INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust. Nonlinear Control 2012; **00**:1–25
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/rnc

Hybrid-type Observer Design Based on a Sufficient Condition for Observability in Switched Nonlinear Systems

Hyungbo Shim[1]*and Aneel Tanwani[2]

[1] *ASRI, Department of Electrical Engineering, Seoul National University, Korea,*
[2] *Team BIPOP, Institut National de Recherche en Informatique et Automatique (INRIA), Rhône-Alpes, France.*

SUMMARY

This paper presents a sufficient condition for observability of continuous-time switched nonlinear systems that also involve state jumps. Without assuming observability of individual modes, the sufficient condition is based on gathering partial information from each mode so that the state is completely recovered after several switchings. Based on the sufficient condition, a hybrid-type observer is designed, which comprises a copy of the actual plant and an error correction scheme at discrete time instants. In order to execute the proposed design, the observable component of the state at each mode needs to be estimated without transients or "peaking" (caused by high-gain observers), and this motivates us to introduce a back-and-forth estimation technique. Under the assumption of persistent switching, analysis shows that the estimate thus generated converges asymptotically to the actual state of the system. Simulation results validate the utility of proposed algorithm.

Received ...

KEY WORDS: Switched nonlinear system; Large-time observability; High-gain observer; Hybrid-type observer; Uniform observability

1. INTRODUCTION

We study observability conditions and observer design for a class of switched systems, that comprise a family of nonlinear control affine dynamical subsystems with state jumps. These subsystems are activated by a switching signal $\sigma : \mathbb{R} \to \mathbb{N}$, which is right-continuous and piecewise constant where the discontinuities occur at discrete time instants $\{t_q\}$, $q \in \mathbb{N}$, called the switching times. Mathematically, the systems under consideration are modeled as:

$$\dot{x}(t) = f_{\sigma(t)}(x(t)) + g_{\sigma(t)}(x(t))u(t), \quad t \neq \{t_q\}, \tag{1a}$$

$$x(t_q) = p_{\sigma(t_q),\sigma(t_q^-)}(x(t_q^-)), \tag{1b}$$

$$y(t) = h_{\sigma(t)}(x(t)), \tag{1c}$$

where $x : \mathbb{R} \mapsto \mathbb{R}^n$ is the state trajectory, $y : \mathbb{R} \mapsto \mathbb{R}^{n_y}$ is the output, the measurable function $u : \mathbb{R} \mapsto \mathbb{R}^{n_u}$ is the input belonging to some input class $\mathcal{U}$ of interest, and the map $p_{\sigma(t_q),\sigma(t_q^-)}(\cdot)$ denotes the jumps in state trajectories at time t_q while switching from mode $\sigma(t_q^-)$ to $\sigma(t_q)$. Such jumps may model impulse effects, state resets, or any other discontinuity in the state trajectory. For the development of results in this paper, we assume that the following hypotheses hold for system (1) throughout the paper.

*Correspondence to: Kwanak PO Box 34, Seoul, Korea. E-mail: hshim@snu.ac.kr

(H1) The switching signal $\sigma(\cdot)$ is assumed to be a function of time only and is considered to be known for designing observers.
(H2) There are a finite number of switching times in any finite time interval, that is, the Zeno phenomenon is not exhibited.
(H3) The solution $x(t)$ remains in a set $\mathcal{X} \subset \mathbb{R}^n$ on the time interval of interest.
(H4) All the vector fields and functions are smooth on $\mathcal{X}$, and there exists a unique solution for every time interval under consideration.

Under the hypotheses (H1)–(H4), the resulting solution is absolutely continuous between any two consecutive switching instants and the possible discontinuities in the state trajectory, appearing at the switching instants, are represented by (1b).

When dealing with observability of nonlinear systems, there are different notions that are involved. The work in [11, 13] talks about observability in local neighborhoods of the state space. Authors in [12, 14] describe the notion of 'large-time' versus 'small-time' observability where the difference lies in the fact whether it is possible to recover the state instantaneously in time or the system becomes observable after certain time. If the system description has exogenous inputs acting on it, then the question arises whether observability holds for all inputs or not [8, 9]; if it does, the system is called uniformly observable.

The concept of observability studied in this paper is a refinement of the 'large-time observability' already considered in the literature (e.g., [12, 14, 24]) and the 'uniform observability' studied in [8, 9]. Switched systems can be thought of as a family of dynamical subsystems, where a switching signal determines the active subsystem at each time instant. It is entirely possible that none of these subsystems is observable in the sense that information about the full state is not immediate in the output signal. But the information available from each mode can be combined in a certain manner so that, under some conditions, it is possible to recover the state vector completely after several switchings have occurred. This explains how the concept of 'large-time' comes into picture when dealing with switched systems and our aim is to derive conditions that make the system large-time observable when the state trajectories are contained in a given set $\mathcal{X}$. Moreover, since our ultimate goal lies in the construction of an observer, the observability for all inputs (i.e., uniform observability) is of concern in order for the observer to be independent of particular inputs. For a formal definition of large-time uniform observability on a given set, see Section 2.

For switched systems, the problems of observability and observer design have been studied primarily for the linear case. Some initial observer results, such as [1, 17], have assumed that each mode in the subsystem is in fact an observable system that admits a state observer; and then analyzed the stability problem of the switched error dynamics using classical tools, such as dwell time, or common Lyapunov function. More relaxed approaches do not assume observability of the individual modes, and the notion of gaining observability by switching has appeared in, e.g., [5, 26, 27]. The sufficient conditions proposed in [5, 26] imply that the full state information is recovered after one or several switchings. Both papers use outputs and their derivatives to recover the state. The work of [27] gives geometric conditions under which there exists at least one switching signal that makes the system observable. However, in spite of being limited to the linear case, it is not clear how the conditions in [5, 26, 27] can lead to feasible observer design. This gap was bridged in our recent work [24], where the idea of a hybrid-type observer is proposed in linear setting for estimating the state trajectories asymptotically.

For observability of switched nonlinear systems, however, there doesn't exist much literature. The authors in [3] address observer design using observability of individual modes. Somewhat related to our approach is the work of [14], where the authors propose some preliminary abstract results for observability. The basic idea is to recover the entire state by collecting the partial information at each mode and transporting it unperturbed to some point in time. As done in [24], this transportation is easily carried out in linear systems because the system state $x(t)$ can always be written as the *sum* of an observable vector $x_o(t)$ and an unobservable vector $x_u(t)$ for the active mode at time t; and the transportation over a certain interval is achieved simply by multiplying the state-transition matrix corresponding to that interval with $x(t) (= x_o(t) + x_u(t))$. Because of the linearity of the state-transition matrix, the transported state is still the sum of two vectors, out of which the unobservable

DOI: 10.1002/rnc

component can be easily annihilated, leaving behind the other part as a linear combination of the observable components of individual modes only, and making it possible to recover the state. This strategy cannot be adopted in nonlinear systems, and instead of transporting x to a future time via the state-transition matrix, we exploit the system structure to construct some nonlinear dynamics, whose integration plays the role of transporting the known partial information. This idea yields a sufficient (but not a necessary) observability condition that renders a particular structure to the switched system, which is amenable for construction of asymptotic observers. A motivating example and further discussion of this concept appear in Section 2. The proposed sufficient condition for observability and an overview of observer construction are provided in Section 3. A geometric characterization of the proposed condition is given in Section 4, along with illustrative examples.

Finally, the design of hybrid-type observer based on the proposed observability condition is studied in Section 5. The proposed observer consists of a copy of plant dynamics that is running synchronously with the plant, and an error correction implemented as a jump in the state of the observer. Before the jump, an estimation algorithm is executed in which several sub-observers are run to process the input and the output data that are recorded in the memory. Unlike the simpler hybrid-type observer in [24], the so-called 'back-and-forth' operation of the observer is newly proposed and specifically required for nonlinear systems as a remedy for transient overshoot that might be caused by the high-gain observers employed in the estimation algorithm. Since the estimation algorithm does not run synchronously to the plant, we also take into account the time consumed by the algorithm, which was not considered in the preliminary version [21] of this paper. Simulation results are also presented in Section 5.

For the observer design, our approach shares the same spirit with [2], and the result of this paper may be regarded as an extension of the linear result in [2], in the sense that, a coordinate-independent condition is derived for observability and nonlinear systems are treated with a new observer design strategy. More discussions and concluding remarks are found in Section 6.

The notation and terminology used in this paper are summarized as follows. For a signal $x(\cdot)$, $x_{[t_1,t_2]}$ means $\{x(t) : t_1 \le t \le t_2\}$. For a vector x, $|x|$ denotes the Euclidean norm of x. $\mathcal{R}(A)$ implies the range space of the columns of matrix A, and $A^\top$ is the transpose of A. We write $[x_1^\top, x_2^\top]^\top$ simply by $\mathrm{col}(x_1, x_2)$. A composite function $\lambda(p(\cdot))$ is denoted by $\lambda \circ p$. With given functions λ_i, $i = 1, \cdots, k$, we denote $\lambda_{\langle k\rangle} := \mathrm{col}(\lambda_1, \ldots, \lambda_k)$. Now let $\mathcal{X}$ be a set in $\mathbb{R}^n$, and whenever we say a property holds 'on $\mathcal{X}$,' we mean that it holds for every $x \in \mathcal{X}$. Smooth functions $\lambda_1(x), \ldots, \lambda_k(x)$, defined on $\mathcal{X}$, are said to be *independent on* $\mathcal{X}$ if their differential one-forms, $d\lambda_1(x), \ldots, d\lambda_k(x)$, are linearly independent on $\mathcal{X}$. In addition, if there exist $n-k$ smooth functions $\lambda_{k+1}, \ldots, \lambda_n$ such that $\mathrm{col}(\lambda_1(x), \ldots, \lambda_n(x))$ becomes a diffeomorphism from $\mathcal{X}$ to $\mathbb{R}^n$, then we say that $\lambda_1, \ldots, \lambda_k$ are *potential coordinate functions on* $\mathcal{X}$. We also recall that the Lie derivative of a function λ along the vector field f is $L_f\lambda(x) := \frac{\partial \lambda(x)}{\partial x} f(x)$ and $L_f(d\lambda) = dL_f\lambda$. The differential of a map p acting on the vector field v is denoted by p_*v. For a distribution $\mathcal{V}$, $p_*\mathcal{V} = \{p_*v \,|\, v \in \mathcal{V}\}$. We call a codistribution $\mathcal{W}$ at x^o *nonsingular* when $\dim \mathcal{W}$ is constant in a neighborhood of x^o. A codistribution $\mathcal{W}$ is *invariant* with respect to (w.r.t.) a vector field v if $L_v\mathcal{W} \subset \mathcal{W}$. If $\mathcal{W} = \mathrm{span}\,\{d\lambda_1, \ldots, d\lambda_k\}$ and $L_v(d\lambda_j) = dL_v\lambda_j \in \mathcal{W}$ for $j = 1, \ldots, k$, then $\mathcal{W}$ is invariant w.r.t. v. *Involutivity* of a codistribution is determined by the involutivity of its kernel which is a distribution [13]. A codistribution generated by the exact one-forms is always involutive. The operator mod denotes modulus after division (over the set of integers).

2. OBSERVABILITY NOTION AND MOTIVATING EXAMPLE

Let us formalize the notion of observability considered in this paper.

Definition 1 (Large-time observability)
System (1) with a switching signal $\sigma(\cdot)$ is *large-time uniformly observable on a set* $\mathcal{X} \subset \mathbb{R}^n$ if there exists a finite time $T > t_0$ so that $x(T)$ is determined uniquely from $y_{[t_0,T]}$, $u_{[t_0,T]}$, and $\sigma_{[t_0,T]}$ for any measurable input $u_{[t_0,T]}$, when the state $x(t)$ remains in $\mathcal{X}$ for $t \in [t_0, T]$. If the time $T > t_0$ is arbitrary, then system (1) is called *small-time uniformly observable on a set* $\mathcal{X}$. $\triangleleft$

Prepared using rncauth.cls
DOI: 10.1002/rnc

In case of no jump map (1b), the knowledge of $x(T)$, $\sigma_{[t_0,T]}$, and $u_{[t_0,T]}$ determines $x_{[t_0,T]}$ uniquely by (1a). This is not the case in general because the jump map (1b) may not be reversible. In this sense, the notion of observability studied in this paper is also referred to as 'determinability' in [27, 24] and 'reconstructability' in [22] (where the systems considered are linear). From the definition, if a certain mode of system (1) is small-time uniformly observable and the switching signal activates that mode at a certain time t_1, then the system is automatically large-time uniformly observable with $T > t_1$. Note that $x(T)$ may be reconstructed using the derivatives of $y(\cdot)$ and $u(\cdot)$ (although differentiation should not be used in the observer construction). It is noted that, although the observability in Definition 1 is uniform[†] w.r.t. the input u, uniformity w.r.t. the switching signal σ is not required.

The following example motivates the forth-coming discussion on the large-time observability.

Example 1 (Large-time observable switched system)
Let $\mathcal{X} := \{x \in \mathbb{R}^3 : x_1 > 0, x_3 > 0\}$ and suppose that the state $x(t)$, with $x(0) \in \mathcal{X}$, evolves according to system (1) that consists of three modes given by

$$\text{mode 1}: \begin{cases} \dot{x} = f_1(x) := \begin{pmatrix} 0.1x_3 \\ x_1^2 - x_3^2 + 2x_1 \\ 0.1(x_1+1) \end{pmatrix} \\ y = h_1(x) := x_2 \end{cases}$$

$$\text{mode 2}: \begin{cases} \dot{x} = f_2(x) := \begin{pmatrix} x_3 \\ -(x_1^2 - x_3^2 + 2x_1)x_2 \\ x_1 + 1 \end{pmatrix} \\ y = h_2(x) := x_1^2 - x_3^2 + 2x_1 \end{cases}$$

$$\text{mode 3}: \begin{cases} \dot{x} = f_3(x) := \begin{pmatrix} x_2^2 \\ -\frac{1}{2}x_2 \\ 0 \end{pmatrix} \\ y = h_3(x) := x_1 + x_2^2 \end{cases}$$

and the jump maps

$$p_{2,1}(x) := \begin{pmatrix} x_1 \\ x_1^2 - x_3^2 + 2x_1 + 0.1x_2 \\ x_3 \end{pmatrix}$$

with all other jump maps being identity (e.g., $p_{3,2}(x) = p_{1,3}(x) = x$). It is easily seen that, for any switching signal, the system has no finite escape time and the set $\mathcal{X}$ is forward invariant. None of the three modes are observable in the classical sense, which can be verified by inspecting the rank of the observability codistribution $\mathcal{O}_i := \mathrm{span}\,\{dh_i, dL_{f_i}h_i, dL_{f_i}^2 h_i, \cdots\}$ for each mode i. While the dimension of the state is 3, we obtain, for mode 1, $L_{f_1}h_1(x) = x_1^2 - x_3^2 + 2x_1$ and $L_{f_1}^k h_1(x) = 0$ for all $k \geq 2$, which ensures that $\mathcal{O}_1 = \mathrm{span}\,\{[0,1,0], [2x_1+2, 0, -2x_3]\}$ has rank 2 for all $x \in \mathcal{X}$. Similarly, it is seen that $\mathcal{O}_2 = \mathrm{span}\,\{[2x_1+2, 0, -2x_3]\}$ and $\mathcal{O}_3 = \mathrm{span}\,\{[1, 2x_2, 0]\}$, both of which indicate that their rank is 1 on $\mathcal{X}$.

