On a Sufficient Condition for Observability of Nonlinear Switched Systems and Observer Design Strategy

Hyungbo Shim and Aneel Tanwani

Abstract—This paper presents a sufficient condition for observability of switched systems that involve state jumps and comprise nonlinear dynamical subsystems affine in control. Without assuming observability of individual modes, the sufficient condition is based on gathering partial information from each mode so that the state is recovered completely after some time. Using this result, an observer is designed which employs a novel ‘back-and-forth’ technique to generate state estimates. Under the assumption of persistent switching, analysis shows that the estimate converges asymptotically to the actual state of the system.

I. INTRODUCTION

We study observability conditions and an observer design for a class of switched nonlinear systems $\Sigma$, described as

$$
\begin{align*}
  \dot{x}(t) &= f_{\sigma(t)}(x(t)) + g_{\sigma(t)}(x(t))u(t), & t \neq \{q\}, \\
  x(t_q) &= p_{\sigma(t_q)}(x(t_q^-)), & (1b) \\
  y(t) &= h_{\sigma(t)}(x(t)), & (1c)
\end{align*}
$$

where $x: \mathbb{R} \mapsto \mathbb{R}^n$ is the state trajectory, $y: \mathbb{R} \mapsto \mathbb{R}^d$ is the output, the measurable function $u: \mathbb{R} \mapsto \mathbb{R}^d$ is the input belonging to some input class $\mathcal{U}$ of interest, and $\sigma: \mathbb{R} \mapsto \mathbb{N}$ is the switching signal that is right-continuous and changes its value at switching times $\{t_q\}, q \in \mathbb{N}$. Let $t_0$ be the initial time and the jump map (1b) applies at $t = t_q, q \geq 1$. It is assumed that there are a finite number of switching times in any finite time interval. The switching mode $\sigma$ and the switching times $\{t_q\}$ may come from a supervisory logic controller, or may be determined internally depending on the system state. In any case, we treat them as a known, external input in this paper. It is assumed that the solution $x(t)$ remains in a compact set $\mathcal{X} \subset \mathbb{R}^n$ on the time interval of interest. All the vector fields and functions are assumed to be smooth, and thus, the existence and uniqueness of the solution, for all times, are guaranteed by the fact that the solution remains in a compact set.

When dealing with observability of nonlinear systems, there are different notions that are involved. The work in [8], [10] talk about observability in local neighborhoods of the state space. Authors in [9] describe the notion of ‘large-time’ versus ‘small-time’ observability where the difference lies in whether it is possible to recover the state instantaneously in time or the system becomes observable after certain time interval. If the system description has exogenous inputs acting on it, then the question arises whether observability holds for all inputs or not [5], [6]; 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., [9]) and the ‘uniform observability’ studied in [5], [6]. 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 [8], [10]. 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 some time. This explains how the concept of ‘large-time’ comes into picture when dealing with switched systems and our goal is to derive conditions that make the system large-time observable on a given set $\mathcal{X}$. Moreover, since we are interested in an observer construction at the end, the observability for all inputs (i.e., uniform observability) is of concern in order for the observer to be independent of particular inputs.

For switched systems, among other structural properties, observability and observer design for linear case have been actively studied during the past decade. Some initial observer results on switched systems, such as [1], [13] for linear case and [3] for nonlinear case, have assumed that each mode in the system is in fact observable admitting a state observer, and have treated the switching as a source of perturbation effect. More relaxed approaches do not assume observability of the individual modes, and the notion of gaining observability by switching for linear systems has appeared in, e.g., [4], [20], [21]. Even though limited to the linear case, it is not clear how the conditions in [4], [20], [21] can lead to feasible observer design. On the other hand, there is not much literature on the observability of switched nonlinear systems.

The main contribution of this paper lies in the unified treatment of observability conditions and observer design which has not been discussed in literature for nonlinear systems, to the authors’ knowledge. For the observer design, our approach shares the same spirit as [2], and the result of this paper can be regarded as an extension of [2], in the sense that, a coordinate-independent condition is derived for observability and nonlinear systems are treated with a new observer design strategy.

The notation used in this paper is summarized as follows.
\( \mathcal{R}(A) \) implies the range space of the columns of matrix \( A \), and \( A^T \) is the transpose of \( A \). We denote \([x_1, x_2]^T\) simply by \( \text{col}(x_1, x_2) \), and \( \lambda_1, \ldots, \lambda_k \). For a signal \( x(t) \), \( x(t, t) \) means \( \{x(t), t : t_1 \leq t \leq t_2\} \). The differential of a map \( p \) acting on the vector field \( v \) is denoted by \( p \cdot v \).

For a distribution \( \mathcal{W} \), \( p \cdot \mathcal{W} = \{ p \cdot v \mid v \in \mathcal{W} \} \). We call a distribution \( \mathcal{W} \) at \( x^0 \) nonsingular" when \( \dim \mathcal{W} \) is constant in a neighborhood of \( x^0 \).

The notation \( L \cdot \text{com} \{ \lambda_1(x), \ldots, \lambda_k(x) \} \) means a set of linear combinations of the functions \( \lambda_i \) with constant coefficients, i.e., \( \{ \sum_{i=1}^k c_i \lambda_i(x) : c_i \in \mathbb{R} \} \). 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 \( \text{col}(\lambda_1(x), \ldots, \lambda_n(x)) \) 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 \( \lambda \) along the vector field \( F_x(\lambda)(x) = \frac{\partial \lambda(x)}{\partial x} f(x) \) and \( L_f(x)(x) = dL_f(x) \). Some proofs have been omitted in this conference paper due to space constraints.

