(\tau)\| + kL_1\frac{\epsilon}{\lambda}\sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}(\tau),(\tau))\|$$

Hence, if $\epsilon^* := \lambda/(KL_1)$ and $\epsilon < \epsilon^*$, we have

$$\sup_{0\leq\tau\leq t}\|\sigma(\hat{\theta}(\tau),\tau)\| \leq k^2L_1\|M_1(0)\|\frac{\epsilon}{\lambda - kL_1\epsilon}\sup_{0\leq\tau\leq t}\|\tilde{\theta}(\tau)\|$$

Using the preceding equation, (72) can be obtained as follows

$$\begin{aligned}\|\tilde{\theta}(t)\| &\leq \frac{\xi^*}{\xi-\xi^*}\|\tilde{\theta}_0\| \\ &+ k^3L_1\|M_1(0)\|^2\frac{\xi\xi^*}{\xi-\xi^*}\frac{\epsilon^2}{\lambda(\lambda-kL_1\epsilon)}\sup_{0\leq\tau\leq t}\|\tilde{\theta}(\tau)\|\end{aligned} \tag{73}$$

There exist $\xi_2^* > \xi^*$ and $\epsilon_2^* < \epsilon^*$ such that by selecting $\xi > \xi_2^*$ and $0 < \epsilon < \epsilon_2^*$, one has $1 - k^3L_1\|M_1(0)\|^2\frac{\xi\xi^*}{\xi-\xi^*}\frac{\epsilon^2}{\lambda(\lambda-kL_1\epsilon)} > 0$. Also, since the right hand side of (73) is nondecreasing, it is also greater than or equal to $\sup\|\tilde{\theta}(\tau)\|$ for $0 \leq \tau \leq t$; thus,

$$\sup_{0\leq\tau\leq t}\|\tilde{\theta}(\tau)\| \leq \frac{\frac{\xi^*}{\xi-\xi^*}\|\tilde{\theta}_0\|}{1 - k^3L_1\|M_1(0)\|^2\frac{\xi\xi^*}{\xi-\xi^*}\frac{\epsilon^2}{\lambda(\lambda-kL_1\epsilon)}}$$

As a result, $\|\tilde{\theta}(t)\|$ is less than or equal to $\mathcal{O}(\|\tilde{\theta}_0\|/\xi)$ when ξ and ϵ are selected large and small enough, respectively. □

Proof of Lemma 4. Considering the effect of noise as (33), we see that in (15)-(18), only (16) and (18) need modification as

$$\dot{\hat{\theta}} = -\epsilon^{n-1}PM_1^TC^T(\epsilon^{n-1}C\eta + \nu) \tag{74}$$

$$\begin{aligned}\dot{\eta}_i(\hat{\theta}) &= \frac{1}{\epsilon}A_1\eta_i(\hat{\theta}) \\ &+ B[f(x,u) - f_0(x - D(\epsilon)\eta, u)] - \frac{1}{\epsilon^n}H_1\nu\end{aligned} \tag{75}$$

where $H_1 = \begin{bmatrix}\kappa_1 & \cdots & \kappa_n\end{bmatrix}^T$. Note that the effect of ν on η_is is similar; therefore, since the ith column of M_1 is $\eta_i(\hat{\theta}) - \eta_N(\hat{\theta})$, M_1 does not depend on ν and (21) is valid. One can take the derivative of (15) and use (21), (74), and (75) to obtain

$$\begin{aligned}\dot{\eta} = \frac{1}{\epsilon}A_1\eta &- \epsilon^{2(n-1)}M_1PM_1^TC^TC\eta \\ &+ B[f(x,u) - f_0(x - D(\epsilon)\eta, u)] \\ &- \epsilon^{n-1}M_1PM_1^TC^T\nu - \frac{1}{\epsilon^n}H_1\nu\end{aligned} \tag{76}$$