Nevertheless, we claim that the switching among these three modes: $1 \to 2 \to 3$, makes it possible to recover complete information about the state. Assume that a particular execution has been observed on some time interval $[0, T]$ including two switching times t_1 (for $1 \to 2$) and t_2 (for $2 \to 3$) such that $0 < t_1 < t_2 < T$. Then, with the output $y(t)$ and its derivatives at hand, it is seen that the state $x(t)$ is recovered immediately after t_2. Indeed, it follows from the system equations at

[†] For nonlinear systems, observability depends on particular inputs in general. Therefore, uniform observability, which means 'observability for any input,' is a stronger notion than observability. See [8, 9, 10].

mode 1 and mode 2 that

$$
\begin{aligned}
x_2(t_2) &= x_2(t_2^-) = e^{-\int_{t_1}^{t_2^-} y(s)ds} x_2(t_1) \\
&= e^{-\int_{t_1}^{t_2^-} y(s)ds}(x_1^2(t_1^-) - x_3^2(t_1^-) + 2x_1(t_1^-) + 0.1x_2(t_1^-)) = e^{-\int_{t_1}^{t_2^-} y(s)ds}(\dot{y}(t_1^-) + 0.1y(t_1^-)).
\end{aligned} \tag{2}
$$

Then, the state $x_2(t_2)$ is determined from known quantities. Also, from mode 3, we determine

$$x_1(t_2) = y(t_2) - x_2^2(t_2).$$

Finally, from mode 2 and the fact that $x_3 > 0$ on $\mathcal{X}$,

$$x_3(t_2) = x_3(t_2^-) = +\sqrt{x_1^2(t_2^-) + 2x_1(t_2^-) - y(t_2^-)} = +\sqrt{x_1^2(t_2) + 2x_1(t_2) - y(t_2^-)}.$$

In this way, we can recover $x(t_2)$ (or $x(T)$ for any $T > t_2$). Thus, the switched system with switching sequence $1 \to 2 \to 3$ is large-time (uniform) observable on $\mathcal{X}$. $\triangleleft$

Let us highlight some interesting aspects of this example in order to motivate the technical details that follow:

1. The main idea of the example was to illustrate that even though the individual modes of the system are not observable, it is possible to extract partial information about the state from each mode. Under certain constraints on the dynamics of the system, it is then possible to accumulate all the information at some time instant in future so that it becomes possible to determine complete knowledge of the state of the system. In this example, we have seen that the information from mode 1 and mode 2 is combined with that of mode 3 at time t_2 to recover the state at that time instant.
2. In equation (2), even though x_2 is unobservable under mode 2 and mode 3, we are able to express x_2 at time t_2 as a function of the output measured during mode 2 and mode 1. This could be done because $x_2(t_1)$, obtained from the jump map $p_{2,1}(x)$, depends only on the observable quantities at t_1^- given by $x_2(= y)$ and $x_1^2 - x_3^2 + 2x_1(= \dot{y})$. Also, the evolution of x_2 under the dynamics of mode 2 over the interval $[t_1, t_2)$ depends only on the measurable signal $y(\cdot)$ and not the unobservable quantities. Since $x_2(t_2)$ is obtained by transporting $x_2(t_1)$ under the dynamics of mode 2, it is now possible to calculate $x_2(t_2)$. This enables collecting all the information at time t_2 to determine the whole state x.

The above mentioned arguments underline the basic ideas of our treatments about observability and observer design. This approach of combining the available information from various modes and preserving parts of it, leads to a sufficient condition for large-time observability, which will be formalized in the next section. Since the condition relies on the system structure, we also provide, in Section 4, the geometric conditions to verify such structural properties. The sufficient condition for large-time observability is in fact closely related to the observer construction and, to put the entire development into perspective, the next section discusses briefly how this condition leads to observer design.

3. OBSERVER SYNOPSIS: IN THE PERSPECTIVE OF OBSERVABILITY

In this section, we present the underlying idea for the observer design, which will be detailed in Section 5. The key aspect of our approach is the transformation of the dynamics at each mode to particular canonical structures, on which the observability is clearly seen assuming that the outputs and its derivatives are available. After presenting the structure in this section, existence of such a structure will be formulated under a geometric condition in Section 4. The design of asymptotic observer (to generate state estimates without using the derivatives of the output) based on this particular structure is given in Section 5.

DOI: 10.1002/rnc

Before proceeding, let us rename the switching sequence for convenience. That is, when the switching signal $\sigma(t)$ takes the mode sequence $\{q_1, q_2, \cdots\}$, we rename them as increasing integers $\{1, 2, 3, \cdots\}$ which is ever increasing even though the same mode is revisited. This way jump maps take the form $p_{2,1}, p_{3,2}, \dots$, so that the jump map at t_q is given by $p_{q+1,q}(\cdot)$, and for brevity we denote it simply as p_q, that is, $x(t_q) = p_q(x(t_q^-))$. When the mode is switched without the state jump (1b) from the mode number q, we take $p_q(x) = x$, and when the state jump occurs at mode q without mode switching we take $f_{q+1}(x) = f_q(x)$, $g_{q+1}(x) = g_q(x)$, and $h_{q+1}(x) = h_q(x)$.

We first note that the individual system at each mode may not be observable, calling for the classical observability decomposition [13]: Changing the coordinates so that the system is explicitly split into the observable part and the unobservable part. Thus, for a fixed mode q that comprises (1a) and (1c) (the jump map (1b) will be treated shortly), we assume the following:

Assumption 1 (Uniform observability decomposition)
For each mode $q \in \mathbb{N}$, there exists a diffeomorphism λ^q defined on $\mathcal{X}$ such that, with $\mathrm{col}(\xi_q, \xi_q') := \lambda^q(x) \in \mathbb{R}^n$, the dynamical equations (1a) and (1c) are transformed into

$$y = H_q(\xi_q), \tag{3a}$$
$$\dot{\xi}_q = F_q(\xi_q) + G_q(\xi_q)u, \quad \xi_q \in \mathbb{R}^{\nu_q}, \tag{3b}$$
$$\dot{\xi}_q' = F_q'(\xi_q', \xi_q) + G_q'(\xi_q', \xi_q)u, \quad \xi_q' \in \mathbb{R}^{n-\nu_q}, \tag{3c}$$

where the ξ_q-subsystem with the output y is small-time uniformly observable on $\Xi_q := \lambda^q_{\langle \nu_q \rangle}(\mathcal{X})$. $\lhd$

In equation (3), the state ξ_q is small-time uniformly observable at mode q so that it can be determined from the input and the output of mode q only. In fact, as long as we restrict our attention to (3a) and (3b), small-time uniform observability becomes the standard uniform observability that has often been studied in the literature (see [10] and references therein), which can be checked in various ways. For instance, if the class of inputs $\mathcal{U}$ consists of smooth functions, then one may try to find a function $\mathcal{E}$ such that

$$\xi_q = \mathcal{E}(y, \dot{y}, \dots, y^{(d_y^q - 1)}, u, \dot{u}, \dots, u^{(d_u^q - 1)}) \tag{4}$$

where $d_y^q \in \mathbb{N}$ and $d_u^q \in \mathbb{N}$, and that the function $\mathcal{E}(\,\cdot\,, u, \dot{u}, \dots, u^{(d_u^q-1)})$ is surjective onto Ξ_q for all $u(\cdot) \in \mathcal{U}$. The existence of such a function $\mathcal{E}$ is used as the definition of uniform observability in [25, 13]. Other ways to check uniform observability can be found in [18], but we will also present a geometric condition in Section 4.

The state ξ_q' denotes unobservable parts at mode q. However, if the information obtained from the previous mode is taken into account, then some components of ξ_q' may be determined. For example, suppose that some part of $\xi_q'(t_{q-1})$ depends only on the known $\xi_{q-1}(t_{q-1}^-)$, and the evolution of that part of $\xi_q'(t)$ over $[t_{q-1}, t_q)$ is governed by a differential equation that depends only on the known or observable quantities. Then, such components of $\xi_q'(t)$ are recovered completely during the interval $[t_{q-1}, t_q)$. This idea is formalized by the following structural assumption.

Assumption 2 (Switched canonical structure)
For each mode $q \in \mathbb{N}$, there exists a diffeomorphism $\mathrm{col}(\vartheta^q_{\langle k_q \rangle}, \omega^q_{\langle l_q \rangle}, \pi^q_{\langle n-k_q-l_q \rangle})$ on $\mathcal{X}$ such that, with $\theta_q := \vartheta^q_{\langle k_q \rangle}(x) \in \mathbb{R}^{k_q}$, $w_q := \omega^q_{\langle l_q \rangle}(x) \in \mathbb{R}^{l_q}$, and $\varsigma_q := \pi^q_{\langle n-k_q-l_q \rangle}(x) \in \mathbb{R}^{n-k_q-l_q}$, system (1) takes the following form:

$$\dot{\theta}_q = F_q^o(\theta_q, w_q) + G_q^o(\theta_q, w_q)u \tag{5a}$$
$$\dot{w}_q = F_q^\dagger(\theta_q, w_q) + G_q^\dagger(\theta_q, w_q)u \tag{5b}$$
$$\dot{\varsigma}_q = F_q^\ddagger(\varsigma_q, \theta_q, w_q) + G_q^\ddagger(\varsigma_q, \theta_q, w_q)u \tag{5c}$$

for $t \in [t_{q-1}, t_q)$ with

$$\begin{bmatrix} \theta_{q+1}(t_q) \\ w_{q+1}(t_q) \\ \varsigma_{q+1}(t_q) \end{bmatrix} = \begin{bmatrix} R_q^o(\theta_q(t_q^-), w_q(t_q^-), \varsigma_q(t_q^-)) \\ R_q^*(w_q(t_q^-), \xi_q(t_q^-), \xi_{q+1}(t_q)) \\ R_q^\ddagger(\theta_q(t_q^-), w_q(t_q^-), \varsigma_q(t_q^-)) \end{bmatrix}, \tag{5d}$$

DOI: 10.1002/rnc

where each component of $\theta_q = \vartheta^q_{\langle k_q \rangle}(x) \in \mathbb{R}^{k_q}$ is a smooth function of $\xi_q = \lambda^q_{\langle \nu_q \rangle}(x) \in \mathbb{R}^{\nu_q}$ of Assumption 1 (and $k_q \le \nu_q$); i.e., there exists a smooth map $\chi_q : \Xi_q \to \vartheta^q_{\langle k_q \rangle}(\mathcal{X})$ such that

$$\theta_q = \chi_q(\xi_q). \tag{6}$$

Let $l_1 = 0$ and w_1 be a null vector. $\triangleleft$

By the assumption, the states θ_q can be thought of as small-time uniformly observable since they depend only on the small-time uniformly observable state ξ_q. Meanwhile, the states $w_q(t)$ may not be observable. However, assuming that $w_{q-1}(t)$ is known for the interval $[t_{q-2}, t_{q-1})$, we note that the initial condition $w_q(t_{q-1}) = R^*_{q-1}(w_{q-1}(t^-_{q-1}), \xi_{q-1}(t^-_{q-1}), \xi_q(t_{q-1}))$ is known because both ξ_{q-1} and ξ_q are known. Moreover, the w_q-subsystem in (5b) shows that the evolution of w_q at mode q is not affected by the unknown component ς_q. It can also be interpreted that, through the evolution of w_q, the accumulated information around the time t_{q-1} (i.e., $w_{q-1}(t^-_{q-1})$, $\xi_{q-1}(t^-_{q-1})$, and $\xi_q(t_{q-1})$) is delivered up to the time t^-_q, and at the next mode, more information is accumulated through the observable component ξ_{q+1}. This idea leads to the following outcome.

Theorem 1 (Large-time uniform observability)
Under Assumptions 1 and 2, if there is a mode number $m \in \mathbb{N}$ such that $k_m + l_m = n$, then system (1) is large-time uniformly observable on $\mathcal{X}$. $\triangleleft$

This is because at any time T after the time t_{m-1}, enough information about the state is accumulated, and therefore, by inverting the diffeomorphism

$$\begin{bmatrix} \vartheta^m_{\langle k_m \rangle}(x(T)) \\ \omega^m_{\langle l_m \rangle}(x(T)) \end{bmatrix} = \begin{bmatrix} \theta_m(T) = \chi_m(\xi_m(T)) \\ w_m(T) \end{bmatrix}, \tag{7}$$

the state $x(T)$ is recovered.

A consequence of Theorem 1 is that system (1) with any switching signal containing the consecutive subsequence $\{1, 2, \ldots, m\}$ is large-time uniformly observable as well.

Remark 1
In equation (5d), the arguments of R^*_q could be just $w_q(t^-_q)$ and $\xi_q(t^-_q)$. However, by allowing $\xi_{q+1}(t_q)$ in R^*_q, the restriction is relaxed because it is basically asking that $\omega^{q+1}_{\langle l_{q+1} \rangle}(p_q(x))$ is a function of $(\omega^q_{\langle l_q \rangle}(x), \lambda^q_{\langle \nu_q \rangle}(x), \lambda^{q+1}_{\langle \nu_{q+1} \rangle}(x))$ rather than just $(\omega^q_{\langle l_q \rangle}(x), \lambda^q_{\langle \nu_q \rangle}(x))$. On the other hand, since $(\theta_q, w_q, \varsigma_q)$ gives a complete coordinate system in $\mathbb{R}^n$, the format of $R^\ddagger_q$ and R^o_q in (5d) is not a restriction. $\triangleleft$

The large-time uniform observability discussed so far inspires a way of constructing an observer. Here we discuss a synopsis of the observer construction, and the details are given in Section 5. First, we observe that, by (6) and from (3a), (3b), (5b), and (5d), there are maps F^*_q and G^*_q such that the evolution of ξ_q and w_q at each mode q is governed by

$$y = H_q(\xi_q) \tag{8a}$$

$$\dot{\xi}_q = F_q(\xi_q) + G_q(\xi_q)u \tag{8b}$$

$$\dot{w}_q = F^*_q(w_q, \xi_q) + G^*_q(w_q, \xi_q)u \tag{8c}$$

for $t \in [t_{q-1}, t_q)$, and

$$w_q(t_{q-1}) = R^*_{q-1}(w_{q-1}(t^-_{q-1}), \xi_{q-1}(t^-_{q-1}), \xi_q(t_{q-1})). \tag{8d}$$

We then design two separate observers for the component ξ_q and w_q to generate the corresponding estimates $\hat{\xi}_q$ and $\hat{w}_q$, respectively. While it is possible to obtain a good estimate of the observable component ξ_q, the variable w_q is not a directly observable quantity at mode q. Therefore, the role of the observer for w_q is not to reduce the error $\hat{w}_q(t) - w_q(t)$, but to deliver the estimates $\hat{\xi}_{q-1}(t^-_{q-1})$ and $\hat{w}_{q-1}(t^-_{q-1})$, that are obtained from the previously active mode and are encoded in the initial condition equation (8d) along with $\hat{\xi}_q(t_{q-1})$, to the next mode through $\hat{w}_q(t)$. Suppose that the input u and the output y are stored during the interval $[t_0, t_m)$, and at time $t = t_m^-$, an observer operates and computes the estimate $\hat{\xi}_q(t)$ of $\xi_q(t)$ for each interval $[t_{q-1}, t_q)$, $q = 1, \ldots, m$. Owing to small-time uniform observability, identifying ξ_q for each switching interval should be theoretically possible, and let us suppose that $\hat{\xi}_q(t) \approx \xi_q(t)$ on $[t_{q-1}, t_q)$ for now. With the estimate $\hat{\xi}_1(t_1^-) \approx \xi_1(t_1^-)$ and $\hat{\xi}_2(t_1) \approx \xi_2(t_1)$ for example, we obtain the estimate $\hat{w}_2(t_1) \approx w_2(t_1)$ by using (8d) (with all the states replaced by their estimates). Then, integration of (8c) for $q = 2$ results in an approximation $\hat{w}_2(t) \approx w_2(t)$ on $[t_1, t_2)$. (The error between $\hat{w}_2$ and w_2 may tend to increase during the interval, though.) This process repeats until we get $\hat{\xi}_m(t_m^-) \approx \xi_m(t_m^-)$ and $\hat{w}_m(t_m^-) \approx w_m(t_m^-)$. Assuming that $k_m + l_m = n$, the estimate $x^\star(t_m^-)$ of $x(t_m^-)$ is now determined uniquely from $\hat{\xi}_m(t_m^-)$ and $\hat{w}_m(t_m^-)$, by the map (7) (with $x(T)$, $\xi_m(T)$, and $w_m(T)$ replaced by $x^\star(t_m^-)$, $\hat{\xi}_m(t_m^-)$, and $\hat{w}_m(t_m^-)$ respectively). The details of the implementation of this idea, as well as error analysis, will be given in Section 5. In particular, for obtaining the estimate $\hat{\xi}_q(t)$ we employ the high-gain observers presented in, e.g., [7, 8, 25]. However, in order to integrate $\hat{w}_q(t)$ by (8c), we need rather good estimate $\hat{\xi}_q(t) \approx \xi_q(t)$ on the whole interval $[t_{q-1}, t_q)$ while the (high-gain) observer usually experiences transients (or, peaking in the estimates) before it provides a good estimate. So if we run the observer from the time t_{q-1} as usual, we may not get a good estimate for the transient period after t_{q-1}. We overcome this difficulty by proposing a novel back-and-forth observation technique in Section 5.