II. Preliminaries

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

**Definition 1:** A system \( \Sigma \) with a switching signal \( \sigma(t) \) is large-time uniformly observable on a set \( \mathcal{X} \subset \mathbb{R}^n \) if there exist a finite time \( T > t_0 \) so that \( x(T) \) is determined uniquely from \( y(t), u(t) \), and \( \sigma(t) \) for any measurable input \( u(t) \), when the state \( x(t) \) remains in \( \mathcal{X} \) for \( t \in [t_0, T] \). If the time \( T > t_0 \) can be chosen arbitrarily, then the system \( \Sigma \) is called small-time uniformly observable on a set \( \mathcal{X} \).

In case of no jump map (1b), the knowledge of \( x(T), \sigma(T) \), and \( u(T) \) determines \( x(T) \) uniquely. This is not the case in general because the jump map (1b) may not be reversible. From the definition, if a certain mode of system \( \Sigma \) is small-time observable and the switching signal activates that mode at a certain time, then the system is automatically large-time observable. Note that (2) may be reconstructed using the derivatives of \( y(t) \) and \( u(t) \) (although differentiation should not be used in the observer construction). It is noted that, although the observability in Definition 1 is uniform with respect to the input \( u \), uniformity with respect to the switching signal \( \sigma \) is not required.

The following lemma will be frequently used in the paper.

**Lemma 1:** Consider a codistribution \( \mathcal{W} \) generated by exact one-forms, i.e., \( \mathcal{W} = \text{span} \{ d\lambda_1, \ldots, d\lambda_k \} \) with \( 1 \leq k \leq n \), where \( \lambda_1, \ldots, \lambda_k \) are smooth potential coordinate functions defined on a set \( \mathcal{X} \subset \mathbb{R}^n \).

1) If the codistribution \( \mathcal{W} \) is invariant with respect to a smooth vector field \( f(x) \), i.e.,

\[ L_f(x) \mathcal{W} \subset \mathcal{W} \]

on \( \mathcal{X} \), then there exists a smooth vector field \( F : \lambda_{1, \ldots, k}(x) \rightarrow \mathbb{R}^k \) such that \( \frac{\partial \lambda_{1, \ldots, k}(x)}{\partial x} f(x) = F(\lambda_{1, \ldots, k}(x)) \) on \( \mathcal{X} \).

2) If a smooth function \( h : \mathcal{X} \rightarrow \mathbb{R} \) satisfies

\[ dh \in \mathcal{W} \]

on \( \mathcal{X} \), then there exists a smooth function \( H : \lambda_{1, \ldots, k}(x) \rightarrow \mathbb{R} \) such that \( h(x) = H(\lambda_{1, \ldots, k}(x)) \) on \( \mathcal{X} \).

3) Let \( \mathcal{W}' \) be another codistribution such that \( \dim (\mathcal{W} + \mathcal{W}') \) is constant on \( \mathcal{X} \), and suppose that \( \mathcal{W} + \mathcal{W}' = \text{span} \{ d\lambda_{1, \ldots, k}, dp' \mid j = 1, \ldots, r' \} \) where \( r' = \dim (\mathcal{W} + \mathcal{W}') - \dim \mathcal{W} \), and the elements of \( \{ \lambda_{1, \ldots, k}, p' \} \) are smooth potential coordinate functions on \( \mathcal{X} \). If a smooth map \( p : \mathcal{X} \rightarrow \mathbb{R}^n \) satisfies

\[ p_k(\ker \mathcal{W} \cap \ker \mathcal{W}') \subset \ker \mathcal{W} \]

on \( \mathcal{X} \), then there exists a smooth map \( P : \lambda_{1, \ldots, k}(x) \rightarrow \mathbb{R}^k \) such that \( \lambda_{1, \ldots, k}(p(x)) = P(\lambda_{1, \ldots, k}(x)) \lambda_{1, \ldots, k}(p(x)) \) while \( x \) and \( p(x) \) are contained in \( \mathcal{X} \).

Before dealing with the switched case, let us consider the system (1a) and (1c) for a fixed mode \( q \), without the jump map (1b) for now. In particular, we note that the individual subsystems may not be observable, which calls for the classical Kalman decomposition [10]: changing the coordinates so that the system is explicitly split into the observable part and the unobservable part. For this, let the observation space \( \mathcal{O}_q \) be the linear space of functions over \( \mathbb{R} \) containing all \( h_{q,i} \)'s (where \( h_{q,i} \) is the \( i \)-th element of \( h_q \)) and all repeated Lie derivatives \( L_{q_1}L_{q_2} \cdots L_{q_n}h_{q,i} \), with \( v_1 \in \{ f_q, q_{1,1}, \ldots, q_{d,q} \} \) (i.e., \( q_{j} \) is the \( j \)-th column of \( q_{j} \)).