We perform the proof using stability analysis of perturbed systems with non-vanishing perturbation [2]. Let the Lyapunov function candidate $V(\eta) = \eta^TP_1\eta$ with $A_1^TP_1 + P_1A_1 = -I$. Then, we employ (76) to get

$$\begin{aligned}\dot{V}(\eta) = -\frac{1}{\epsilon}\eta^T\eta &- 2\epsilon^{2(n-1)}\eta^TP_1M_1PM_1^TC^TC\eta \\ &+ 2\eta^TP_1B[f(x,u) - f_0(x - D(\epsilon)\eta, u)] \\ &- 2\epsilon^{n-1}\eta^TP_1M_1PM_1^TC^T\nu - \frac{2}{\epsilon^n}\eta^TP_1H_1\nu\end{aligned}$$

By considering (17), (21), and (28), we have $\|M_1(t)\| \leq k\|M_1(0)\|e^{-\frac{\lambda}{\epsilon}t}$ and $\|P(t)\| \leq \gamma$. Therefore, (24) and $\|\nu\| \leq \mu$ can be utilized to obtain

$$\begin{aligned}\dot{V}(\eta) \leq &(-\frac{1}{\epsilon} + \rho_1\gamma\epsilon^{2(n-1)} + 2L_1\|P_1\|)\|\eta\|^2 \\ &+ (\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}t} + \frac{\rho_2}{\epsilon^n})\mu\|\eta\|\end{aligned} \tag{77}$$

where $\rho_1 = 2k^2\|P_1\|\|M_1(0)\|^2$ and $\rho_2 = 2\|P_1\|\|H_1\|$. There exists $\epsilon^* > 0$ such that by selecting $0 < \epsilon \leq \epsilon^*$, one has $\rho_1\gamma\epsilon^{2(n-1)} + 2L_1\|P_1\| \leq \frac{1}{2\epsilon}$. Consequently, it is obtained $\dot{V}(\eta) \leq -\frac{1}{2\epsilon}\|\eta\|^2 + (\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}t} + \frac{\rho_2}{\epsilon^n})\mu\|\eta\|$, and by employing (29), we have

$$\dot{V}(\eta) \leq -\frac{V(\eta)}{2\epsilon\lambda_{\max}(P_1)} + (\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}t} + \frac{\rho_2}{\epsilon^n})\frac{\mu\sqrt{V(\eta)}}{\sqrt{\lambda_{\min}(P_1)}}$$

To proceed with the analysis, it is needed to convert the preceding equation into a linear differential inequality and employ the Comparison Lemma [2]. For that, let $W(t) = \sqrt{V(t)}$; hence, for $V(t) \neq 0$, $\dot{W} = \dot{V}/(2\sqrt{V})$ can be used to get

$$\dot{W} \leq -\frac{W}{4\epsilon\lambda_{\max}(P_1)} + (\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}t} + \frac{\rho_2}{\epsilon^n})\frac{\mu}{2\sqrt{\lambda_{\min}(P_1)}} \tag{78}$$

In order to use the comparison lemma, we need to show that the upper right hand derivative $D^+W(t)$ satisfies (78). If $V(t) \neq 0$, $W(t)$ is differentiable, and in turn, $D^+W(t) = \dot{W}(t)$. To show that $D^+W(t)$ satisfies (78) when $V(t) = 0$, using $W(t) = \sqrt{V(t)} = 0$, it is required to show

$$D^+W(t) \leq (\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}t} + \frac{\rho_2}{\epsilon^n})\frac{\mu}{2\sqrt{\lambda_{\min}(P_1)}} \tag{79}$$

In this regard, let the definition of the upper right hand derivative $D^+W(t) = \limsup_{h\to0^+}\frac{W(t+h)-W(t)}{h}$. Note that since $V(t) = 0$, we have $W(t) = 0$ and $\eta(t) = 0$. Therefore, by using (29), one has

$$D^+W(t) \leq \lim\sup_{h\to0^+}\frac{\sqrt{\lambda_{\max}(P_1)}}{h}\|\eta(t+h)\| \tag{80}$$