4. GEOMETRIC CONDITIONS FOR CANONICAL STRUCTURE AND OBSERVABILITY

In this section, we discuss a geometric condition that yields the switched canonical structure of Assumption 2 as well as the uniform observability decomposition of Assumption 1, and thus, leads to a sufficient condition for large-time uniform observability. For simplicity, we study the case of single input and single output ($n_u = n_y = 1$).

4.1. Geometric Condition for Large-time Uniform Observability: Local Version

According to [11, 16, 13], a geometric condition to decompose system (1) at a mode q into the form of (3) is as follows. Let the observability codistribution be $\tilde{\mathcal{O}}_q := \mathrm{span}\,\{dL_{v_1}L_{v_2}\cdots L_{v_j}h_q : v_j \in \{f_q, g_q\}, j \geq 0\}$. Then, the condition that the codistribution $\tilde{\mathcal{O}}_q$ is nonsingular at some $x^o \in \mathcal{X}$ (i.e., $\dim \tilde{\mathcal{O}}_q(x)$ is constant around x^o) guarantees, by Frobenius theorem, that there exist a local neighborhood $\mathcal{X}' \subset \mathcal{X}$ of x^o and a diffeomorphism defined on $\mathcal{X}'$ such that system (1) is represented as (3). Note that the codistribution $\tilde{\mathcal{O}}_q$ is invariant w.r.t. f_q and g_q by construction. However, since we are interested in the 'uniform observability' of the observable part in (3), the condition needs to be strengthened. For this, let us assume that the input class $\mathcal{U}$ includes the zero input ($u(t) \equiv 0$), and define $\mathcal{O}_q^j := \mathrm{span}\,\{dh_q, dL_{f_q}h_q, \ldots, dL_{f_q}^{j-1}h_q\}$.

Assumption 3 (Uniform observability of the observable part)
For each mode q, there is an integer $\nu_q (\leq n)$ such that

1. $\mathcal{O}_q^{\nu_q}$ is nonsingular at $x^o \in \mathcal{X}$ and $\dim \mathcal{O}_q^{\nu_q} = \nu_q$,
2. $\mathcal{O}_q^{\nu_q}$ is invariant w.r.t. f_q,
3. $\mathcal{O}_q^j$ is invariant w.r.t. g_q for $j = 1, \ldots, \nu_q$. ◁

By Assumption 3, there exists a neighborhood $\mathcal{X}'$ of x^o where Assumption 1 holds [16, 13]. In fact, by Assumption 3.1, there is a set $\mathcal{X}'$ on which $\xi_{q,j} = L_{f_q}^{j-1}h_q$, $j = 1, \ldots, \nu_q$, are potential coordinate functions. With this choice of coordinates, we obtain

$$\begin{aligned} y &= \xi_{q,1}, \\ \dot{\xi}_{q,j} &= \xi_{q,j+1} + L_{g_q}L_{f_q}^{j-1}h_q \cdot u, \qquad j = 1, \ldots, \nu_q - 1, \\ \dot{\xi}_{q,\nu_q} &= L_{f_q}^{\nu_q}h_q + L_{g_q}L_{f_q}^{\nu_q-1}h_q \cdot u. \end{aligned} \tag{9}$$

By Assumptions 3.2 and 3.3, Lemma 1.1 in the Appendix can be employed repeatedly to show that $L_{f_q}^{\nu_q} h_q$ depends only on $\xi_q \in \mathbb{R}^{\nu_q}$, and $L_{g_q} L_{f_q}^{j-1} h_q$ depends only on $\xi_{q,1}, \cdots, \xi_{q,j}$. Gauthier *et al.* [8] have shown that this triangular structure is sufficient (and necessary as well in the case of single output) for (small-time) uniform observability of ξ_q. (Or, by taking successive derivatives of the output in ξ_q-coordinates, one can simply compute the map (4).)

Now let us denote $\mathcal{O}_q^{\nu_q}$ simply by $\mathcal{O}_q$. For $q \geq 2$, define $\mathcal{O}'_q := \operatorname{span}\{d(h_q \circ p_{q-1}), d(L_{f_q} h_q \circ p_{q-1}), \cdots, d(L_{f_q}^{\nu_q - 1} h_q \circ p_{q-1})\}$. We then define a sequence of codistributions $\mathcal{W}_q$ for $q \in \mathbb{N}$, with $\mathcal{W}_0 := \{0\}$. In particular,

> let $\mathcal{W}_q$ be *the largest nonsingular and involutive codistribution, invariant with respect to f_q and g_q, contained in* $(\mathcal{O}_q + \mathcal{W}_{q-1})$ *such that* $(p_q)_*(\ker \mathcal{W}_q \cap \ker \mathcal{O}'_{q+1}) \subset \ker \mathcal{W}_q$.

Following observations are immediate: (a) By Assumption 3, the codistribution $\mathcal{O}_q$ itself is invariant w.r.t. f_q and g_q. (b) If $p_q(x) = x$, so that there is no state jump, then the condition $(p_q)_*(\ker \mathcal{W}_q \cap \ker \mathcal{O}'_{q+1}) \subset \ker \mathcal{W}_q$ automatically holds. (c) The “largest” codistribution is well-defined because involutivity and invariance of a codistribution generated by exact one-forms is preserved under the addition, and if two smooth nonsingular codistributions $\mathcal{W}_a$ and $\mathcal{W}_b$ satisfy $p_*(\ker \mathcal{W}_i \cap \mathcal{D}) \subset \ker \mathcal{W}_i$, where $i \in \{a, b\}$, for any differentiable map p and any distribution $\mathcal{D}$, then $p_*(\ker(\mathcal{W}_a + \mathcal{W}_b) \cap \mathcal{D}) \subset \ker(\mathcal{W}_a + \mathcal{W}_b)$.[‡]

Theorem 2 (Large-time uniform observability: Local version)
If Assumption 3 holds; and the codistributions $\mathcal{W}_q$, $\mathcal{O}_q + \mathcal{W}_{q-1}$, and $\mathcal{W}_q + \mathcal{O}'_{q+1}$, $q \in \mathbb{N}$, are nonsingular at $x^o \in \mathcal{X}$, then there exist diffeomorphisms on a neighborhood $\mathcal{X}' \subset \mathcal{X}$ of x^o which induce the structures (3) and (5) proposed in Assumptions 1 and 2, respectively. Furthermore, if there is an integer $m \in \mathbb{N}$ such that

$$\dim(\mathcal{O}_m + \mathcal{W}_{m-1})(x^o) = n \tag{10}$$

(or simply $\dim \mathcal{W}_m(x^o) = n$ because $\mathcal{W}_m = \mathcal{O}_m + \mathcal{W}_{m-1}$ by construction), then system (1) is large-time uniformly observable on $\mathcal{X}'$. $\triangleleft$

The proof of the theorem appears after the statement of Theorem 3 in the next subsection.

4.2. *Geometric Condition for Large-time Uniform Observability: Global Version*

Compared to the local observability in the previous subsection, the condition in this subsection may be thought of as a “global” version in the sense that the observability holds over the set $\mathcal{X}$ given *a priori*. In this case, the condition explicitly assumes the existence of coordinate functions, which makes its statements no more compact.

Assumption 4
For each mode q,

(a) there is an integer $\nu_q (\leq n)$ such that

 (1) $\{h_q, L_{f_q} h_q, \ldots, L_{f_q}^{\nu_q - 1} h_q\}$ are potential coordinate functions on $\mathcal{X}$,
 (2) $\mathcal{O}_q = \mathcal{O}_q^{\nu_q}$ is invariant w.r.t. f_q,
 (3) $\mathcal{O}_q^j$ is invariant w.r.t. g_q for $j = 1, \ldots, \nu_q$,

(b) there are potential coordinate functions $\{\vartheta^q_{\langle k_q \rangle}, \omega^q_{\langle l_q \rangle} : k_q + l_q = \dim \mathcal{W}_q\}$ on $\mathcal{X}$ such that

$$\mathcal{W}_q = \operatorname{span}\{d\vartheta_1^q, \cdots, d\vartheta_{k_q}^q, d\omega_1^q, \cdots, d\omega_{l_q}^q\}, \quad d\vartheta_j^q \in \mathcal{O}_q, \quad d\omega_j^q \notin \mathcal{O}_q, \tag{11}$$

[‡] For each $i \in \{a, b\}$, $\ker(\mathcal{W}_a + \mathcal{W}_b) \subset \ker \mathcal{W}_i$, so that $(\ker(\mathcal{W}_a + \mathcal{W}_b) \cap \mathcal{D}) \subset \ker \mathcal{W}_i \cap \mathcal{D}$, which in turn implies that $p_*(\ker(\mathcal{W}_a + \mathcal{W}_b) \cap \mathcal{D}) \subset p_*(\ker \mathcal{W}_i \cap \mathcal{D}) \subset \ker \mathcal{W}_i$ by the assumption. Therefore, we have that $p_*(\ker(\mathcal{W}_a + \mathcal{W}_b) \cap \mathcal{D}) \subset \ker \mathcal{W}_a \cap \ker \mathcal{W}_b = \ker(\mathcal{W}_a + \mathcal{W}_b)$.

(c) there are potential coordinate functions $\{\mu^q_{\langle r_q \rangle} : r_q = \dim(\mathcal{O}_q + \mathcal{W}_{q-1})\}$ on $\mathcal{X}$ such that

$$\begin{aligned}\mathcal{O}_q + \mathcal{W}_{q-1} &= \operatorname{span}\{dh_q, dL_{f_q}h_q, \cdots, dL_{f_q}^{\nu_q-1}h_q, \\ &\qquad d\vartheta_1^{q-1}, \cdots, d\vartheta_{k_{q-1}}^{q-1}, d\omega_1^{q-1}, \cdots, d\omega_{l_{q-1}}^{q-1}\} \\ &= \operatorname{span}\{d\mu_1^q, \ldots, d\mu_{r_q}^q\}\end{aligned} \tag{12}$$

where each μ_j^q is a function of $\{h_q, \cdots, L_{f_q}^{\nu_q-1}h_q, \vartheta_1^{q-1}, \cdots, \vartheta_{k_{q-1}}^{q-1}, \omega_1^{q-1}, \cdots, \omega_{l_{q-1}}^{q-1}\}$,

(d) there are potential coordinate functions $\{\mu'^q_{\langle r'_q \rangle} : r'_q = \dim(\mathcal{W}_q + \mathcal{O}'_{q+1}) - \dim \mathcal{W}_q\}$ on $\mathcal{X}$ such that

$$\mathcal{W}_q + \mathcal{O}'_{q+1} = \operatorname{span}\{d\vartheta_1^q, \cdots, d\vartheta_{k_q}^q, d\omega_1^q, \cdots, d\omega_{l_q}^q, d\mu_1'^q, \ldots, d\mu_{r'_q}'^q\} \tag{13}$$

where each $\mu_j'^q$ is a function of $\{h_{q+1} \circ p_q, \cdots, L_{f_{q+1}}^{\nu_{q+1}-1}h_{q+1} \circ p_q\}$. ◁

Remark 2
Reason that k_q may be less than ν_q is because of the condition $(p_q)_*(\ker \mathcal{W}_q \cap \ker \mathcal{O}'_{q+1}) \subset \ker \mathcal{W}_q$. For example, since $\mathcal{O}_1$ is involutive (because it is generated by exact one-forms), nonsingular, and invariant w.r.t. f_1 and g_1 on $\mathcal{X}$ (by Assumption 4.(a)), the largest involutive and invariant codistribution $\mathcal{W}_1$ contained in $\mathcal{O}_1 + \mathcal{W}_0 = \mathcal{O}_1$ is $\mathcal{O}_1$ itself (i.e., $k_1 = \nu_1$) if the condition $(p_1)_*(\ker \mathcal{W}_1 \cap \ker \mathcal{O}'_2) \subset \ker \mathcal{W}_1$ is not taken into account. ◁

Theorem 3 (Large-time uniform observability: Global version)

1. If Assumption 4 holds, then the canonical structures (3) and (5) from Assumptions 1 and 2, respectively, exist globally on the set $\mathcal{X}$.
2. If Assumption 4 holds and there is an integer $m \in \mathbb{N}$ such that

$$\dim(\mathcal{O}_m + \mathcal{W}_{m-1}) = n \quad \text{on } \mathcal{X},$$

then system (1) is large-time uniformly observable on $\mathcal{X}$. ◁

Proof of Theorem 3 is found in the Appendix while the proof of Theorem 2 is given here.

Proof of Theorem 2
When the codistributions $\mathcal{O}_q$, $\mathcal{W}_q$, $\mathcal{O}_q + \mathcal{W}_{q-1}$, and $\mathcal{W}_q + \mathcal{O}'_{q+1}$ are nonsingular at a point x^o, then Assumptions 4.(a).(1), (b), (c), and (d) hold in a neighborhood $\mathcal{X}'$ of x^o. Indeed, since the smooth codistribution $\mathcal{W}_q$ is nonsingular and involutive, then by Frobenius theorem, there exist potential coordinate functions $\vartheta^q_{\langle k_q \rangle}$ and $\omega^q_{\langle l_q \rangle}$, whose differential one-forms span $\mathcal{W}_q$ in a local neighborhood of x^o (see the proof of [13, Theorem 1.4.1]), thus satisfying Assumption 4.(b). By the same argument, involutivity and nonsingularity of $\mathcal{O}_q + \mathcal{W}_{q-1}$ and $\mathcal{W}_q + \mathcal{O}'_{q+1}$ yield independent smooth functions that satisfy Assumptions 4.(c) and 4.(d), respectively, in a neighborhood of x^o. Let $\mathcal{X}'$ be the intersection of such neighborhoods. From the statement of Theorem 3.1, with $\mathcal{X} = \mathcal{X}'$, the existence of diffeomorphisms yielding (3) and (5) is now guaranteed. Now, if (10) holds, then $\dim(\mathcal{O}_m + \mathcal{W}_{m-1})(x') = n$ for each $x' \in \mathcal{X}'$, and the statement of Theorem 3.2 proves the large-time uniform observability of system (1) on a local neighborhood $\mathcal{X}'$ of x^o. □

Another interpretation of Theorems 2 and 3 in terms of "distribution" is also possible. In order to recover the system state $x(t)$, the partial information from each mode is quantified in terms of the maximal integral submanifold of the distribution $\mathcal{O}_q^\perp$ which has the property that the states on the slices (or "leaves") of this submanifold are not distinguishable by the output of mode q. If there are certain states on this submanifold which were observable under the previous mode $q-1$, and are also decoupled from the remaining indistinguishable states of the current mode q, then we can carry this additional information forward, thereby reducing the uncertainty about the unknown state. This is exactly the intuition formalized in constructing $\mathcal{W}_q$ and later developed in Assumption 4. This process is continued at each mode q and if at some point in time, the uncertainty is reduced to a point, we term the system observable as the entire state can now be reconstructed.

DOI: 10.1002/rnc

Remark 3 (Multiple-input multiple-output (MIMO) case)
The geometric conditions given in Assumptions 3 and 4.(a) enable the transformation of the individual mode into a particular triangular form, which eventually leads to the desired canonical form (3). However, for MIMO case, there are several structures that guarantee uniform observability; and each one of these structures has different geometric conditions associated to them (for example [20]). Therefore, without giving the geometric condition, if we assume that the individual modes of the system are endowed with uniform observability property in the form of structure (3), then the results hold for MIMO case as well.