**Assumption 1:** For each mode \( q \), the codistribution \( d\mathcal{O}_q = \text{span} \{ \lambda : \lambda \in \mathcal{O}_q \} \) has constant dimension \( k_q \); \( \dim d\mathcal{O}_q = k_q \) on the set \( \mathcal{X} \). In addition, there are \( k_q \) smooth potential coordinate functions \( \lambda_{j,q}, j = 1, \ldots, k_q \), such that

\[ d\mathcal{O}_q = \text{span} \{ \lambda_{q,1}, \ldots, \lambda_{q,k_q} \} \text{ on } \mathcal{X} \]

Under Assumption 1, additional functions \( \lambda_{q,k_q+1}, \ldots, \lambda_{q,n} \) can be found to yield a diffeomorphism \( \lambda_q := \text{col}(\lambda_{1,q}(x), \ldots, \lambda_{n,q}(x)) \) on \( \mathcal{X} \). Then, Lemma 1.1 implies that the system (1a) and (1c) for mode \( q \) can be written in the new coordinates \( \text{col}(\xi_{q,1}, \xi_{q,2}) \) as

\[ \begin{align*}
\dot{\xi}_q &= F_q(\xi_q, \xi_q) + G_q(\xi_q, \xi_q) u \\
\dot{\xi}_q &= F_q(\xi_q) + G_q(\xi_q) u \\
y &= H_q(\xi_q).
\end{align*} \]

This representation is valid on the set \( \lambda_q(\mathcal{X}) \).

**Assumption 2:** The reduced-order subsystem (2b) and (2c) is small-time uniformly observable on the set \( \mathcal{X}_q := \lambda_{1,\ldots,k_q}(\mathcal{X}) \), which is the projection of \( \lambda_q(\mathcal{X}) \) onto the \( \xi_q \)-coordinates.

**Remark 1:** As a matter of fact, the simple condition that \( \dim d\mathcal{O}_q(x) = k_q \) in a neighborhood of some \( x^0 \in \mathcal{X} \) guarantees, by Frobenius theorem, that there exists a local neighborhood \( \mathcal{X}'(\subset \mathcal{X}) \) of \( x^0 \) such that Assumption 1 holds. Compared to this local observability (studied in, e.g., [8], [10], [12]), Assumptions 1 and 2 may be thought of as global versions ("global" in the sense of the whole region \( \mathcal{X} \)).
As long as we restrict our attention to (2b) and (2c) for each mode, Assumption 2 becomes the standard uniform observability assumption that has often been studied in the literature (see [7] and references therein). Assumption 2 can be checked in various ways; for instance, if the class of inputs \( \mathcal{U} \) consists of smooth functions only, then one may try to find a function \( \mathcal{E} \) such that
\[
\mathcal{E}(y, ar{y}, \ldots, y^{(n_y-1)}, u, \bar{u}, \ldots, u^{(n_u-1)})
\]
where \( n_y \in \mathbb{N} \) and \( n_u \in \mathbb{N} \), and that the function \( \mathcal{E}(\cdot, u, \bar{u}, \ldots, u^{(n_u-1)}) \) is surjective onto \( \mathbb{R}^n \) for all \( u(\cdot) \in \mathcal{U} \). The existence of such a function \( \mathcal{E} \) to reconstruct \( x(T) \) is used as the definition of uniform observability in [10], [19]. Other ways to check Assumption 2 can be found in [14].

### III. SUFFICIENT CONDITION FOR OBSERVABILITY

In deriving the sufficient condition for observability, we do not assume the individual modes of the system to be observable. Hence, in order to recover the system state \( x(t) \), partial information obtained from each mode is accumulated. This partial information is quantified in terms of the maximal integral submanifold of the distribution \( d\mathcal{O}_q \), which has the property that the states on the slices (or “leaves”) of this submanifold are not distinguishable by the output of mode \( q \). As soon as a switch occurs, the indistinguishable states must be contained in the intersection of the integral submanifolds of both modes before and after the switching. In this way, switching reduces the uncertainty in the state. Continuing in this manner, with subsequent switching, we expect to reduce the size of submanifold that contains the indistinguishable states. Eventually, if the corresponding intersections reduce to a point, we obtain observability. However, while the intersections are taken at the same time, the information contained in the integral submanifolds is scattered in time because each one of them becomes available sequentially as time goes on. This suggests that the partial information, obtained at each mode, should evolve uncorrupted along the dynamics of subsequent modes until all the information is gathered to compute the state. Inspired by this intuition, we present structural conditions which guarantee that the evolution of the partial information is feasible without being affected by the unknown quantities in subsequent modes.

Before presenting the condition, 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 description also incorporates cases where there is a state jump without change in dynamics or the mode change does not involve state jumps.

**Theorem 1:** Suppose that Assumptions 1 and 2 hold, and define \( d\mathcal{O}_q := \text{span} \{d(\lambda_{q,i} \circ p_{q-1}) : i = 1, \ldots, k_q \} \) for each \( q \geq 2 \). On \( \mathcal{X} \), define a sequence of codistributions \( \mathcal{W}_q \), with \( \mathcal{W}_0 := \{0\} \), as: \( \mathcal{W}_q \) is the largest nonsingular and involutive codistribution, invariant with respect to \( f_q \) and \( g_q \), contained in \( (d\mathcal{O}_q + W_{q-1}) \) such that \( (p_q)_* (\ker \mathcal{W}_q \cap \ker d\mathcal{O}_{q+1}) \subset \ker \mathcal{W}_{q+1} \).