Also, one can use the Taylor series, $\eta(t) = 0$, and (76) to get

$$\|\eta(t+h)\| \leq h\| - \epsilon^{n-1}M_1PM_1^TC^T\nu - \frac{1}{\epsilon^n}H_1\nu\| + h^2\|\sum_{k=2}^{\infty}\frac{h^{k-2}}{k!}\eta^{(k)}(t)\|$$

Employing the upper bounds of $\|P\|$, $\|M_1\|$, and $\|\nu\|$ together with the definitions of ρ_1 and ρ_2 in (77), we have

$$\|\eta(t+h)\| \leq h(\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}t} + \frac{\rho_2}{\epsilon^n})\frac{\mu}{2\|P_1\|} + h^2\|\sum_{k=2}^{\infty}\frac{h^{k-2}}{k!}\eta^{(k)}(t)\|$$

By using the preceding equation and (80), it can be obtained

$$D^+W(t) \leq \sqrt{\lambda_{\max}(P_1)}(\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}t} + \frac{\rho_2}{\epsilon^n})\frac{\mu}{2\|P_1\|}$$

Since $\|P_1\| = \lambda_{\max}(P_1)$ and $\sqrt{\lambda_{\min}(P_1)} \leq \sqrt{\lambda_{\max}(P_1)}$, it can be concluded from the previous equation that (79) is valid.

Therefore, $D^+W(t)$ satisfies (78), and using the comparison lemma, the following equation can be obtained

$$W(t) \leq e^{-\frac{1}{4\epsilon\lambda_{\max}(P_1)}t}W(0) + \int_0^t e^{-\frac{t-\tau}{4\epsilon\lambda_{\max}(P_1)}}(\rho_1\gamma\epsilon^{n-1}e^{-2\frac{\lambda}{\epsilon}\tau} + \frac{\rho_2}{\epsilon^n})\frac{\mu}{2\sqrt{\lambda_{\min}(P_1)}}d\tau$$

By calculating the integral and using the definition $\lambda = 1/(2\lambda_{\max}(P_1))$, we have

$$W(t) \leq e^{-\frac{\lambda}{2\epsilon}t}W(0) + \Big[\frac{1}{3}\rho_1\gamma\epsilon^n(e^{-\frac{\lambda}{2\epsilon}t} - e^{-2\frac{\lambda}{\epsilon}t}) + \frac{\rho_2}{\epsilon^{n-1}}(1 - e^{-\frac{\lambda}{2\epsilon}t})\Big]\frac{2\lambda_{\max}(P_1)}{\sqrt{\lambda_{\min}(P_1)}}\mu$$

Finally, (34) can be obtained by employing the preceding equation, $\sqrt{\lambda_{\min}(P_1)}\|\eta\| \leq W \leq \sqrt{\lambda_{\max}(P_1)}\|\eta\|$, (67), and the following definitions.

$$k_1 = \frac{\sqrt{\lambda_{\max}(P_1)}}{\sqrt{\lambda_{\min}(P_1)}}, k_2 = \rho_1\frac{2\lambda_{\max}(P_1)}{3\lambda_{\min}(P_1)}, k_3 = \rho_2\frac{2\lambda_{\max}(P_1)}{\lambda_{\min}(P_1)} \tag{81}$$

□

Proof of Lemma 5. First, we consider the proposed scheme between two resetting time instants t_{k-1} and t_k. During this time interval $[t_{k-1}, t_k)$, the conditions of Lemma 4 are satisfied, and (76) is valid. Analogous to the proof of Lemma 4, one can obtain

$$\begin{aligned} W(t) \leq\ & e^{-\frac{\lambda}{2\epsilon}(t-t_{k-1})}W(t_{k-1}) \\ & + \Big[\frac{1}{3}\rho_1\gamma\epsilon^n(e^{-\frac{\lambda}{2\epsilon}(t-t_{k-1})} - e^{-2\frac{\lambda}{\epsilon}(t-t_{k-1})}) \\ & + \frac{\rho_2}{\epsilon^{n-1}}(1 - e^{-\frac{\lambda}{2\epsilon}(t-t_{k-1})})\Big]\frac{2\lambda_{\max}(P_1)}{\sqrt{\lambda_{\min}(P_1)}}\mu \end{aligned} \tag{82}$$