4.3. *Example: Linear Case*

To get a better understanding of the conditions given in the previous subsection, let us consider the case of linear systems. If the system is linear, then Assumption 4 and Theorem 3 (or, Assumption 3 and Theorem 2) become much simpler leading to a global result (i.e., $\mathcal{X} = \mathbb{R}^n$). Consider a linear version of (1) as

$$\dot{x}(t) = A_{\sigma(t)}x(t) + B_{\sigma(t)}u(t), \qquad t \neq \{t_q\}, \tag{14a}$$
$$x(t_q) = P_{\sigma(t_q^-)}x(t_q^-), \qquad x \in \mathbb{R}^n, u \in \mathbb{R}, y \in \mathbb{R}, \tag{14b}$$
$$y(t) = C_{\sigma(t)}x(t). \tag{14c}$$

Define the matrices O_q and O'_q as

$$O_q := \mathrm{col}(C_q, C_qA_q, \ldots, C_qA_q^{\nu_q-1}), \qquad \text{and} \qquad O'_q := O_qP_{q-1},$$

where ν_q is the observability index of (C_q, A_q) so that $\mathrm{rank}\, O_q = \nu_q$. Then the codistribution $\mathcal{O}_q$ of the previous subsection corresponds to the range space $\mathcal{R}(O_q^\top)$, and $\mathcal{O}'_q$ to $\mathcal{R}(O_q'^\top)$. It is noted that the codistribution, related to the vector subspace in this way, is always nonsingular on $\mathcal{X} = \mathbb{R}^n$.

Before presenting the linear version, a terminology is defined which is taken from [4]. A subspace $\mathcal{V}$ is a conditioned invariant subspace under a linear transformation given by a matrix A, w.r.t. another subspace $\mathcal{V}'$ (or, briefly, an $(A, \mathcal{V}')$-*conditioned invariant*) if

$$A(\mathcal{V} \cap \mathcal{V}') \subset \mathcal{V}.$$

Corollary 1
Define, with $\mathcal{V}_0 = \{0\}$, the sequence of subspaces:

> $\mathcal{V}_q$ is the largest $A_q^\top$-invariant subspace contained in $(\mathcal{R}(O_q^\top) + \mathcal{V}_{q-1})$ such that $\mathcal{V}_q^\perp$ is $(P_q, \ker O'_{q+1})$-conditioned invariant.

If there exists an $m \in \mathbb{N}$ such that $\dim(\mathcal{R}(O_m^\top) + \mathcal{V}_{m-1}) = n$, then system (14) is large-time observable[§] on $\mathbb{R}^n$.

Proof
We can simply show that Assumption 4 holds with the subspace $\mathcal{V}_q$ corresponding to the codistribution $\mathcal{W}_q$. Indeed, Assumption 4.(a) trivially holds with $\mathcal{O}_q = \mathrm{span}\{d(\eta_j x) = \eta_j :$ η_j is each row of $O_q\}$, $f_q = A_q x$, and $g_q = B_q$. Now we can always find a basis for $\mathcal{V}_q$ as $\{\varphi_1^q, \cdots, \varphi_{k_q}^q, \varpi_1^q, \cdots, \varpi_{l_q}^q\}$ with $k_q + l_q = \dim \mathcal{V}_q$ $(k_q \leq \nu_q)$ such that $\varphi_j^q \in \mathcal{R}(O_q^\top)$ and $\varpi_j^q \notin \mathcal{R}(O_q^\top)$. Assumption 4.(b) holds with $\mathcal{W}_q$ being the span of those basis vectors of $\mathcal{V}_q$. We also take some vectors λ_j^q in $\mathcal{R}(O_q^\top)$ such that $\mathcal{R}(O_q^\top) + \mathcal{V}_{q-1}$ has the basis $\{\lambda_1^q, \cdots, \lambda_\rho^q, \varphi_1^{q-1}, \cdots, \varphi_{k_{q-1}}^{q-1}, \varpi_1^{q-1}, \cdots, \varpi_{l_{q-1}}^{q-1}\}$ where $k_{q-1} + l_{q-1} + \rho = \dim(\mathcal{R}(O_q^\top) + \mathcal{V}_{q-1})$; then Assumption 4.(c) holds. Similarly, we take some vectors $\mu_j'^q$ among the columns of $O_{q+1}'^\top$ such

[§]For linear systems, uniform observability is equivalent to the standard observability.

that $\mathcal{V}_q + \mathcal{R}(O_{q+1}'^{\top})$ has the basis $\{\varphi_1^q, \cdots, \varphi_{k_q}^q, \varpi_1^q, \cdots, \varpi_{l_q}^q, \mu_1'^q, \cdots, \mu_{r_q'}'^q\}$ where $k_q + l_q + r_q' = \dim(\mathcal{V}_q + \mathcal{R}(O_{q+1}'^{\top}))$, which guarantees Assumption 4.(d).

On the other hand, invariance of $\mathcal{V}_q$ w.r.t. $A_q^{\top}$ implies the invariance of $\mathcal{W}_q$ w.r.t. f_q and g_q. (Invariance w.r.t. g_q trivially follows because $g_q = B_q$ is a constant vector field.) Finally, $(P_q, \ker O_{q+1}')$-conditioned invariance of $\mathcal{V}_q^{\perp}$ implies that

$$P_q(\mathcal{V}_q^{\perp} \cap \ker O_{q+1}') \subset \mathcal{V}_q^{\perp}, \tag{15}$$

which corresponds to the condition $(p_q)_*(\ker \mathcal{W}_q \cap \ker \mathcal{O}_{q+1}') \subset \ker \mathcal{W}_q$. □

For convenience, let us construct the linear version of (8) here. For this, suppose that $\mathcal{V}_q = \operatorname{span}\{\varphi_1^q, \cdots, \varphi_{k_q}^q, \varpi_1^q, \cdots, \varpi_{l_q}^q\}$ where all column vectors φ_j^q's and ϖ_j^q's are a basis of $\mathcal{V}_q$ such that $\varphi_j^q \in \mathcal{R}(O_q^{\top})$ and $\varpi_j^q \notin \mathcal{R}(O_q^{\top})$. With $\Theta_q := [\varphi_1^q, \cdots, \varphi_{k_q}^q]^{\top}$ and $W_q := [\varpi_1^q, \cdots, \varpi_{l_q}^q]^{\top}$, define the coordinates $\xi_q := O_q x$, $\theta_q := \Theta_q x$, and $w_q := W_q x$. We also note that, by construction, there is a matrix $X_q \in \mathbb{R}^{k_q \times \nu_q}$ such that $\Theta_q = X_q O_q$.

By (15), and the fact that $P_q^{-1}\mathcal{V}_q^{\perp} = \ker[\Theta_q^{\top}\ W_q^{\top}]^{\top} P_q$, there exist two matrices $S_q^{\diamond}$ and $S_q^{\diamond\diamond}$ such that

$$\begin{bmatrix}\Theta_q \\ W_q\end{bmatrix} P_q = S_q^{\diamond} \begin{bmatrix}\Theta_q \\ W_q\end{bmatrix} + S_q^{\diamond\diamond} O_{q+1} P_q.$$

Likewise, $A_q^{\top}$-invariance of $\mathcal{V}_q$ implies the existence of matrices F_q^* such that

$$A_q^{\top} W_q^{\top} = [\Theta_q^{\top}\ W_q^{\top}] F_q^{*\top},$$

and, since $\mathcal{V}_q \subset O_q + \mathcal{V}_{q-1}$, there are two matrices $S_q^{\dagger}$ and $S_q^{\ddagger}$ such that

$$W_q^{\top} = O_q^{\top} S_q^{\dagger\top} + [\Theta_{q-1}^{\top}\ W_{q-1}^{\top}] S_q^{\ddagger\top}.$$

Taking all these equations into account, we obtain that

$$\begin{aligned}
y &= H_q \xi_q \\
\dot{\xi}_q &= F_q \xi_q + G_q u \\
\dot{w}_q &= W_q A_q x + W_q B_q u \\
&= F_q^* \begin{bmatrix}\Theta_q \\ W_q\end{bmatrix} x + W_q B_q u = F_q^* \begin{bmatrix} X_q \xi_q \\ w_q \end{bmatrix} + W_q B_q u \\
w_q(t_{q-1}) &= W_q x(t_{q-1}) \\
&= S_q^{\dagger} O_q x(t_{q-1}) + S_q^{\ddagger} \begin{bmatrix}\Theta_{q-1} \\ W_{q-1}\end{bmatrix} (P_{q-1} x(t_{q-1}^-)) \\
&= S_q^{\dagger} \xi_q(t_{q-1}) + S_q^{\ddagger} S_{q-1}^{\diamond} \begin{bmatrix}\Theta_{q-1} \\ W_{q-1}\end{bmatrix} x(t_{q-1}^-) + S_q^{\ddagger} S_{q-1}^{\diamond\diamond} O_q P_{q-1} x(t_{q-1}^-) \\
&= S_q^{\dagger} \xi_q(t_{q-1}) + S_q^{\ddagger} S_{q-1}^{\diamond} \begin{bmatrix} X_{q-1}\xi_{q-1}(t_{q-1}^-) \\ w_{q-1}(t_{q-1}^-)\end{bmatrix} + S_q^{\ddagger} S_{q-1}^{\diamond\diamond} \xi_q(t_{q-1}),
\end{aligned}$$

where the first two equations are from observability decomposition so that (H_q, F_q) is an observable pair. These equations correspond to the strucuture presented in (8).

4.4. *Revisiting Example 1*

As a showcase, let us verify Assumption 4 for the switched system of Example 1. First of all, since $\dim \mathcal{O}_1 = \nu_1 = 2$, $\dim \mathcal{O}_2 = \nu_2 = 1$, and $\dim \mathcal{O}_3 = \nu_3 = 1$, the functions $\lambda_j^q = L_{f_q}^{j-1} h_q$ are listed as

$$\lambda_1^1(x) = x_2, \quad \lambda_2^1(x) = x_1^2 - x_3^2 + 2x_1,$$
$$\lambda_1^2(x) = x_1^2 - x_3^2 + 2x_1, \quad \lambda_1^3(x) = x_1 + x_2^2,$$

DOI: 10.1002/rnc

and

$$\lambda_1^2(p_1(x)) = x_1^2 - x_3^2 + 2x_1, \quad \lambda_1^3(p_2(x)) = x_1 + x_2^2.$$

This leads to

$$\begin{aligned} \mathcal{O}_1 &= \operatorname{span}\{[0,1,0],[2x_1+2,0,-2x_3]\}, & & \\ \mathcal{O}_2 &= \operatorname{span}\{[2x_1+2,0,-2x_3]\}, & \mathcal{O}_2' &= \operatorname{span}\{[2x_1+2,0,-2x_3]\}, \\ \mathcal{O}_3 &= \operatorname{span}\{[1,2x_2,0]\}, & \mathcal{O}_3' &= \operatorname{span}\{[1,2x_2,0]\}. \end{aligned}$$

Starting with $\mathcal{W}_0 = \{0\}$, we pick $\mathcal{W}_1 = \mathcal{O}_1$ since $\mathcal{O}_1$ is the largest, f_1-invariant, nonsingular, and involutive codistribution contained in $\mathcal{O}_1$ on $\mathcal{X} = \{x \in \mathbb{R}^3 : x_1 > 0, x_3 > 0\}$, and satisfies that $(p_1)_*(\ker \mathcal{W}_1 \cap \ker \mathcal{O}_2') \subset \ker \mathcal{W}_1$ because¶

$$\begin{bmatrix} 1 & 0 & 0 \\ 2x_1+2 & 0.1 & -2x_3 \\ 0 & 0 & 1 \end{bmatrix} \cdot \operatorname{span}\left\{\begin{bmatrix} x_3 \\ 0 \\ x_1+1 \end{bmatrix}\right\} \subset \operatorname{span}\left\{\begin{bmatrix} x_3 \\ 0 \\ x_1+1 \end{bmatrix}\right\}.$$

Here, we simply let $\mu_j^1 = \vartheta_j^1 = \lambda_j^1$, $j = 1, 2$, so that $k_1 = r_1 = 2$ and $l_1 = 0$. Since $\mathcal{W}_1 + \mathcal{O}_2' = \mathcal{W}_1$, $r_1' = 0$.

Noting that $\mathcal{W}_2 \subset \mathcal{O}_2 + \mathcal{W}_1 = \mathcal{W}_1$, we take $\vartheta_1^2(x) = \lambda_1^2(x) = x_1^2 - x_3^2 + 2x_1$ and $\omega_1^2(x) = x_2$ so that

$$\mathcal{W}_2 = \operatorname{span}\{d\vartheta_1^2, d\omega_1^2\} = \operatorname{span}\{[2x_1+2,0,-2x_3],[0,1,0]\} = \mathcal{W}_1 = \mathcal{O}_1,$$

which is, on $\mathcal{X}$, the largest nonsingular and involutive codistribution, invariant w.r.t. f_2, contained in $\mathcal{O}_2 + \mathcal{W}_1$, and satisfies that $(p_2)_*(\ker \mathcal{W}_2 \cap \ker \mathcal{O}_3') \subset \ker \mathcal{W}_2$ because $\ker \mathcal{W}_2 \cap \ker \mathcal{O}_3' = \{0\}$ (or, because $p_2(x) = x$). Indeed, invariance w.r.t. f_2 is verified as

$$\begin{aligned} L_{f_2} d\omega_1^2 &= L_{f_2} dx_2 = dL_{f_2} x_2 = d([0,1,0]\cdot f_2) \\ &= [-(2x_1+2)x_2, -(x_1^2 - x_3^2 + 2x_1), 2x_3x_2] \in \mathcal{W}_2, \\ L_{f_2} d\vartheta_1^2 &= d\left(\frac{\partial \vartheta_1^2}{\partial x}\cdot f_2\right) = d([2x_1+2,0,-2x_3]\cdot f_2) = 0 \in \mathcal{W}_2. \end{aligned}$$

As a result, $k_2 = l_2 = 1$, $r_2 = \dim(\mathcal{O}_2 + \mathcal{W}_1) = 2$, $\mu_1^2 = \vartheta_1^2$, $\mu_2^2 = \omega_1^2$, $r_2' = \dim(\mathcal{W}_2 + \mathcal{O}_3') - \dim \mathcal{W}_2 = 1$, and $\mu_1'^2(x) = \lambda_1^3(p_2(x)) = x_1 + x_2^2$.

Finally, $\mathcal{W}_3 \subset \mathcal{O}_3 + \mathcal{W}_2 = \mathbb{R}^3$ so that we take $\mathcal{W}_3 = \mathbb{R}^3$ and $k_3 = 1$ and $l_3 = 2$. Therefore, the hypothesis in Theorem 3 holds with $m = 3$.

These functions can now be used to arrive at the form presented in (8). For mode 1, let $\xi_{1,1} := \lambda_1^1(x)$ and $\xi_{1,2} := \lambda_2^1(x)$. Then we get

$$\begin{aligned} y = \xi_{1,1}, \qquad \dot{\xi}_{1,1} &= x_1^2 - x_3^2 + 2x_1 = \xi_{1,2}, \\ \dot{\xi}_{1,2} &= 0.2x_1x_3 - 0.2x_3(x_1+1) + 0.2x_3 = 0. \end{aligned} \tag{16a}$$

Similarly, for mode 2, we introduce the following coordinates, $\xi_2 := \lambda_1^2(x) = x_1^2 - x_3^2 + 2x_1$ and $w_2 := \omega_1^2(x) = x_2$. The dynamics of mode 2 then take the following form:

$$\begin{aligned} y = \xi_2, \qquad \dot{\xi}_2 &= 2x_1x_3 - 2x_3(x_1+1) + 2x_3 = 0, \\ \dot{w}_2 &= -(x_1^2 - x_3^2 + 2x_1)x_2 = -\xi_2 w_2, \\ w_2(t_1) &= x_1^2(t_1^-) - x_3^2(t_1^-) + 2x_1(t_1^-) + 0.1x_2(t_1^-) = \xi_{1,2}(t_1^-) + 0.1\xi_{1,1}(t_1^-). \end{aligned} \tag{16b}$$

¶It actually shows that $(p_1)_* \ker \mathcal{W}_1 \subset \ker \mathcal{W}_1$, which implies that, because of Lemma 1.3, the map R_{q-1}^* in (8d) does not depend on ξ_q, or in this particular example, $w_2(t_1)$ does not depend on ξ_2 as shown in (16b).