If

1) \( \exists \ m \geq 1 \) such that, on \( \mathcal{X} \),
\[
\dim(d\mathcal{O}_m + W_{m-1}) = n,
\]
(or simply \( \dim W_m = n \) because \( W_m = d\mathcal{O}_m + W_{m-1} \) by construction),
2) the codistributions \( W_q \) (1 \( \leq q \leq m \), \( d\mathcal{O}_q + W_{q-1} \) (2 \( \leq q \leq m \), and \( W_q + d\mathcal{O}_{q+1} \) (1 \( \leq q \leq m - 1 \) are nonsingular on \( \mathcal{X} \). Moreover,

(a) \( \exists \) smooth potential coordinate functions \( \{\phi_{q,i}, \omega_{q,j} : i = 1, \ldots, k_q, j = 1, \ldots, \bar{l}_q, \bar{k}_q + \bar{l}_q = \dim \mathcal{W}_q\} \) on \( \mathcal{X} \) such that
\[
\mathcal{W}_q = \text{span} \{d\phi_{q,1}, \ldots, d\phi_{q,k_q}, d\omega_{q,1}, \ldots, d\omega_{q,\bar{l}_q}\},
\]
\[
d\phi_{q,i} \in d\mathcal{O}_q, \quad d\omega_{q,j} \notin d\mathcal{O}_q \quad (3)
\]
(b) \( \exists \) smooth potential coordinate functions \( \{\mu_{q,i} : i = 1, \ldots, \bar{r}_q, \bar{r}_q = \dim(d\mathcal{O}_q + W_{q-1})\} \) on \( \mathcal{X} \) such that
\[
d\mathcal{O}_q + W_{q-1} = \text{span} \{d\mu_{1,1}, \ldots, d\mu_{1,k_q}\},
\]
\[
\mu_{q,i} \in \text{com} \{\lambda_{q,1}, \ldots, \lambda_{q,k_q}, \phi_{q,1}, \ldots, \}
\]
\[
\phi_{q-1-k_q+1, \omega_{q-1,1}, \omega_{q-1,2}, \ldots, \omega_{q-1,\bar{l}_q}}, \quad (5)
\]
(c) \( \exists \) smooth potential coordinate functions \( \{\mu'_{q,j} : j = 1, \ldots, \bar{r}'_q, \bar{r}'_q = \dim(\mathcal{W}_q + d\mathcal{O}'_{q+1}) - \dim \mathcal{W}_q\} \) on \( \mathcal{X} \) such that
\[
\mathcal{W}_q + d\mathcal{O}'_{q+1} = \text{span} \{d\mu'_{1,1}, \ldots, d\mu'_{1,k_q}\},
\]
\[
d\omega_{q,1}, \ldots, d\omega_{q,\bar{l}_q}, d\mu'_{1,1}, \ldots, d\mu'_{1,k_q}, \quad (6)
\]
\[
\mu'_{q,j} \in \text{com} \{\lambda_{q+1,1}, \ldots, \lambda_{q+1,k_q}, \phi_{q,1}, \ldots, \}
\]
\[
p_{q,j} \quad (5)
\]

then the system (1) is large-time uniformly observable on \( \mathcal{X} \) for all the switching signals containing the consecutive subsequence \( \{1, 2, \ldots, m\} \).

Following observations are immediate: (a) \( d\mathcal{O}_q \) itself is invariant with respect to \( f_q \) and \( g_q \) by construction, if \( p_q(x) = x \), so that there is no state jump, then the condition \( (p_{q,j})_* (\ker \mathcal{W}_q \cap \ker d\mathcal{O}_{q+1}) \subset \ker \mathcal{W}_{q+1} \) automatically holds, the resulting codistribution in the assumption of Theorem 1 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 \( \mathcal{p}_i \circ \mathcal{W}_i \subset \ker \mathcal{W}_i \), where \( i \in \{a, b\} \), then the corresponding codistribution \( \mathcal{W}_a + \mathcal{W}_b \) satisfies the conditions above, and any differentiable map \( p \) and any distribution \( D \), then \( p_* (\ker (\mathcal{W}_a + \mathcal{W}_b) \cap D) \subset \ker (\mathcal{W}_a + \mathcal{W}_b) \).

The compactness of the set \( \mathcal{X} \) guarantees the solution without finite escape time, and will be used for observer construction in the next section. If all the assumptions hold with \( \mathcal{X} = \mathbb{R}^n \), then the observability property becomes global in case the solution has no finite escape time. On the other hand, if local observability is of interest, then the assumptions get simpler by removing the items 2(a), 2(b), and 2(c).

**Corollary 1:** Suppose that Assumptions 1 and 2 hold in a neighborhood of a point \( x_0 \in \mathcal{X} \). If each of the codistributions \( \mathcal{W}_q, d\mathcal{O}_q + W_{q-1}, \) and \( \mathcal{W}_q + d\mathcal{O}'_{q+1} \) are
nonsingular at $x^o$, $W_p$ is smooth and involutive at $x^o$, and 
$\dim(dO_m + W_m-1)(x^o) = n$, then the system is large-time
uniformly observable in some neighborhood of $x^o$ for all
the switching signals containing the consecutive subsequence
$\{1, 2, \ldots, m\}$. <

Now we present the proof of Theorem 1, which is con-
structive in the sense that a technique to recover $x(t)$ at some
time $t = T > t_{m-1}$ is revealed (rather than discussing the
indistinguishability of two different states). This way paves
the road to the observer design in the next section.

Proof: [Proof of Theorem 1] Consider the interval prior to
the first switching, $[t_0, t_1)$. Since $W_1 \subset dO_1$, we have
that

$$W_1 = \text{span } \{d\phi_{1,1}, d\phi_{1,2}, \ldots, d\phi_{1,k_1}\}, \quad \bar{k}_1 \leq k_1.$$

Because $d\phi_{1,i}$, for each $i = 1, \ldots, \bar{k}_1$, is an element of
d$O_1$ that is generated by the differentials of $\lambda_{1,l}$, $l = 1, \ldots, k_1$, and $\lambda_{1,l}$'s are potential coordinate functions on $X$ (by Assumption 1), $\phi_{1,i}$ is a function of $\lambda_{1,1-k_1}$ only (by Lemma 1.2). Since $\xi_1 := \lambda_{1,1-k_1}(x)$ is small-time uniformly
observable on $\lambda_{1,1-k_1}(X)$ (by Assumption 2), the value of
the vector $\xi_1(t) = \lambda_{1,1-k_1}(x(t))$, and thus, $\phi_{1,1-k_1}(x(t))$ are recovered for $t \in [t_0, t_1)$.