where $t \in [t_{k-1}, t_k)$. At $t = t_k$, MHGO is re-initialized using (35) such that $\hat{x}_o(t_k) = \hat{x}_o(t_k^-)$. As a result, we have $\eta(t_k) = \eta(t_k^-)$ and $W(t_k) = W(t_k^-)$. By exploiting this equality and (82), it is obtained

$$\begin{aligned} W(t) \leq\ & e^{-\frac{\lambda}{2\epsilon}t}W(0) \\ & + \Big\{\sum_{j=1}^{k-1} e^{-\frac{\lambda}{2\epsilon}(t-t_j)}\Big[\frac{1}{3}\rho_1\gamma\epsilon^n(e^{-\frac{\lambda}{2\epsilon}(t_j-t_{j-1})} - e^{-2\frac{\lambda}{\epsilon}(t_j-t_{j-1})}) \\ & + \frac{\rho_2}{\epsilon^{n-1}}(1 - e^{-\frac{\lambda}{2\epsilon}(t_j-t_{j-1})})\Big] \\ & + \frac{1}{3}\rho_1\gamma\epsilon^n(e^{-\frac{\lambda}{2\epsilon}(t-t_{k-1})} - e^{-2\frac{\lambda}{\epsilon}(t-t_{k-1})}) \\ & + \frac{\rho_2}{\epsilon^{n-1}}(1 - e^{-\frac{\lambda}{2\epsilon}(t-t_{k-1})})\Big\}\frac{2\lambda_{\max}(P_1)}{\sqrt{\lambda_{\min}(P_1)}}\mu \end{aligned}$$

The proof is concluded by employing the definitions of k_1, k_2, and k_3 as (81). □

Mehran Shakarami received the B.Sc. degree in electronics engineering from Shamsipour College and the M.Sc. degree in electrical engineering/control systems from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. He is currently pursuing the Ph.D. degree in systems and control with the University of Groningen, The Netherlands. His current research interests include game theory, dynamical networks, and control theory.

Kasra Esfandiari (S'14) received the B.Sc. degree in electrical engineering/control systems from Shiraz University, Shiraz, Iran, in 2012, and the M.Sc. degree in electrical engineering/control systems from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2014. In 2018, he received the second M.Sc. degree in electrical engineering from Yale University, New Haven, CT, USA, where he is currently pursuing the Ph.D. degree in electrical engineering. His current research interests include adaptive systems, neural networks, nonlinear control, fault diagnosis and recovery, and robotics.

Amir Abolfazl Suratgar (M'09–SM'14) received his B.S. degree in electrical engineering from Isfahan University of technology, Isfahan, Iran, in 1996. He was honored M.S. and Ph.D. degrees in control engineering from Amirkabir University of Technology (Tehran polytechnic), Tehran, Iran, in 1999 and 2002, respectively. Currently, he is an Associate Professor in Amirkabir University of Technology (Tehran polytechnic), Tehran, Iran. The areas of his research interests are optimal control and optimization, computational intelligence, neuroscience, distributed optimization, and machine learning. He has published more than 140 papers in scientific journals and conferences.

Heidar A. Talebi (M'02–SM'08) received the B.Sc. degree in electrical engineering from Ferdowsi University, Mashhad, Iran, in 1988, the M.Sc. degree in electrical engineering from Tarbiat Modarres University, Tehran, Iran, in 1991, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, QC, Canada, in 1997.

He was a Postdoctoral Fellow and a Research Fellow at Concordia University and The University of Western Ontario. In 1999, he joined Amirkabir University of Technology, Tehran, Iran. He is currently a professor in the Department of Electrical Engineering, and his research interests include control, robotics, medical robotics, fault diagnosis and recovery, intelligent systems, adaptive control, nonlinear control, and real-time systems.