Finally, for the third mode, the new coordinates are $\xi_3 := \lambda_1^3(x) = x_1 + x_2^2$, $w_{3,1} := x_2$, and $w_{3,2} := x_1^2 - x_3^2 + 2x_1$, and the resulting system dynamics become:

$$\begin{aligned} y = \xi_3, \qquad \dot{\xi}_3 &= x_2^2 + 2x_2(-\frac{1}{2}x_2) = 0, \\ \dot{w}_{3,1} &= -\frac{1}{2}x_2 = -\frac{1}{2}w_{3,1}, & w_{3,1}(t_2) &= x_2(t_2^-) = w_2(t_2^-), \\ \dot{w}_{3,2} &= 2x_1x_2^2 + 2x_2^2 = 2(x_1+1)x_2^2 & w_{3,2}(t_2) &= x_1^2(t_2^-) - x_3^2(t_2^-) + 2x_1(t_2^-) \\ &= 2(\xi_3 - w_{3,1}^2 + 1)w_{3,1}^2, & &= \xi_2(t_2^-). \end{aligned} \tag{16c}$$

It is seen that the system dynamics in the new coordinates indeed follow the structure prescribed in (8), which is crucial for the observer design in the next section.

5. OBSERVER DESIGN

Based on the study of large-time observability, let us now discuss the design of an asymptotic observer for system (1). By asymptotic observer, we mean an observer whose state $\hat{x}(t)$ converges to the plant state $x(t)$ as time tends to infinity. It is shown in this section that an asymptotic observer can be constructed based on the following assumptions.

Assumption 5
1. The solution $x(t)$ of the plant (1) remains in a compact set $\mathcal{X} \subset \mathbb{R}^n$, and the input $u(t)$ (and its derivatives, whenever necessary) is uniformly bounded. Let $|u(t)| \leq C_u$ for $t \geq t_0$.
2. The switching is persistent and occurs within the duration D; that is,

$$\tau_q := t_q - t_{q-1} \leq D, \qquad \forall q \in \mathbb{N} \tag{17}$$

   where t_q is the switching time.
3. There are a finite number of modes in system (1), and the same mode sequence repeats. We denote each cycle of modes as $\{1, 2, \cdots, m\}$, so that $\sigma(t) = ((q-1) \bmod m) + 1$ for $t \in [t_{q-1}, t_q)$, $q \in \mathbb{N}$, and the mode q and $q + m$ refer to the same mode.
4. Assumptions 1 and 2 hold on $\bar{\mathcal{X}}$, and $k_m + l_m = n$ so that system (1) is large-time uniformly observable on $\bar{\mathcal{X}}$, where $\bar{\mathcal{X}}$ is any set that properly contains the set $\mathcal{X}$ and their boundaries do not intersect. $\triangleleft$

In Assumption 5, item 1 is often the outcome of control, which is not an observer problem but makes the observer problem much easier. It should not be assumed *a priori* when the estimated state $\hat{x}(t)$ is used for a state-feedback control, in a certainty-equivalence manner, and the stability of the overall closed-loop system needs to be analyzed differently, which is however beyond the scope of this paper. Item 2 is introduced for a practical reason: As discussed before, when the system is not (small-time) observable at any mode, only switching can provide new and fresh information into the observer. Therefore, assuming that the switching is persistent is natural in order to construct an observer whose estimate converges to the true value in infinite time. More discussions on items 2 and 3 can be found in Section 6.1. The requirement that the assumptions hold on a slightly larger set $\bar{\mathcal{X}}$ in item 4 is because we will utilize the Lipschitz extension below.

For a given map $f : \bar{\mathcal{A}} \to \mathbb{R}^k$, a *Lipschitz extension* of f from a compact set $\mathcal{A} (\subset \bar{\mathcal{A}} \subset \mathbb{R}^l)$ is a map $\bar{f} : \mathbb{R}^l \to \mathbb{R}^k$ such that it is globally Lipschitz, $\bar{f}(x) = f(x)$ for every $x \in \mathcal{A}$, and it preserves the structure of $f(x)$ (for example, if the arguments of $f_j(x)$ consist only of x_1 and x_2 on $\bar{\mathcal{A}}$, then the arguments of $\bar{f}_j(x)$ are the same on $\mathbb{R}^l$). When a compact set $\mathcal{A}$ is contained in $\bar{\mathcal{A}}$ and their boundaries do not intersect, there always exists a Lipschitz extension of f from $\mathcal{A}$ (see [19]). One simple choice of Lipschitz extension of f is to saturate its arguments, i.e., $\bar{f}(x) = f(\mathrm{sat}_1(x_1), \cdots, \mathrm{sat}_l(x_l))$, with each saturation being inactive ($\mathrm{sat}_j(x_j) = x_j$) on $\mathcal{A}$. See [18] for more practical methods to get Lipschitz extensions. In this section, whenever we put an overbar on a map, we imply the Lipschitz extension of the map.

Under Assumption 5, the observer we propose is of hybrid-type, and has the form

$$\dot{\hat{x}}(t) = \bar{f}_q(\hat{x}(t)) + \bar{g}_q(\hat{x}(t))u(t), \qquad t \in [t_{q-1}, t_q), \tag{18a}$$

$$\hat{x}(t_q) = \bar{p}_q(\hat{x}(t_q^-)), \qquad t = t_q, \quad q \in \mathbb{N}, \tag{18b}$$

$$\hat{x}(t_{im}^\star) = \hat{x}^\star(t_{im}^\star), \qquad t_{im}^\star = t_{im} + T_{\mathsf{comp}}, \quad i \in \mathbb{N}, \tag{18c}$$

with an initial condition $\hat{x}(t_0) \in \mathcal{X}$, where $\bar{f}_q$, $\bar{g}_q$, and $\bar{p}_q$ are Lipschitz extensions of f_q, g_q, and p_q from $\mathcal{X}$, respectively. Let us denote the Lipschitz coefficients of $\bar{f}_q$, $\bar{g}_q$, and $\bar{p}_q$ by C_f, C_g, and C_p, respectively, for all $q \in \mathbb{N}$. Note that, since the plant state $x(t)$ remains in $\mathcal{X}$, we can also treat system (1) as if the vector fields f_q, g_q, and p_q in (1) are replaced with their Lipschitz extensions, so that the plant (1) and the observer (18) are both globally Lipschitz. Then, it is seen that the observer equations (18a) and (18b) are just a copy of system (1) without any error correction. Instead, the estimation error is corrected through the vector $\hat{x}^\star$ at each $t_{im}^\star$, $i = 1, 2, \cdots$. The estimate $\hat{x}^\star$ is computed by an estimation algorithm to be proposed, and the computation of $\hat{x}^\star(t_{im}^\star)$ begins at every t_{im} (i.e., right after every m-th switch occurs). The delay time T_{comp} represents an upper bound of the time required for the computation, and we implicitly assume that $t_{im} + T_{\mathsf{comp}} < t_{(i+1)m}$ and that $t_{im}^\star$ does not coincide with any switching instance t_q.

Now we propose an estimation algorithm for $\hat{x}^\star(t_{im}^\star)$, which guarantees that $\lim_{t\to\infty} |\hat{x}(t) - x(t)| = 0$.

5.1. *Estimation Algorithm*

The proposed estimation algorithm is based on the representation (8) of system (1). It consists of four steps; (i) running ξ_q-observers, (ii) running w_q-observers, (iii) taking inverse of a map, and (iv) performing a catch-up process to compensate for computational delay T_{comp}. We discuss each of these steps in this subsection one by one. The algorithm begins at every time t_{im}, $i \in \mathbb{N}$, and processes the past data (i.e., $u_{[t_{(i-1)m}, t_{im}]}$, $y_{[t_{(i-1)m}, t_{im}]}$, $\sigma_{[t_{(i-1)m}, t_{im}]}$, and $\hat{x}^\star(t_q)$, $q = (i-1)m, \cdots, im-1$) that is stored in the memory.

For notational simplicity, we denote the mode $q = (i-1)m + k$ just by $q = k$ (as if $i = 1$) in this subsection. Therefore, when we say $q = m + 1$ it should be interpreted as $q = im + 1$.

5.1.1. ξ_q-observer and the back-and-forth technique. The ξ_q-observer means an observer for the small-time uniformly observable system (8a) and (8b). Instead of presenting a particular form of observer, we display the required minimal property of the observer (in Assumption 6) that will be used in our algorithm. In this way, we incorporate many observer design techniques available in the literature into the proposed algorithm. For this, we suppose that an observer for (8a) and (8b) is generally written as (here the superscript 'f' indicates 'forward' whose meaning will become clear soon)

$$\begin{aligned} \dot{\zeta}_q^{\mathsf{f}} &= \Sigma_q^{\mathsf{f}}(\zeta_q^{\mathsf{f}}, y, \upsilon), \qquad t \in [t_{q-1}, t_q), \quad \zeta_q^{\mathsf{f}} \in \mathbb{R}^{n_{\zeta_q}} \\ \hat{\xi}_q^{\mathsf{f}} &= \Upsilon_q(\zeta_q^{\mathsf{f}}, y, \upsilon), \qquad \zeta_q^{\mathsf{f}}(t_{q-1}) = \Lambda_q(\hat{x}^\star(t_{q-1}), \upsilon(t_{q-1})) \end{aligned} \tag{19}$$

where $\hat{\xi}_q^{\mathsf{f}}$ is the estimate of the state ξ_q, and υ represents the input u and its derivatives[‖]. The dynamics of the state $\hat{x}^\star$ will be discussed in detail when we introduce the catch-up process, and for now, let us suppose that $\hat{x}^\star(t_{q-1}) = \hat{x}(t_{q-1})$ where $\hat{x}$ is the state of (18). The maps Λ_q and Υ_q have the property that

$$\lambda_{\langle \nu_q \rangle}^q(x) = \Upsilon_q(\Lambda_q(x, \upsilon), H_q(x), \upsilon) \quad \text{for all admissible } \upsilon \text{ and all } x \in \mathcal{X}. \tag{20}$$

It can be interpreted that $\zeta_q^{\mathsf{f}}(t)$ is an estimate of $\Lambda_q(x(t), \upsilon(t))$. Since $\mathcal{X}$ is compact (and with the boundedness of υ), we assume, without loss of generality, that $\Upsilon_q(\cdot, y, \upsilon)$ and $\Lambda_q(\cdot, \upsilon)$ are globally

[‖]The reason for introducing the derivatives of the input u through υ is to incorporate the high-gain observers, studied in, e.g., [7, 25], where the input derivatives take part in the change of coordinates. When υ contains the derivatives of u as well as the input u itself, Assumption 5.1 is strengthened by adding that $\upsilon(t)$ is uniformly bounded.

DOI: 10.1002/rnc

Lipschitz uniformly w.r.t. y and υ (otherwise, take the Lipschitz extension). Let their Lipschitz coefficients be C_Υ and C_Λ for all $q \in \mathbb{N}$.

When system (1) has single input and single output ($n_u = n_y = 1$), the system (8a) and (8b) can take the form of (9), and thus, a particular example of the observer (19) is the high-gain observer studied in [8]. This corresponds to the case that $\hat{\xi}_q^{\mathsf{f}} = \Upsilon_q(\zeta_q^{\mathsf{f}}, y, \upsilon) = \zeta_q^{\mathsf{f}}$ and

$$\begin{aligned} \dot{\hat{\xi}}_q^{\mathsf{f}} &= \bar{F}_q(\hat{\xi}_q^{\mathsf{f}}) + \bar{G}_q(\hat{\xi}_q^{\mathsf{f}})u + K_q^{\mathsf{f}} \cdot (y - \bar{H}_q(\hat{\xi}_q^{\mathsf{f}})), \qquad t \in [t_{q-1}, t_q), \\ \hat{\xi}_q^{\mathsf{f}}(t_{q-1}) &= \bar{\lambda}^q_{\langle \nu_q \rangle}(\hat{x}^\star(t_{q-1})), \end{aligned} \tag{21}$$

where K_q^{f} is a constant injection gain, $\bar{\lambda}^q_{\langle \nu_q \rangle}$ is a Lipschitz extension of $\lambda^q_{\langle \nu_q \rangle}$ from $\mathcal{X}$, and $\bar{F}_q$, $\bar{G}_q$, and $\bar{H}_q$ are Lipschitz extensions of F_q, G_q, and H_q from $\Xi_q = \lambda^q_{\langle \nu_q \rangle}(\mathcal{X})$, respectively.

Now we state the property required for the proposed algorithm.

Assumption 6 (High-gain observer property)
For any constants $b > 0$ and $\delta > 0$, there exist an (forward) observer (19) and a class-$\mathcal{KL}$ function β_q^{f} satisfying

1. $|\zeta_q^{\mathsf{f}}(t) - \Lambda_q(x(t), \upsilon(t))| \le \beta_q^{\mathsf{f}}\left(|\zeta_q^{\mathsf{f}}(t_{q-1}) - \Lambda_q(x(t_{q-1}), \upsilon(t_{q-1}))|, t - t_{q-1}\right)$ for $t \in [t_{q-1}, t_q)$,
2. $\beta_q^{\mathsf{f}}(a, t) < \delta a$ for all $a > 0$ and $b \le t \le \tau_q$. $\lhd$

In the assumption, the first item implies that the observer (19) is an asymptotic observer while the second item states that the convergence rate can be made arbitrarily fast, which are characteristics of the high-gain observers. In fact, most of the nonlinear observer designs in the literature, such as [15, 8, 7, 25, 20], give rise to exponential functions that play the role of β_q^{f}; that is, $\beta_q^{\mathsf{f}}(a, t) = ke^{-\alpha t}a$ and, by choosing the observer gains appropriately, the constant α can be made arbitrarily large while the constant k increases with the polynomial order of α [23], which satisfies Assumption 6.2 with α such that $ke^{-\alpha b} < \delta$. This is because the designs are based on a quadratic error Lyapunov function in particular coordinates, so that the error satisfies an exponential convergence property in those coordinates.

Although the observer (19) can provide quite a good estimate $\hat{\xi}_q(t)$ after a relatively short transient period beginning at t_{q-1}, the proposed estimation algorithm will require a “good” estimate of $\xi_q(t)$ for the *entire* interval $[t_{q-1}, t_q)$ (the reason for this requirement will become clear when we discuss w_q-observer shortly). The task of obtaining good estimates for the whole interval is not achieved solely by the forward observer even though a high observer gain is used. (This point has been well studied in [23].) To solve this problem, we propose a ‘back-and-forth observation technique’ here. First, note that the solution $\xi_q(t)$ on the interval $[t_{q-1}, t_q)$ satisfies the following backward differential equation: with $\xi_q^{\mathsf{b}}(t) := \xi_q(t_q - t)$ and $\xi_q^{\mathsf{b}}(0) = \xi_q(t_q^-)$,

$$\dot{\xi}_q^{\mathsf{b}} = -F_q(\xi_q^{\mathsf{b}}) - G_q(\xi_q^{\mathsf{b}})u(t_q - t), \quad y(t_q - t) = H_q(\xi_q^{\mathsf{b}}), \quad t \in (0, \tau_q]. \tag{22}$$

Then, by slightly modifying the forward observer (19), we may design a backward observer for (22) as follows:

$$\begin{aligned} \dot{\zeta}_q^{\mathsf{b}} &= \Sigma_q^{\mathsf{b}}(\zeta_q^{\mathsf{b}}, y(t_q - t), \upsilon^{\mathsf{b}}(t_q - t)), \qquad t \in (0, \tau_q], \quad \zeta_q^{\mathsf{b}} \in \mathbb{R}^{n_{\zeta_q}} \\ \hat{\xi}_q^{\mathsf{b}} &= \Upsilon_q(\zeta_q^{\mathsf{b}}, y(t_q - t), \upsilon(t_q - t)), \qquad \zeta_q^{\mathsf{b}}(0) = \zeta_q^{\mathsf{f}}(t_q^-) \end{aligned} \tag{23}$$

where the vector υ^{b} contains $u(t_q - t)$ and its derivatives (which differs from υ by the signs of certain components; odd number of differentiation yields negative sign). It should be noted that the initial condition of ζ_q^{b} is set to be the final value of the forward observer (19). This implies that the backward observer runs after the forward observer has been executed. A backward observer, for example, could be obtained by a simple modification of (21) as follows:

$$\begin{aligned} \dot{\hat{\xi}}_q^{\mathsf{b}} &= -\bar{F}_q(\hat{\xi}_q^{\mathsf{b}}) - \bar{G}_q(\hat{\xi}_q^{\mathsf{b}})u(t_q - t) - K_q^{\mathsf{b}} \cdot (y(t_q - t) - \bar{H}_q(\hat{\xi}_q^{\mathsf{b}})), \\ \hat{\xi}_q^{\mathsf{b}}(0) &= \hat{\xi}_q^{\mathsf{f}}(t_q^-), \qquad t \in (0, \tau_q], \end{aligned} \tag{24}$$

DOI: 10.1002/rnc

where the injection gain K_q^{b} may not be related to K_q^{f} and needs to be redesigned. For the backward observer (23), we also assume the following:

Assumption 6′ (Backward high-gain observer property)
For any constants $b > 0$ and $\delta > 0$, there exist an (backward) observer (23) and a class-$\mathcal{KL}$ function β_q^{b} satisfying

1. $|\zeta_q^{\mathsf{b}}(t) - \Lambda_q(x(t_q - t), \upsilon(t_q - t))| \le \beta_q^{\mathsf{b}}\left(|\zeta_q^{\mathsf{b}}(0) - \Lambda_q(x(t_q^-), \upsilon(t_q^-))|, t\right)$ for $t \in (0, \tau_q]$,
2. $\beta_q^{\mathsf{b}}(a, t) < \delta a$ for all $a > 0$ and $b \le t \le \tau_q$. $\lhd$

Once Assumption 6 holds for (19), this additional requirement is mild. For example, the designs of [15, 8, 7, 25, 20] readily satisfy this requirement (because the system structure does not change and only the signs of vector fields and signals are reversed).