Now Lemma 1.3, with the item 2(c) and (p)$_1$(ker $W_1 \cap
\ker dO_2 \subset$ ker $W_1$, implies the existence of a function $P_1$
(and then $P_2$ below, since $\phi_{1,i}$ is a function of $\lambda_{1,1-k_1}$) such that

$$\phi_{1,1-k_1}(x(t_1)) = \phi_{1,1-k_1}(P_1(x(t_1)))
= P_1(\lambda_{1,1-k_1}(x(t_1)), \lambda_{1,1-k_1} \circ p_1(x(t_1)))
= (\lambda_{1,1-k_1}(x(t_1)), \lambda_{1,1-k_1} \circ p_1(x(t_1)))
= (\lambda_{1,1-k_1}(x(t_1)), \lambda_{1,1-k_1}(x(t_1))), $$(8)

where the third equality follows from (3) and (7).

Next, consider the interval $[t_1, t_2)$. For $i = 1, \ldots, \bar{k}_2$,
using Lemma 1.2, the condition $d\phi_{2,i} \in dO_2 = \text{span } \{d\lambda_{2,1}, \ldots, d\lambda_{2,k_2}\}$ guarantees that $\phi_{2,i}$ is a function of $\lambda_{2,1-k_2}$ only. Again by Assumption 2, the vector $\xi_2(t) := \lambda_{2,1-k_2}(x(t))$, and thus, $\phi_{2,1-k_2}(x(t))$ are recovered for the interval $[t_1, t_2)$.

Now observing that $W_2$ is invariant w.r.t. $f_2$ and $g_2$,
$W_2 = \text{span } \{d\phi_{2,1}, \ldots, d\phi_{2,k_2}, d\omega_{2,1}, \ldots, d\omega_{2,k_2} : \bar{k}_2 + \bar{k}_2 = \dim W_2\}$, and $\{d\phi_{2,1-k_2}, d\omega_{2,1-k_2}\}$ are potential coordinate functions on $X$, we apply Lemma 1.1 and obtain smooth vector fields $\tilde{F}_2$ and $G_2$ such that, with $z_2 := \omega_{2,1-k_2}(x)$,

$$\dot{z}_2 = \frac{\partial \omega_{2,1-k_2}(x)}{\partial x} \cdot (f(x) + g(x))u
= \tilde{F}_2(z_2, \phi_{2,1-k_2}(x)) + G_2(z_2, \phi_{2,1-k_2}(x))u \quad (9)$$

over $[t_1, t_2)$. In this interval, the vector $\xi_2(t) = \lambda_{1,2-k_2}(x(t))$
is recovered. Hence, if the initial condition $z_2(t_1) = \omega_{2,1-k_2}(x(t_1))$ is recovered, the vector $\xi_2(t)$ on the interval $[t_1, t_2)$ is also known by solving the differential equation (9).

Note that $d\omega_{2,j} \in W_2 \subset (dO_2 + W_1) = \text{span } \{d\mu_{2,1}, \ldots, d\mu_{2,r_2}\}, j = 1, \ldots, r_2$, by the definition of $W_2$ and the item 2(b). Therefore, by Lemma 1.2, $\omega_{2,j}$
can be written as a function of $\mu_{2,i}$’s, which leads to

$$z_2(t_1) = \omega_{2,1-k_2}(x(t_1)) = \tilde{S}_2^*(\mu_{2,1-k_2}(x(t_1)))
= \tilde{S}_2^*(\xi_2(t_1), \phi_{2,1-k_2}(x(t_1)))
= \tilde{S}_2^*(\xi_2(t_1), \lambda_{1,1-k_1}(x(t_1)), \lambda_{2,1-k_2}(x(t_1)))
= \tilde{S}_2^*(\xi_2(t_1), \xi_2(t_1), \lambda_{1,1-k_1}(x(t_1)))$$

in which, the third equality uses (5) with $\omega_1$ being null, and
the fourth equality follows from (8).

This process is repeated to find $P_q^*, G_q^*,$ and $q$. For
instance, we can find $P_2$ such that

$$\begin{align*}
\tilde{P}_2(\phi_{2,1-k_2}(x(t_2))), \omega_{2,1-k_2}(x(t_2)), \mu_{2,1-k_2}(x(t_2))
= \tilde{P}_2(\lambda_{2,1-k_2}(x(t_2)), \omega_{2,1-k_2}(x(t_2)), \lambda_{2,1-k_2} \circ p_2(x(t_2)))
= P_2(\lambda_{2,1-k_2}(x(t_2)), \omega_{2,1-k_2}(x(t_2)), \lambda_{2,1-k_2} \circ p_2(x(t_2)))
= P_2(\xi_2(t_2), \lambda_{2,1-k_2}(x(t_2)), \lambda_{2,1-k_2}(x(t_2))),
\end{align*}$$

and find $S_q^*$ such that

$$\begin{align*}
z_3(t_2) = \omega_{3,1-k_3}(x(t_2)) = \tilde{S}_3^*(\mu_{3,1-k_3}(x(t_2)))
= \tilde{S}_3^*(\xi_3(t_2), P_2(\xi_2(t_2), \omega_{2,1-k_2}(x(t_2))), \lambda_{2,1-k_2}(x(t_2)))
= S_q^*(\xi_3(t_2), P_2(\xi_2(t_2), z_2(t_2)), \lambda_{2,1-k_2}(x(t_2))).
\end{align*}$$