The operation of ξ_q-observers is now summarized. Suppose, for now, that

$$\hat{x}^\star(t_{q-1}) = \hat{x}(t_{q-1}) \quad \text{for } q = 1, \cdots, m \tag{25}$$

are stored (we will remove (25) later in Section 5.1.4). When the m-th switch occurs at time t_m, the algorithm integrates the ξ_q-observer (19) using the stored data from the past, followed by integrating (23), for $q = 1, \cdots, m$ sequentially. The estimate $\hat{\xi}_q(t)$ is taken to be

$$\hat{\xi}_q(t) := \begin{cases} \hat{\xi}_q^{\mathsf{b}}(t_q - t), & t \in [t_{q-1}, t_{q-1} + \frac{\tau_q}{2}), \\ \hat{\xi}_q^{\mathsf{f}}(t), & t \in [t_{q-1} + \frac{\tau_q}{2}, t_q). \end{cases} \tag{26}$$

Since the estimation transients have been removed in (26), arbitrarily small estimation error can now be obtained for the entire interval $[t_{q-1}, t_q)$. Indeed, suppose that, with $b = \tau_q/2$ and a given $\delta \in (0, 1)$, the observers (19) and (23) are suitably designed under Assumptions 6 and 6′. Then, with $\tilde{\xi}_q := \hat{\xi}_q - \xi_q$,

$$\begin{aligned}
\sup_{t \in [t_{q-1} + \frac{\tau_q}{2}, t_q)} |\tilde{\xi}_q(t)| &= \sup_{t \in [t_{q-1} + \frac{\tau_q}{2}, t_q)} |\hat{\xi}_q^{\mathsf{f}}(t) - \xi_q(t)| && \text{(27a)}\\
&= \sup_{t \in [t_{q-1} + \frac{\tau_q}{2}, t_q)} |\Upsilon_q(\zeta_q^{\mathsf{f}}(t), y(t), \upsilon(t)) - \Upsilon_q(\Lambda_q(x(t), \upsilon(t)), y(t), \upsilon(t))| && \text{(27b)}\\
&\le \sup_{t \in [t_{q-1} + \frac{\tau_q}{2}, t_q)} C_\Upsilon \beta_q^{\mathsf{f}}(|\zeta_q^{\mathsf{f}}(t_{q-1}) - \Lambda_q(x(t_{q-1}), \upsilon(t_{q-1}))|, t - t_{q-1}) && \text{(27c)}\\
&\le C_\Upsilon \delta |\zeta_q^{\mathsf{f}}(t_{q-1}) - \Lambda_q(x(t_{q-1}), \upsilon(t_{q-1}))| && \text{(27d)}\\
&= \delta C_\Upsilon |\Lambda_q(\hat{x}^\star(t_{q-1}), \upsilon(t_{q-1})) - \Lambda_q(x(t_{q-1}), \upsilon(t_{q-1}))| && \text{(27e)}\\
&\le \delta C_\Upsilon C_\Lambda |\hat{x}^\star(t_{q-1}) - x(t_{q-1})|. && \text{(27f)}
\end{aligned}$$

And, similarly,

$$\begin{aligned}
\sup_{t \in [t_{q-1}, t_{q-1} + \frac{\tau_q}{2})} |\tilde{\xi}_q(t)| &= \sup_{t \in [t_{q-1}, t_{q-1} + \frac{\tau_q}{2})} |\hat{\xi}_q^{\mathsf{b}}(t_q - t) - \xi_q(t)| && \text{(28a)}\\
&= \sup_{t \in [t_{q-1}, t_{q-1} + \frac{\tau_q}{2})} |\Upsilon_q(\zeta_q^{\mathsf{b}}(t_q - t), y(t), \upsilon(t)) - \Upsilon_q(\Lambda_q(x(t), \upsilon(t)), y(t), \upsilon(t))| && \text{(28b)}\\
&\le \sup_{t \in [t_{q-1}, t_{q-1} + \frac{\tau_q}{2})} C_\Upsilon |\zeta_q^{\mathsf{b}}(t_q - t) - \Lambda_q(x(t), \upsilon(t))| && \text{(28c)}\\
&= \sup_{t \in (\frac{\tau_q}{2}, \tau_q]} C_\Upsilon |\zeta_q^{\mathsf{b}}(t) - \Lambda_q(x(t_q - t), \upsilon(t_q - t))| && \text{(28d)}\\
&\le \sup_{t \in (\frac{\tau_q}{2}, \tau_q]} C_\Upsilon \beta_q^{\mathsf{b}}(|\zeta_q^{\mathsf{b}}(0) - \Lambda_q(x(t_q^-), \upsilon(t_q^-))|, t) && \text{(28e)}\\
&\le \delta C_\Upsilon |\zeta_q^{\mathsf{f}}(t_q^-) - \Lambda_q(x(t_q^-), \upsilon(t_q^-))| && \text{(28f)}\\
&\le \delta C_\Upsilon \beta_q^{\mathsf{f}}(|\zeta_q^{\mathsf{f}}(t_{q-1}) - \Lambda_q(x(t_{q-1}), \upsilon(t_{q-1}))|, t_q - t_{q-1}) && \text{(28g)}\\
&\le \delta C_\Upsilon \delta C_\Upsilon C_\Lambda |\hat{x}^\star(t_{q-1}) - x(t_{q-1})|. && \text{(28h)}
\end{aligned}$$

DOI: 10.1002/rnc

Therefore, combining (27) and (28), the operation of ξ_q-observer for the mode q leads to

$$\sup_{t\in[t_{q-1},t_q)} |\tilde{\xi}_q(t)| \le \delta \cdot \max\{C_\Upsilon C_\Lambda, C_\Upsilon^2 C_\Lambda\} \cdot |\hat{x}^\star(t_{q-1}) - x(t_{q-1})|. \tag{29}$$

Let us define $M_{\Upsilon\Lambda} := \max\{C_\Upsilon C_\Lambda, C_\Upsilon^2 C_\Lambda\}$ for simplicity.

5.1.2. w_q-observer. The w_q-observer for $q = 2, \cdots, m$ is just a replication of (8c) and (8d), given by

$$\dot{\hat{w}}_q = \bar{F}_q^*(\hat{w}_q, \hat{\xi}_q) + \bar{G}_q^*(\hat{w}_q, \hat{\xi}_q)u, \tag{30a}$$

with the initial condition

$$\hat{w}_q(t_{q-1}) = \bar{R}_{q-1}^*(\hat{w}_{q-1}(t_{q-1}^-), \hat{\xi}_{q-1}(t_{q-1}^-), \hat{\xi}_q(t_{q-1})). \tag{30b}$$

and $\hat{w}_1 := 0$ for convenience. Letting $\Omega_q := \omega_{\langle l_q\rangle}^q(\mathcal{X})$, the vector fields $\bar{F}_q^*$ and $\bar{G}_q^*$ are Lipschitz extensions of F_q^* and G_q^* from the set $\Omega_q \times \Xi_q$, and $\bar{R}_{q-1}^*$ is Lipschitz extension of R_{q-1}^* from $\Omega_{q-1} \times \Xi_{q-1} \times \Xi_q$. Let C_{F^*}, C_{G^*}, and C_{R^*} be their Lipschitz coefficients, respectively. Again we emphasize that the role of w_q-observer is not to reduce the error $\tilde{w}_q(t) := \hat{w}_q(t) - w_q(t)$, but to deliver the estimates $\hat{\xi}_{q-1}(t_{q-1}^-)$ and $\hat{w}_{q-1}(t_{q-1}^-)$, that are obtained from the previously active mode, to the next mode through $\hat{w}_q(t)$ along the copy of system dynamics (30a). In order for the information not to be corrupted too much during the delivery, we require that $\hat{\xi}_q(t)$ remains near its true value $\xi_q(t)$ for the entire interval $[t_{q-1}, t_q)$. This is the reason why we required the reduction in estimation error $\tilde{\xi}_q(t)$ for the entire interval in the previous subsection. Here, we emphasize that, shrinking the transient period by increasing the observer gain may worsen the situation because of the peaking phenomenon [23]; that is, the peaking in $\hat{\xi}_q(t)$ may damage the delivery role of (30a).

The operation of w_q-observers (30) for $q = 2, \cdots, m$ begins after the operation of ξ_q-observers, and the error caused by the operation of w_q-observers will be analyzed in Section 5.2.

5.1.3. Inversion. After the integrations of ξ_q-observers and w_q-observers, the estimates $\hat{\xi}_m(t_m^-) \in \mathbb{R}^{k_m}$ and $\hat{w}_m(t_m^-) \in \mathbb{R}^{l_m}$ become available. Since, by Assumption 5.4, the map $\mathrm{col}(\vartheta_{\langle k_m\rangle}^m, \omega_{\langle l_m\rangle}^m)$ is a diffeomorphism on $\mathcal{X}$, let its inverse on the range be given, referring to (7), by $x = \Phi(\chi_m(\xi_m), w_m)$. Now define a map $\Psi(\xi_m, w_m)$ as a Lipschitz extension of the map $p_m \circ \Phi(\chi_m(\xi_m), w_m)$ from $\vartheta_{\langle k_m\rangle}^m(\mathcal{X}) \times \omega_{\langle l_m\rangle}^m(\mathcal{X})$ where p_m is the jump map of (1). Then, we set the estimate $\hat{x}^\star(t_m)$ for the plant state $x(t_m)$ as

$$\hat{x}^\star(t_m) = \Psi(\hat{\xi}_m(t_m^-), \hat{w}_m(t_m^-))). \tag{31}$$

Let the Lipschitz coefficient of Ψ be C_Ψ.

5.1.4. Catch-up process. The algorithm to compute $\hat{x}^\star(t_m)$ begins at time $t = t_m$, and takes some time until the value of $\hat{x}^\star(t_m)$ is calculated. Therefore, in order to update the observer state $\hat{x}$ of (18) by $\hat{x}^\star$, we should translate $\hat{x}^\star(t)$ from t_m to the current time. This can be done by integrating a copy of (18a) and (18b) from time t_m onwards with the initial condition $\hat{x}^\star(t_m)$. This separate integration should be sufficiently fast, compared to the on-line observer (18) running synchronously in real time, so that the expressions resulting from both the integrations coincide after some time $t_m^\star = t_m + T_{\mathsf{comp}}$. We call this procedure as a 'catch-up process,' which is summarized as follows. When $\hat{x}^\star(t_m)$ is obtained by inversion from (31), integration of the following equation starts:

$$\dot{\hat{x}}^\star(t) = \bar{f}_q(\hat{x}^\star(t)) + \bar{g}_q(\hat{x}^\star(t))u(t), \qquad t \in [t_{q-1}, t_q), \tag{32a}$$
$$\hat{x}^\star(t_q) = \bar{p}_q(\hat{x}^\star(t_q^-)), \tag{32b}$$

for $q = m+1, \cdots, 2m$ with the initial condition $\hat{x}^\star(t_m)$. This integration is performed fast during $[t_m, t_m + T_{\mathsf{comp}})$, and (18c) is updated at $t = t_m + T_{\mathsf{comp}}$. Then, the integration continues synchronously with real time until t_{2m}^-, or just store $\hat{x}^\star(t) = \hat{x}(t)$ for rest of the time (which is possible because both (32) and (18) yield the same result). Reason for the integration after

DOI: 10.1002/rnc

$t_m + T_{\mathsf{comp}}$ is that the knowledge of $\hat{x}^\star(t_q)$ for $q = m, \cdots, 2m-1$ will be used for ξ_q-observers in the next cycle of the algorithm that starts at $t = t_{2m}$.

Now we can remove the statement (25) because we have specified $\hat{x}^\star(t_q)$ here.

5.2. Convergence of Estimation Error

We first look at the error $\tilde{x}^\star(t) := \hat{x}^\star(t) - x(t)$, and then $\tilde{x}(t) := \hat{x}(t) - x(t)$. It can be thought that the state $\hat{x}^\star(t)$ obeys (32) and is updated at every $t = t_{im}$, in computer time, $i \in \mathbb{N}$, by the inverse relation (31). Since there is no error correction between t_{im} and $t_{(i+1)m}$, the estimation error $\tilde{x}^\star(t)$ may increase during the interval. However, its growth is limited and can be computed by Gronwall-Bellman's inequality and the Lipschitz property of (32) as follows: for any $t^\dagger$ and $t^\ddagger$ such that $t_{im} \le t^\dagger < t^\ddagger \le t_{(i+1)m}$,

$$|\tilde{x}^\star(t^\ddagger)| \le (C_p)^\rho \exp\left((C_f + C_g C_u)(t^\ddagger - t^\dagger)\right) |\tilde{x}^\star(t^\dagger)| =: M(t^\ddagger, t^\dagger)|\tilde{x}^\star(t^\dagger)| \tag{33}$$

where ρ is the number of switches in the interval $(t^\dagger, t^\ddagger]$, and for convenience, let

$$M_k^i := M(t_{im+k}, t_{im}) = (C_p)^k \exp\left((C_f + C_g C_u)\sum_{j=1}^{k} \tau_{j+im}\right),$$

and $M_0^i := 1$ for convenience. This in turn implies from (29) that, for $q = im+1, \ldots, im+m$,

$$\sup_{t\in[t_{q-1},t_q)} |\tilde{\xi}_q(t)| \le \delta M_{\Upsilon\Lambda} |\tilde{x}^\star(t_{q-1})| \le \delta M_{\Upsilon\Lambda} M_{q-1-im}^i |\tilde{x}^\star(t_{im})|. \tag{34}$$

It then follows from (30a) that $|\dot{\tilde{w}}_q| \le M_*(|\tilde{w}_q| + |\tilde{\xi}_q|)$, where $M_* := C_{F^*} + C_{G^*}C_u$. For $q = im+2, \ldots, im+m$, this leads to,

$$\begin{aligned} |\tilde{w}_q(t_q^-)| &\le e^{M_*\tau_q}|\tilde{w}_q(t_{q-1})| + (e^{M_*\tau_q} - 1) \sup_{t\in[t_{q-1},t_q)} |\tilde{\xi}_q(t)| \\ &\le e^{M_*\tau_q}|\tilde{w}_q(t_{q-1})| + (e^{M_*\tau_q} - 1)\delta M_{\Upsilon\Lambda} M_{q-1-im}^i |\tilde{x}^\star(t_{im})|. \end{aligned} \tag{35}$$

From (30b) and (34), with $\tilde{w}_{im+1} := 0$,

$$\begin{aligned} |\tilde{w}_q(t_{q-1})| &\le C_{R^*}|\tilde{\xi}_q(t_{q-1})| + C_{R^*}|\tilde{\xi}_{q-1}(t_{q-1}^-)| + C_{R^*}|\tilde{w}_{q-1}(t_{q-1}^-)| \\ &\le \delta C_{R^*} M_{\Upsilon\Lambda}(M_{q-1-im}^i + M_{q-2-im}^i)|\tilde{x}^\star(t_{im})| + C_{R^*}|\tilde{w}_{q-1}(t_{q-1}^-)|. \end{aligned} \tag{36}$$

Putting (36) into (35), we obtain

$$|\tilde{w}_q(t_q^-)| \le C_{R^*} e^{M_*\tau_q} |\tilde{w}_{q-1}(t_{q-1}^-)| + \delta \tilde{N}_{q-1-im}^i \, |\tilde{x}^\star(t_{im})|$$

where $\tilde{N}_{q-1-im}^i := (e^{M_*\tau_q} - 1)M_{\Upsilon\Lambda}M_{q-1-im}^i + e^{M_*\tau_q}C_{R^*}M_{\Upsilon\Lambda}(M_{q-1-im}^i + M_{q-2-im}^i)$. From this, it is not difficult to derive that, for $q = im+2, \ldots, im+m$,

$$|\tilde{w}_q(t_q^-)| \le \delta N_{q-1-im}^i |\tilde{x}^\star(t_{im})| \tag{37}$$

where

$$N_{q-1-im}^i := \tilde{N}_{q-1-im}^i + \sum_{j=2}^{q-1-im} (C_{R^*})^{q-j-im} \tilde{N}_j^i \exp\left(M_* \sum_{k=q+2-j}^{q} \tau_k\right).$$

So far, we have obtained $|\tilde{\xi}_{(i+1)m}(t_{(i+1)m}^-)| \le \delta M_{\Upsilon\Lambda} M_{m-1}^i |\tilde{x}^\star(t_{im})|$ and $|\tilde{w}_{(i+1)m}(t_{(i+1)m}^-)| \le \delta N_{m-1}^i |\tilde{x}^\star(t_{im})|$ from (34) and (37), which, by (31), finally lead to,

$$\begin{aligned} |\tilde{x}^\star(t_{(i+1)m})| &\le |\Psi(\hat{\xi}_{(i+1)m}(t_{(i+1)m}^-), \hat{w}_{(i+1)m}(t_{(i+1)m}^-)) - \Psi(\xi_{(i+1)m}(t_{(i+1)m}^-), w_{(i+1)m}(t_{(i+1)m}^-))| \\ &\le C_\Psi(|\tilde{\xi}_{(i+1)m}(t_{(i+1)m}^-)| + |\tilde{w}_{(i+1)m}(t_{(i+1)m}^-)|) \\ &\le \delta C_\Psi(M_{\Upsilon\Lambda}M_{m-1}^i + N_{m-1}^i)|\tilde{x}^\star(t_{im})|. \end{aligned}$$

Suppose that δ is chosen such that

$$\delta C_\Psi (M_{\Upsilon\Lambda} M^i_{m-1} + N^i_{m-1}) \leq \gamma < 1 \tag{38}$$

with a constant γ. Then, $\lim_{i\to\infty} |\tilde{x}^\star(t_{im})| = 0$. This implies that $\lim_{t\to\infty} |\tilde{x}^\star(t)| = 0$ because $t_{(i+1)m} - t_{im} \leq mD$ and

$$\begin{aligned} \sup_{t\in[t_{im}, t_{(i+1)m})} |\tilde{x}^\star(t)| &\leq \sup_{t\in[t_{im}, t_{(i+1)m})} M(t, t_{im}) |\tilde{x}^\star(t_{im})| \\ &\leq (C_p)^{m-1} \exp((C_f + C_g C_u) mD) |\tilde{x}^\star(t_{im})| \end{aligned} \tag{39}$$

for all $i \in \mathbb{N}$, due to (33) and Assumption 5.2.

Finally, since $\lim_{t\to\infty} |\tilde{x}^\star(t)| = 0$ and $\tilde{x}^\star(t^\star_{im}) = \tilde{x}(t^\star_{im})$, we have that $\lim_{i\to\infty} |\tilde{x}^\star(t^\star_{im})| = \lim_{i\to\infty} |\tilde{x}(t^\star_{im})| = 0$. Therefore,

$$\begin{aligned} \sup_{t\in[t^\star_{im}, t^\star_{(i+1)m})} |\tilde{x}(t)| &\leq \sup_{t\in[t^\star_{im}, t^\star_{(i+1)m})} M(t, t^\star_{im}) |\tilde{x}(t^\star_{im})| \\ &\leq (C_p)^{m} \exp((C_f + C_g C_u) mD) |\tilde{x}(t^\star_{im})| \end{aligned} \tag{40}$$

which ensures that $\lim_{t\to\infty} |\tilde{x}(t)| = 0$.

Summarizing the discussions so far, we arrive at the following theorem.

Theorem 4 (Asymptotic observer)
Under Assumptions 5, 6, and 6′, system (1) admits an asymptotic observer; and the observer gains are designed according to the criteria given in (38) under Assumptions 6 and 6′. ◁

Convergence rate of the proposed observer can be increased by reducing γ in (38), but is limited by the fact that the estimation error cannot be reduced until the first update time $t^\star_m$. This is natural because the observer is based on the property of large-time, and not small-time, observability.

5.3. *Simulation Results of Example 1*

The proposed observer has been implemented and simulated for Example 1, and the results are depicted in Fig. 1. For this, the ξ_q-observers (21) and (24), which are linear for this example, are employed with $K^{\mathsf{f}}_1 = \mathrm{col}(2, 1)$, $K^{\mathsf{f}}_2 = K^{\mathsf{f}}_3 = 2$, $K^{\mathsf{b}}_1 = \mathrm{col}(-2, 1)$, and $K^{\mathsf{b}}_2 = K^{\mathsf{b}}_3 = -2$.

For the simulation, we did not apply Lipschitz extensions to any maps or functions. This is justified by regarding that the region $\mathcal{X}$ for large-time observability is so large that both the plant trajectory and the observer trajectory remain in the inactive region $\mathcal{X}$ of the Lipschitz extension. As a matter of fact, Lipschitz extension becomes critical when the plant trajectory is operating near the boundary of $\mathcal{X}$, or the observer gain is so high that the initial peaking transient often goes beyond the set $\mathcal{X}$.

MATLAB source code is available upon request from the authors.

6. DISCUSSIONS AND CONCLUDING REMARKS

6.1. *Discussions*

1. Large-time observability for switched linear systems is studied in our previous work [24], where a necessary and sufficient condition for it is presented. It turns out that the switching times also affect the observability because it has been shown that there are singular switching times which destroy observability even though the same mode sequence with different switching times ensures observability. The conditions in this paper are independent of switching times, which is one evidence of sufficiency. In [24], one can also find a condition independent of switching times, which is of course a sufficient condition, but it can be seen that the linear version of the proposed condition in Section 4.3 is still stronger than [24]. It is actually due to the restrictive structure of w_q-dynamics in this paper.

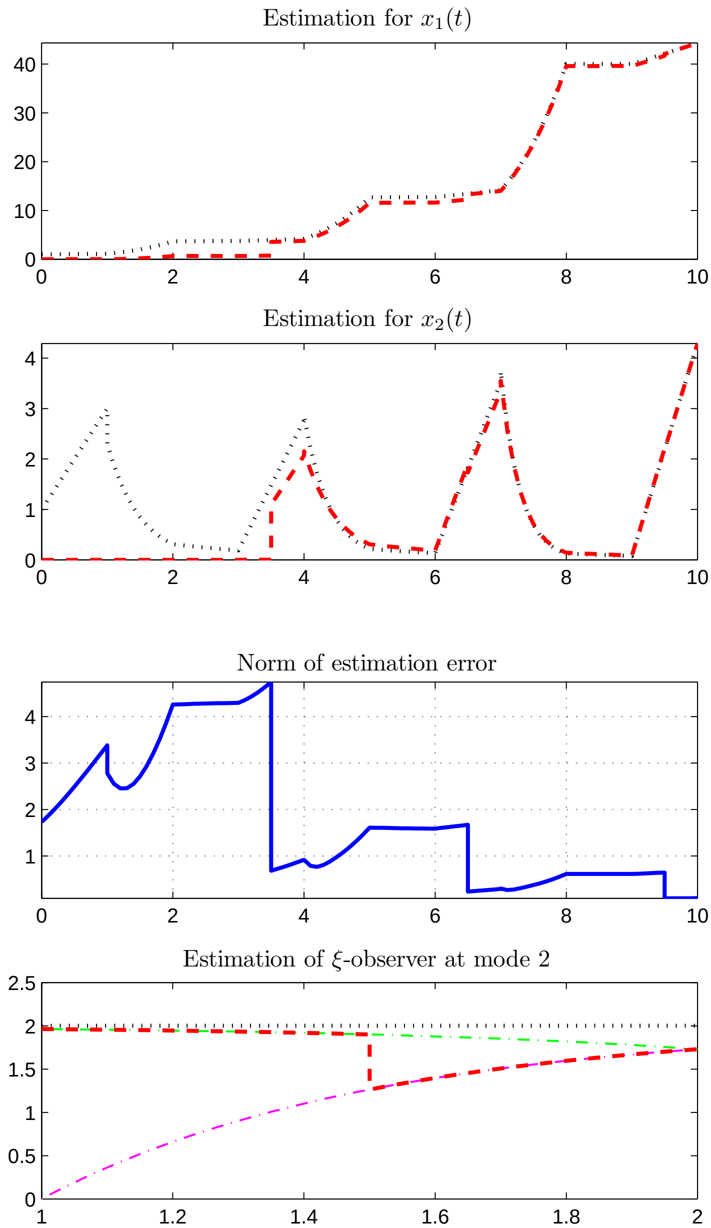

Figure 1. The first two figures show the plant state $x(t)$ (dotted black) and its estimate $\hat{x}(t)$ (dashed red), initiated from $x(0) = \mathrm{col}(1, 1, 1)$ and $\hat{x}(0) = \mathrm{col}(0, 0, 0)$, respectively. With $T_{\mathsf{comp}} = 0.5$, the update (18c) occurs at 3.5, 6.5, and 9.5 seconds while the mode switches after every second. This is clearly seen from the third figure, which plots $|\hat{x}(t) - x(t)|$. The convergence of error to zero is also observed. The fourth plot illustrates the operation of back-and-forth ξ_q-observer for mode 2 over the interval $[1, 2)$. In this mode, ξ_2 is a scalar variable (dotted black). It is observed that the estimate $\hat{\xi}_2$ (dashed red) is obtained from $\hat{\xi}_2^{\mathsf{f}}$ of the forward observer (dashdot magenta) over the interval $[1.5, 2)$ and $\hat{\xi}_2^{\mathsf{b}}$ of the backward observer (dashdot green) over the interval $[1, 1.5)$.

2. In the convergence analysis of state estimation error in Section 5.2, the constants M^i_{m-1} and N^i_{m-1} of (38) are dependent on the switching intervals τ_q, but since they are nondecreasing w.r.t. τ_q, one can make the condition (38) independent of τ_q by setting all $\tau_q = D$. In this way, δ can be chosen off-line and all ξ_q-observers (i.e., observer gains) are designed before

DOI: 10.1002/rnc

the actual operation. Else, δ is chosen on-line and ξ_q-observers are designed whenever the estimation algorithm is executed. The latter case tends to yield lower gain than the former.

3. The convergence analysis in Section 5.2 suggests that the mode sequence $\{1, 2, \cdots, m\}$ need not be the same and repeat. But it is important that, within a sequence of every m modes, there should be a subsequence of modes that guarantees large-time uniform observability. Then, at each execution of the estimation algorithm, the value δ is computed on-line for (38) which guarantees error convergence. This relaxes Assumption 5.4.
4. In an ideal case of no uncertainty and no disturbance, Assumption 5.2 can also be relaxed in theory. The idea is to reuse the past data in the back-and-forth ξ_q-observers. Looking at (19) and (23) with (29), it is seen that one more forward and backward operation with the initial value $\xi_q^{\mathsf{f}}(t_{q-1}) = \xi_q^{\mathsf{b}}(\tau_q)$ results in further reduction of the error. Repeating the back-and-forth operation, followed by longer catch-up process, one can update (18c) at every period of D (if computation is fast enough) even though the actual switching does not occur. This approach is, of course, not practical because the estimation is based on outdated data and does not reflect current information.

6.2. Concluding Remarks

This paper presented a sufficient condition for large-time uniform observability of switched nonlinear systems. Compared to the existing literature on linear systems, this condition is independent of switching times and depends primarily on the mode sequence determined by the switching signal. The proof reveals how the partial information available from each mode can be combined to recover the state. The observer, based on the proposed sufficient condition, generates an estimate that converges to the actual state of the system.

One limitation of the current research is that the switching signal $\sigma(t)$ is known, which is sometimes not the case. Mode detection as well as the detection of switching times is one of the future research directions.

Also, there are potential applications for state estimators of switched nonlinear systems, e.g. [6], and exploring them in the context of real-time systems using our design techniques remains a future work.

ACKNOWLEDGEMENT

The authors are grateful to Daniel Liberzon for helpful discussions on this paper. This work was supported by National Research Foundation Grant funded by the Korean Government (NRF-2011-220-D00043).

APPENDIX

The following lemma is frequently used in the paper.

Lemma 1
Consider a codistribution $\mathcal{W}$ generated by exact one-forms, i.e., $\mathcal{W} = \mathrm{span}\,\{d\lambda_1, \ldots, d\lambda_k\}$ where $\lambda_1, \ldots, \lambda_k$, $1 \le k \le n$, are potential coordinate functions defined on a set $\mathcal{X} \subset \mathbb{R}^n$.

1. If the codistribution $\mathcal{W}$ is invariant w.r.t. a smooth vector field f, i.e.,

$$L_f \mathcal{W} \subset \mathcal{W} \quad \text{on } \mathcal{X},$$

then there exists a smooth vector field $F : \lambda_{\langle k \rangle}(\mathcal{X}) \to \mathbb{R}^k$ such that $\frac{\partial \lambda_{\langle k \rangle}}{\partial x} \cdot f(x) = F(\lambda_{\langle k \rangle}(x))$ for every $x \in \mathcal{X}$.

2. If a smooth function $h : \mathcal{X} \to \mathbb{R}$ satisfies

$$dh \in \mathcal{W} \quad \text{on } \mathcal{X},$$

then there exists a smooth function $H : \lambda_{\langle k\rangle}(\mathcal{X}) \to \mathbb{R}$ such that $h(x) = H(\lambda_{\langle k\rangle}(x))$ for every $x \in \mathcal{X}$.

3. Let $\mathcal{W}'$ be another codistribution such that $\dim(\mathcal{W} + \mathcal{W}') = k + r$ on $\mathcal{X}$, and suppose that there are functions $\mu_1, \ldots, \mu_r$ such that $\mathcal{W} + \mathcal{W}' = \mathrm{span}\{d\lambda_i, d\mu_j : i = 1, \ldots, k, j = 1, \ldots, r\}$ and $\{\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_r\}$ are potential coordinate functions on $\mathcal{X}$. If a smooth map $p : \mathcal{X} \to \mathbb{R}^n$ satisfies
$$p_*(\ker \mathcal{W} \cap \ker \mathcal{W}') \subset \ker \mathcal{W}$$
on $\mathcal{X}$, then there exists a smooth map $P : \lambda_{\langle k\rangle}(\mathcal{X}) \times \mu_{\langle r\rangle}(\mathcal{X}) \to \mathbb{R}^k$ such that $\lambda_{\langle k\rangle}(p(x)) = P(\lambda_{\langle k\rangle}(x), \mu_{\langle r\rangle}(x))$ whenever x and $p(x)$ are contained in $\mathcal{X}$. $\triangleleft$

Proof

Since λ_i, $i = 1, \ldots, k$, are potential coordinate functions on $\mathcal{X}$, we can find $\lambda_{k+1}, \ldots, \lambda_n$ such that $\lambda_{\langle n\rangle}(x)$ becomes a diffeomorphism on $\mathcal{X}$. Let $z = \lambda_{\langle n\rangle}(x)$. In the z-coordinates, it is seen that $\mathcal{W} = \mathrm{span}\{dz_1, \ldots, dz_k\}$, and thus,
$$\ker \mathcal{W} = \mathrm{span}\left\{\frac{\partial}{\partial z_{k+1}}, \cdots, \frac{\partial}{\partial z_n}\right\}. \tag{41}$$
Also, the vector field $f(x)$ is represented in z-coordinates as:
$$\frac{\partial \lambda_{\langle n\rangle}}{\partial x}\bigg|_{x=\lambda_{\langle n\rangle}^{-1}(z)} \cdot f(\lambda_{\langle n\rangle}^{-1}(z)) =: \bar{f}(z) = \begin{bmatrix} \bar{f}_a(z_a, z_b) \\ \bar{f}_b(z_a, z_b) \end{bmatrix},$$
where $z_a = [z_1, \ldots, z_k]^\top$, $z_b = [z_{k+1}, \ldots, z_n]^\top$, $\bar{f}_a(z) \in \mathbb{R}^k$, and $\bar{f}_b(z) \in \mathbb{R}^{n-k}$. Then, since the codistribution $\mathcal{W}$ is invariant w.r.t. f, the distribution $\ker \mathcal{W}$ is also invariant w.r.t. f on $\mathcal{X}$.** Since $\ker \mathcal{W}$ is invariant w.r.t. f on $\mathcal{X}$, or equivalently w.r.t. $\bar{f}$ on $\lambda_{\langle n\rangle}(\mathcal{X})$, it follows that
$$\left[\bar{f}, \frac{\partial}{\partial z_i}\right] = -\sum_{j=1}^{n} \frac{\partial \bar{f}_j}{\partial z_i} \frac{\partial}{\partial z_j} \in \ker \mathcal{W}, \qquad i = k+1, \ldots, n.$$
Hence,
$$\frac{\partial \bar{f}_j}{\partial z_i} = 0, \quad {}^\forall j = 1, \ldots, k, \ \ i = k+1, \ldots, n.$$
This implies $\bar{f}_a(z_a, z_b) = \bar{f}_a(z_a)$. Taking $F = \bar{f}_a$ proves item 1.