In summary, for each time interval $[t_{q-1}, t_q)$, $q = 1, \ldots, m$, we have the differential equation (with $z_1$ being null)

$$\begin{align*}
\dot{\xi}_1 &= \hat{F}_q(\xi_1) + \hat{G}_q(\xi_1)u, \quad \xi_1 \in \mathbb{R}^{k_1}, \quad (11a) \\
y &= \hat{H}_q(\xi_1), \quad (11b) \\
\dot{z}_q &= \hat{F}_q(z_q, \xi_1) + \hat{G}_q(z_q, \xi_1)u, \quad \xi_1 \in \mathbb{R}^{k_1}, \quad (11c) \\
z_q(t_{q-1}) &= \hat{S}_q^*(\xi_1(t_{q-1}), \xi_2(t_{q-1}), \xi_3(t_{q-1})), \quad (11d) \\
z_{q-1}(t_{q-1}) &= \hat{S}_q^*(\xi_1(t_{q-1}), \xi_2(t_{q-1}), \xi_3(t_{q-1})),
\end{align*}$$

in which, $\xi_q(t)$ and $z_q(t)$ are completely determined.

At any time $t = T > t_{m-1}$, it follows under the
assumption in item 1, i.e. $\dim W_m = n$, that the vectors
$\lambda_m(T)$ and $z_m(T)$ are completely recovered or equivalently
$\text{col}(\phi_{m,1-k_m}(x(T)), \omega_{m,1-k_m}(x(T)))$ is determined. In this
way $x(T)$ is recovered uniquely by the inverse, because
$\text{col}(\phi_{m,1-k_m}, \omega_{m,1-k_m})$ is a diffeomorphism.

IV. OBSERVER DESIGN

Based on the study of large-time observability, let us now
discuss the design of an asymptotic observer for the sys-
tem (1). By asymptotic observer, we mean that the estimate
$\hat{x}(t)$ that it generates, converges to the plant state $x(t)$ as
time tends to infinity. In order to achieve this, we introduce
the following assumptions.

**Assumption 3:**

1) The switching is persistent and happens within the duration $D$; that is,

$$t_q - t_{q-1} \leq D, \quad \forall q \in \mathbb{N} \quad (12)$$

where $t_q$ is the switching time.
2) The solution $x(t)$ of the plant (1) remains in a compact set $\mathcal{X} \subset \mathbb{R}^n$, and the input $u(t)$ is uniformly bounded; $|u(t)| \leq M_u$.

3) There is an $m \in \mathbb{N}$ such that the assumption of Theorem 1 holds on a set $\mathcal{X}$ that properly contains $\mathcal{X}$, and the mode sequence repeats the particular modes $\{1, 2, \ldots, m\}$; that is, $\sigma(t) = ((q-1) \mod m) + 1$ for $[t_{q-1}, t_q)$, $q \in \mathbb{N}$.

We do not consider the time consumed for computation by assuming that the data processor is fairly fast compared to the plant process. The computation time, however, needs to be considered in real-time application if the plant itself is fast.

The observer we propose is of hybrid-type, and has the form

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

$$\hat{x}(t_q) = \begin{cases} \hat{p}_q(\hat{x}(t_q)), & (q \mod m) \neq 0, \\ \hat{p}qL_q(y[t_{q-m}, t_q), u[t_{q-m}, t_q)), & (q \mod m) = 0, \end{cases}$$

(13)

with an initial condition $\hat{x}(t_0) \in \mathcal{X} \subset \mathcal{X}$, where $\hat{f}_q$, $\hat{g}_q$, and $\hat{p}_q$ are globally Lipschitz and they have the same values as $f_q$, $g_q$, and $p_q$, respectively, inside the compact set $\mathcal{X}$. Their global Lipschitz property can always be obtained by modifying them outside the set $\mathcal{X}$, using the so-called ‘Lipschitz extension’.\(^1\) It is seen that the observer consists of a plant copy with an estimate update law by some operator $L_q$, which we design in this section. In fact, we present a design of $L_q$, using some dynamic observers for partial states at each mode and an inversion algorithm logic in order to achieve,

$$|\hat{x}(t_m)| \leq \gamma |\hat{x}(t_0)|$$

(14)

where $0 < \gamma < 1$ and $\hat{x} := \hat{x} - x$. The Lipschitz property of (13) and Assumptions 3.1 and 3.2 guarantee that

$$\sup_{t \in [t_{j-1}, t_j)} |\dot{x}(t)| \leq \Gamma |\dot{x}(t_{j-1})|$$

with a constant $\Gamma$ and $j \in \mathbb{N}$. In this way, if (14) holds then its repeated application leads to $\lim_{t \to \infty} |\dot{x}(t)| = 0$.