For proving item 2, the function h is represented in the z-coordinates as $\bar{h}(z) = h \circ \lambda_{\langle n\rangle}^{-1}(z)$. Since
$$d\bar{h} = \frac{\partial \bar{h}}{\partial z_1} dz_1 + \cdots + \frac{\partial \bar{h}}{\partial z_n} dz_n \in \mathcal{W}, \qquad \text{on } \lambda_{\langle n\rangle}(\mathcal{X}),$$
it is seen that $\frac{\partial \bar{h}}{\partial z_i} = 0$, $i = k+1, \ldots, n$. Taking $H = \bar{h}$ proves item 2.

For item 3, we find $\lambda_{k+r+1}, \ldots, \lambda_n$ such that
$$z = \lambda(x) := \mathrm{col}(\lambda_{\langle k\rangle}(x), \mu_{\langle r\rangle}(x), \lambda_{k+r+1}(x), \ldots, \lambda_n(x))$$
becomes a diffeomorphism on $\mathcal{X}$. Then, in the coordinates of z, we have that $\mathcal{W} = \mathrm{span}\{dz_1, \ldots, dz_k\}$ and $\mathcal{W} + \mathcal{W}' = \mathrm{span}\{dz_1, \ldots, dz_{k+r}\}$. This way $\ker \mathcal{W}$ and $\ker \mathcal{W} \cap \ker \mathcal{W}' = (\mathcal{W} + \mathcal{W}')^\perp$ can be equivalently written as $\mathrm{span}\{e_{k+1}, \ldots, e_n\}$ and $\mathrm{span}\{e_{k+r+1}, \ldots, e_n\}$, respectively, on $\lambda(\mathcal{X})$, where e_j is the elementary basis vector (i.e., all elements are zero except the j-th element which is one). With $\bar{p}(z) = \lambda \circ p \circ \lambda^{-1}(z)$, the

** Let $\sigma \in \mathcal{W}$ and $v \in \ker \mathcal{W}$. Then, $\sigma \cdot v = 0$ and $(L_f \sigma) \cdot v = 0$. By the equality $L_f(\sigma \cdot v) = (L_f \sigma) \cdot v + \sigma \cdot [f, v]$, it is seen that $\sigma \cdot [f, v] = 0$. See [16].

DOI: 10.1002/rnc

condition $p_*(\ker \mathcal{W} \cap \ker \mathcal{W}') \subset \ker \mathcal{W}$ implies that

$$\frac{\partial \bar{p}}{\partial z} e_j \in \operatorname{span}\{e_{k+1}, \dots, e_n\} \quad \text{for } j = k+r+1, \dots, n$$

on $\lambda(\mathcal{X})$. This implies that the first k functions of $\bar{p}$ do not depend on $z_{k+r+1}, \dots, z_n$, so that we obtain $P = \bar{p}_{\langle k \rangle}$. This completes the proof. $\square$

Proof of Theorem 3
By Assumption 4.(a).(1), system (1a) and (1c) is converted into (9) with $\xi_{q,j} = L_{f_q}^{j-1} h_q(x)$, $j = 1, \cdots, \nu_q$, on the set $\mathcal{X}$. Then, the same argument as in Section 4.1 (after Assumption 3) proves that Assumption 1 holds.

To see that Assumption 2 holds, we take $\theta_q := \vartheta^q_{\langle k_q \rangle}(x)$, $w_q := \omega^q_{\langle l_q \rangle}(x)$, and $\varsigma_q := \pi^q_{\langle n-k_q-l_q \rangle}(x)$ in which the existence of $\pi^q_{\langle n-k_q-l_q \rangle}$ on the set $\mathcal{X}$ follows from Assumption 4.(b). For simplicity, we denote $\lambda^q_j(x) := L_{f_q}^{j-1} h_q(x)$, so that $\xi_q = \lambda^q_{\langle \nu_q \rangle}(x)$. Here, because $d\vartheta^q_j \in \mathcal{O}_q$ (Assumption 4.(b)) and $\mathcal{O}_q$ is generated by the differentials of $\lambda^q_1, \cdots, \lambda^q_{\nu_q}$ that are potential coordinate functions on $\mathcal{X}$ (by Assumption 4.(a)), each ϑ^q_j is a function of $\lambda^q_{\langle \nu_q \rangle}$ (by Lemma 1.2) so that the map χ_q of (6) exists.

Because $\mathcal{W}_q$ is invariant w.r.t. f_q and g_q and $\mathcal{W}_q = \operatorname{span}\{d\vartheta^q_1, \dots, d\vartheta^q_{k_q}, d\omega^q_1, \dots, d\omega^q_{l_q}\}$ with potential coordinate functions $\{\vartheta^q_{\langle k_q \rangle}, \omega^q_{\langle l_q \rangle}\}$ on $\mathcal{X}$, Lemma 1.1 is applied to obtain smooth vector fields F^o_q, G^o_q, $F^\dagger_q$, and $G^\dagger_q$ such that

$$\begin{aligned} \dot{\theta}_q &= \frac{\partial \vartheta^q_{\langle k_q \rangle}}{\partial x}(x) \cdot (f(x) + g(x)u) = F^o_q(\theta_q, w_q) + G^o_q(\theta_q, w_q)u \\ \dot{w}_q &= \frac{\partial \omega^q_{\langle l_q \rangle}}{\partial x}(x) \cdot (f(x) + g(x)u) = F^\dagger_q(\theta_q, w_q) + G^\dagger_q(\theta_q, w_q)u \end{aligned} \tag{42}$$

which match the equations (5a) and (5b). For the remaining coordinate function ς_q, the equations (5c) and (5d) are naturally derived except

$$w_{q+1}(t_q) = R^*_q(w_q(t_q^-), \xi_q(t_q^-), \xi_{q+1}(t_q)), \tag{43}$$

which is proved henceforth.

Lemma 1.3, along with Assumption 4.(d) and $(p_q)_*(\ker \mathcal{W}_q \cap \ker \mathcal{O}'_{q+1}) \subset \ker \mathcal{W}_q$, implies the existence of a function $P^\dagger_q$ (and then P^*_q below, since $\vartheta^q_{\langle k_q \rangle}$ and $\mu'^q_{\langle r'_q \rangle}$ are functions of $\lambda^q_{\langle \nu_q \rangle}$ and $\lambda^{q+1}_{\langle \nu_{q+1} \rangle} \circ p_q$ respectively) such that

$$\begin{aligned} \begin{bmatrix} \vartheta^q_{\langle k_q \rangle}(x(t_q)) \\ \omega^q_{\langle l_q \rangle}(x(t_q)) \end{bmatrix} &= \begin{bmatrix} \vartheta^q_{\langle k_q \rangle}(p_q(x(t_q^-))) \\ \omega^q_{\langle l_q \rangle}(p_q(x(t_q^-))) \end{bmatrix} \\ &= P^\dagger_q(\vartheta^q_{\langle k_q \rangle}(x(t_q^-)), \omega^q_{\langle l_q \rangle}(x(t_q^-)), \mu'^q_{\langle r'_q \rangle}(x(t_q^-))) \\ &= P^*_q(\lambda^q_{\langle \nu_q \rangle}(x(t_q^-)), \omega^q_{\langle l_q \rangle}(x(t_q^-)), \lambda^{q+1}_{\langle \nu_{q+1} \rangle} \circ p_q(x(t_q^-))) \\ &= P^*_q(\xi_q(t_q^-), w_q(t_q^-), \xi_{q+1}(t_q)). \end{aligned} \tag{44}$$

Note that $d\omega^q_j \in \mathcal{W}_q \subset (\mathcal{O}_q + \mathcal{W}_{q-1}) = \operatorname{span}\{d\mu^q_1, \dots, d\mu^q_{r_q}\}$, $j = 1, \dots, l_q$, by the construction of $\mathcal{W}_q$ and Assumption 4.(c). Therefore, by Lemma 1.2, there exist $S^\dagger_q$ and S^*_q such that

$$\begin{aligned} w_{q+1}(t_q) = \omega^{q+1}_{\langle l_{q+1} \rangle}(x(t_q)) &= S^\dagger_{q+1}(\mu^{q+1}_{\langle r_{q+1} \rangle}(x(t_q))) \\ &= S^*_{q+1}(\lambda^{q+1}_{\langle \nu_{q+1} \rangle}(x(t_q)), \vartheta^q_{\langle k_q \rangle}(x(t_q)), \omega^q_{\langle l_q \rangle}(x(t_q))) \\ &= S^*_{q+1}\big(\xi_{q+1}(t_q), P^*_q\big(\xi_q(t_q^-), w_q(t_q^-), \xi_{q+1}(t_q)\big)\big) \\ &=: R^*_q(w_q(t_q^-), \xi_q(t_q^-), \xi_{q+1}(t_q)) \end{aligned}$$

in which, the third equality follows from Assumption 4.(c), and the fourth equality follows from (44). With this definition of R_q^* the equation (43) holds true.

Here, w_1 is a null vector and $l_1 = 0$ because $\mathcal{W}_1 \subset \mathcal{O}_1 + \mathcal{W}_0 = \mathcal{O}_1$, and therefore,

$$\mathcal{W}_1 = \operatorname{span}\{d\vartheta_1^1, d\vartheta_2^1, \ldots, d\vartheta_{k_1}^1\}, \qquad k_1 \leq \nu_1,$$

with potential coordinate functions ϑ_j^1 such that $d\vartheta_j^1 \in \mathcal{O}_1$. (Without loss of generality, let $k_1 \geq 1$ because, if $k_1 = 0$, we proceed to the next mode $q = 2$ with $\mathcal{W}_1 = \{0\}$.) Hence Assumption 2 holds.

Finally, if $\dim(\mathcal{O}_m + \mathcal{W}_{m-1}) = n$ with some m, then $\mathcal{W}_m = \mathcal{O}_m + \mathcal{W}_{m-1}$ by construction. This implies that $k_m + l_m = n$, and from here onwards, the statement of Theorem 1 completes the proof. □

REFERENCES

1. A. Alessandri and P. Coletta. Switching observers for continuous-time and discrete-time linear systems. In *American Control Conference*, pages 2516–2521, 2001.
2. A. Balluchi, L. Benvenuti, M. D. Di Benedetto, and A. L. Sangiovanni-Vincentelli. Observability for hybrid systems. In *IEEE Conf. on Decision and Control*, pages 1159–1164, 2003.
3. J. P. Barbot, H. Saadaoui, M. Djemai, and N. Manamanni. Nonlinear observer for autonomous switching systems with jumps. *Nonlinear Analysis: Hybrid Systems*, 1:537–547, 2007.
4. G. Basile and G. Marro. Controlled and Conditioned Invariant Subspaces in Linear System Theory. *J. of Optimization Theory and Applications*, 3(5):306–315, 1969.
5. P. Collins and J. H. van Schuppen. Observability of piecewise-affine hybrid systems. In R. Alur and G. J. Pappas (Eds.), *Hyb. Sys: Comp. & Control*, ser., *LNCS*, 2993:265–279. Springer, Berlin, 2004.
6. M. Defoort, M. Djemai, T. Floquet and W. Perruquetti. Robust finite time observer design for multicellular converter. *Int. Journal of Systems Science*, 42(11):1859-1868, 2011.
7. F. Esfandiari and H. K. Khalil. Output feedback stabilization of fully linearizable systems. *Int. J. Control*, 56:1007–1037, 1992.
8. J. P. Gauthier, H. Hammouri, and S. Othman. A simple observer for nonlinear systems: applications to bioreactors. *IEEE Trans. on Automatic Control*, 37:875–880, 1992.
9. J. P. Gauthier and I. Kupka. Observability and observers for nonlinear systems. *SIAM J. on Control*, 32(4):975–994, 1994.
10. J. P. Gauthier and I. Kupka. *Deterministic observation theory and applications*. Cambridge University Press, 2001.
11. R. Hermann and A. Krener. Nonlinear controllability and observability. *IEEE Trans. on Automatic Control*, 22(5):728–740, 1977.
12. J. P. Hespanha, D. Liberzon, D. Angeli, and E. D. Sontag. Nonlinear norm-observability notions and stability of switched systems. *IEEE Trans. on Automatic Control*, 50(2):154–168, 2005.
13. A. Isidori. *Nonlinear Control System*. Springer, 3rd edition, 1995.
14. W. Kang, J. P. Barbot, and L. Xu. On the observability of nonlinear and switched systems. In B. Ghosh, C. Martin and Y. Zhou (Eds.), *Emergent Problems in Nonlinear Systems and Control*, ser., *LNCIS*, 393:199–216. Springer, 2009.
15. A. J. Krener and A. Isidori. Linearization by output injection and nonlinear observers. *Systems & Control Letters*, 3:47–52, 1983.
16. H. Nijmeijer and A. J. van der Schaft. *Nonlinear Dynamical Control Systems*. Springer, 1990.
17. S. Pettersson. Designing switched observers for switched systems using multiple Lyapunov functions and dwell-time switching. In *IFAC Conf. on Analysis and Design of Hybrid Systems*, pages 18–23, 2006.
18. H. Shim. *A passivity-based nonlinear observer and a semi-global separation principle*. PhD thesis, Seoul National University, 2000.
19. H. Shim and J. H. Seo. Nonlinear output feedback stabilization on a bounded region of attraction. *Int. J. of Control*, 73(5):416–426, 2000.
20. H. Shim, Y.I. Son, and J.H. Seo. Semi-global observer for multi-output nonlinear systems. *Systems & Control Letters*, 42(3):233–244, 2001.
21. H. Shim and A. Tanwani. On a Sufficient Condition for Observability of Nonlinear Switched Systems and Observer Design Strategy. In *American Control Conference*, pages 1206–1211, 2011.
22. Z. Sun and S. S. Ge. *Switched Linear Systems: Control and Design*. Springer-Verlag London, 2005.
23. H. Sussmann and P. V. Kokotovic. The peaking phenomenon and the global stabilization of nonlinear systems. *IEEE Trans. on Automatic Control*, 36(4):424–440, 1991.
24. A. Tanwani, H. Shim, and D. Liberzon. Observability implies observer design for switched linear systems. *Conf. on Hyb. Sys: Comp. & Control*, pages 3–12, 2011. (Extended version "Observability for switched linear systems: characterization and observer design" has been submitted and is under review.)
25. A. Teel and L. Praly. Global stabilizability and observability imply semiglobal stabilizability by output feedback. *Systems & Control Letters*, 22:313–325, 1994.
26. R. Vidal, A. Chiuso, S. Soatto, and S. Sastry. Observability of linear hybrid systems. In O. Maler and A. Pnueli (Eds.), *Hyb. Sys: Comp. & Control*, ser., *LNCS*, 2623: 526–539. Springer, Berlin, 2003.
27. G. Xie and L. Wang. Necessary and sufficient conditions for controllability and observability of switched impulsive control systems. *IEEE Trans. on Automatic Control*, 49:960–966, 2004.

DOI: 10.1002/rnc