The proposed observer construction is based on the representation (11) of the plant (1). The idea is that, for each interval $[t_{q-1}, t_q)$, $q = 1, \ldots, m$, a conventional nonlinear observer, which we call $\xi_q$-observer, is employed to obtain the estimate $\hat{\xi}_q(t)$ for that interval. At the same time, a $z_q$-observer, replicating (11c) and (11d), is constructed as follows:

$$\dot{\hat{z}}_q = \hat{F}_q^{*}(\hat{\xi}_q, \hat{\dot{z}}_q, \hat{z}_q)u, \quad 2 \leq q \leq m,$$

(15)

with the initial condition given by:

$$\hat{z}_q(t_{q-1}) = \hat{S}_q(\hat{\xi}_q(t_{q-1}), \hat{\dot{z}}_q(t_{q-1}), \hat{z}_q(t_{q-1})), \quad \hat{z}_{q-1}(t_{q-1}), \hat{\xi}_q(t_{q-1})),$$

(16)

and $\hat{\xi}_1 := 0$ for convenience. Here $F_q^{*}$ is the Lipschitz extension of $F_q$ with respect to the set $Z_q \times \Xi_q$ and so

\(^1\)Since the plant state $x(t)$ remains in $\mathcal{X}$, this modification can also be applied to the plant model (1). See [5] for its utility in observer construction. Detailed procedures to the modification have been discussed in [14], [15].

on ($Z_q$ is the image of $\mathcal{X}$ through $\omega_{q,1} \sim \omega_q^{*}$ and $\Xi_q$ through $\lambda_{q,1} \sim \lambda_{k_0}$). In fact, the variable $z_q$ is not an observable quantity for the mode $q$. Intuitively speaking, the role of $z_q$-observer is not to reduce the error $\hat{z}_q(t) := \hat{z}_q(t) - z_q(t)$, but to deliver the estimates $\hat{\xi}_{q-1}(t_{q-1})$ and $\hat{z}_{q-1}(t_{q-1})$, that are obtained from the previously active mode and are encoded in the initial condition (16), to the next mode through $\hat{z}_q(t)$ along the system dynamics. Suppose that, seen at time $t = t_m$, an ideal observer provides the exact information of $\xi_q(t)$ on each interval $[t_{q-1}, t_q)$, $q = 1, \ldots, m$, using the stored input $u$ and the output $y$ with the model (11a) and (11b). For example, with exact values of $\xi_1(t_{1-1}) = \xi_1(t_{1})$ and $\xi_2(t_{1}) = \xi_2(t_{1})$, we obtain the exact value of $\hat{z}_2(t_{1}) = \hat{z}_2(t_{1})$ by (16). Then, integration of (15) for $q = 2$ results in exact values of $\hat{z}_2(t) = z_2(t)$ on $[t_1, t_2)$. This process repeats until we get $\hat{\xi}_m(t_m) = \xi_m(t_m)$ and $\hat{z}_m(t_m) = z_m(t_m)$, with Assumption 3.3, i.e. dim $W_m = n$, $x(t_m)$ is now determined uniquely from $\xi_m(t_m)$ and $z_m(t_m)$, as the map

$$x(t_m) \mapsto \begin{pmatrix} \phi_m,1-k_m(x(t_m)) \\ \omega_m,1-k_m(x(t_m)) \end{pmatrix} = \begin{pmatrix} \chi(\xi_m(t_m)) \\ z_m(t_m) \end{pmatrix}$$

(17)

is invertible; here $\chi$ is a function such that $\chi(\lambda_{m,1-k_m}(x)) = \phi_m,1-k_m(x)$ whose existence is guaranteed by Lemma 1.2. For convenience let us denote the inverse map by $\bar{\Psi}$, so that

$$x(t_m) = \bar{\Psi}(\xi_m(t_m), z_m(t_m))$$

(18)

As a result, we choose the estimate update law in (13) to be:

$$\hat{z}(t_m) = \bar{\Psi}(\hat{\xi}_m(t_m), \hat{z}_m(t_m)) = : L_q(y[t_{q-1}, t_q), u[t_{q-1}, t_q]),$$

(19)

where $\bar{\Psi}$ is Lipschitz extension of $\Psi$. Through this relation, the plant state is recovered as $\hat{z}(t_m) = x(t_m)$ with exact information $\xi_m(t_m) = \xi_m(t_m)$ and $\hat{z}_m(t_m) = z_m(t_m)$.

However, asymptotic observers in practice inevitably introduce some error in $\hat{\xi}_q(t)$ while estimating $\xi_q(t)$. Moreover, the estimation of $\xi_q(t)$ on the entire interval $[t_{q-1}, t_q)$ needs more attention because the conventional observers, initiated at the time $t = t_{q-1}$, often experience the transient overshoot before they converge to the proper estimates. Reducing the transient period by increasing observer gain may worsen the situation because of the peaking phenomenon [17]; that is, the peaking in $\xi_q(t)$ may damage the role of (15) because large error in $|\hat{z}_q(t_{q-1}) - z_q(t_{q-1})|$ may occur in spite of small error in $|\hat{\xi}_q(t_{q-1}) - z_q(t_{q-1})|$.

We overcome this obstacle by employing a new back-and-forth estimation technique. Suppose that the $\xi_q$-observer over the interval $[t_{q-1}, t_q)$ is written as

$$\hat{\xi}_q^{f} = \hat{F}_q^{f}(\hat{\xi}_q^{f}), \quad \hat{\xi}_q^{f}(t_{q-1}) = \hat{\lambda}_{q,1-k_q}(\hat{x}(t_{q-1})), \quad 1 \leq q \leq m$$

(19)

where $\hat{\lambda}_{q,1-k_q}$ is the Lipschitz extension of $\lambda_{q,1-k_q}$ and the superscript ‘$f$’ indicates ‘forward’, whose meaning will soon become clear from the context. Here, $F_q$, $G_q$, and $H_q$ can be any modified vector fields or functions as long as $\hat{F}_q^{f}(\hat{\xi}_q) = F_q(\hat{\xi}_q)$, $\hat{G}_q(\hat{\xi}_q) = G_q(\hat{\xi}_q)$, and $\hat{H}_q(\hat{\xi}_q) = H_q(\hat{\xi}_q)$ for $\xi_q \in \Xi_q$. In this way, we allow many observer design techniques
available in the literature (which may utilize the Lipschitz extension of $F_q$, $G_q$, and $H_q$ in another coordinate system).

**Assumption 4:** Under the assumption that $\xi_q(t) \in \Xi_q$ for $[t_{q-1}, t_q)$, $q = 1, \ldots, m$, the subsystem (11a) and (11b) admits an observer of the form (19), which can be made to converge to the state $\xi_q(t)$ arbitrarily fast, that is, for arbitrarily small constants $b > 0$ and $c > 0$, there exist an injection gain $K^b_q(\cdot)$ and a class-KC function $\beta^b_q(\cdot)$ satisfying

$$\beta^b_q(\alpha, t) < c\alpha$$

for all $\alpha > 0$ and $b \leq t \leq \tau_q$.

$$|\dot{\tilde{\xi}}^b_q(t) - \dot{\xi}_q(t)| \leq \beta^b_q(|\dot{\tilde{\xi}}^b_q(t_{q-1}) - \dot{\xi}_q(t_{q-1})|), \quad t \in [t_{q-1}, t_q)$$

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

**Remark 2:** Many results in the literature, such as [5], [11], yield an observer satisfying Assumption 4 with $\beta^b_q$ being an exponential function.

Now consider another (backward) observer described as

$$\dot{\tilde{\xi}}^b_q = -\tilde{F}_q(\tilde{\xi}_q) - \tilde{G}_q(\tilde{\xi}_q)u(t_q - t) - K^b_q(\tilde{\xi}_q, y(t_q - t))\|y(t_q - t) - \dot{H}_q(\tilde{\xi}_q)|,

\tilde{\xi}_q^b(0) = \tilde{\xi}_q^b(t_q), \quad t \in (0, \tau_q), \quad 1 \leq q \leq m.$$  

Actually, the trajectory $\tilde{\xi}_q^b(t) := \xi_q(t_q - t)$ satisfies the differential equation

$$\dot{\tilde{\xi}}^b_q = -F_q(\tilde{\xi}_q^b) - G_q(\tilde{\xi}_q^b)u(t_q - t), \quad y(t_q - t) = H_q(\tilde{\xi}_q^b),$$

with $\tilde{\xi}_q^b(0) = \tilde{\xi}_q(t_q)$, for $t \in (0, \tau_q)$, and therefore, 22 can be thought of as one possible observer for it. We further assume that Assumption 4 holds for this case as well, with $\tilde{\xi}_q^b, \xi_q, \beta^b_q$ and $K^b_q$ replaced by $\tilde{\xi}_q^b, \xi_q, \beta^b_q$ and $K^b_q$, respectively. Once Assumption 4 holds for (19), this additional requirement is mild. For example, the designs of [11] and [5] readily satisfy this requirement. Then, using the input $u$ and the output $y$ stored over the interval $[t_{q-1}, t_q)$, we run the observer (19) first from the initial condition $\tilde{\xi}_q^b(t_{q-1}) = \lambda_{q-1, q}(\dot{x}(t_{q-1}))),$ followed by integrating (22) from 0 to $\tau_q$. After that, we take our final estimate $\xi_q(t)$ as

$$\dot{\xi}_q(t) = \begin{cases} \tilde{\xi}_q^b(t_{q-1}), & t \in [t_{q-1}, t_{q-1} + \tau_q/2), \\ \tilde{\xi}_q^b(t), & t \in [t_{q-1} + \tau_q/2, t_{q-1} + \tau_q/2). \end{cases}$$

From Assumption 4 let us assume that, with $b = \tau_q/2$ and a given $c \in (0, 1)$, both $K^b_q$ and $K^b_q$ are designed. With $\xi_q := \dot{\xi}_q - \xi_q$, $\xi_q^b := \dot{\xi}_q^b - \xi_q$, and $\beta^b_q := \beta^b_q - \beta^b_q$, it is seen that

$$\max_{t \in [t_{q-1}, t_{q-1} + \tau_q/2)} |\dot{\xi}_q(t)| \leq \sup_{\dot{\xi}_q^b(0) \in [\xi_{q-1}, \xi_{q-1} + \tau_q/2)} |\dot{\xi}_q^b(t)| \leq \max_{t \in [t_{q-1}, t_{q-1} + \tau_q/2)} |\dot{\xi}_q^b(t)| \leq \sup_{t \in [t_{q-1}, t_{q-1} + \tau_q/2)} |\dot{\xi}_q^b(t)| \leq \beta^b_q(|\dot{\xi}_q^b(t_{q-1})|, \tau_q, t_{q-1} - t_q)$$

Therefore, implementation of the observers in (19) and (22) leads to

$$\sup_{t \in [t_{q-1}, t_q)} |\dot{\xi}_q(t)| \leq c|\dot{\xi}_q^b(t_{q-1})|$$

$$c|\dot{\xi}_q^b(t_{q-1})| e^{\lambda_{q-1, q}(\dot{x}(t_{q-1}))) - \lambda_{q-1, q}(\dot{x}(t_{q-1})))} |(\xi_{q-1}(t_{q-1}) - \xi_{q-1}(t_{q-1}))|.$$  

**Theorem 2:** Under the assumptions so far, there is $c^* > 0$ such that, for each $c \in (0, c^*)$, the inequality (14) holds. This theorem concludes the asymptotic error convergence to zero. (Proof is omitted but available from the authors.)

**Acknowledgement**

The authors are grateful to Daniel Liberzon for helpful discussions on this paper.

**References**


