Observability through delayed measurements: a new approach to state observers design

Filippo Cacace\textsuperscript{a}; Alfredo Germani\textsuperscript{b}; Costanzo Manes\textsuperscript{b}
\textsuperscript{a} CIR - Centro Integrato di Ricerca, Università Campus Bio-Medico di Roma, Rome, Italy
\textsuperscript{b} Dipartimento di Ingegneria Elettrica e dell'Informazione, Università degli Studi dell'Aquila, L'Aquila, Italy

First published on: 29 October 2010

To cite this Article Cacace, Filippo , Germani, Alfredo and Manes, Costanzo(2010) 'Observability through delayed measurements: a new approach to state observers design', International Journal of Control,, First published on: 29 October 2010 (iFirst)

To link to this Article DOI: 10.1080/00207179.2010.528035
URL: http://dx.doi.org/10.1080/00207179.2010.528035
Observability through delayed measurements: a new approach to state observers design

Filippo Cacace*, Alfredo Germaniab and Costanzo Manesb

aCIR - Centro Integrato di Ricerca, Università Campus Bio-Medico di Roma, Rome, Italy; bDipartimento di Ingegneria Elettrica e dell’Informazione, Università degli Studi dell’Aquila, Via G. Gronchi, 18, 67100, L’Aquila, Italy

(Received 9 February 2010; final version received 27 September 2010)

An original approach for the construction of state observers for dynamic systems is presented in this article. New notions of observability (delay-observability and delay-drift-observability), useful for the development of the observer equations and for the convergence proof, are introduced and compared with other standard observability concepts. In particular, it is shown that delay-drift-observability is an extension of the well-known drift-observability property. The construction of the proposed observer is straightforward and does not require any coordinate transformation of the system into a canonical form, which is a common feature of many other approaches. A numerical example is reported, where a system that is not drift-observable is considered. In this case, the proposed observer is easily constructed and performs well.

Keywords: state observers; nonlinear systems; delayed measurements

1. Introduction

It is known that state observers are useful in virtually all monitoring and control applications of the systems engineering, and for this reason observability theory and observer construction methods for nonlinear systems are currently very active research areas. Although many techniques have been proposed in the literature to solve the problem of state observer design, all existing methods have their own advantages, disadvantages and applicability restrictions. For these reasons, new approaches are constantly proposed and new classes of systems are considered by authors.

A largely studied approach to state observer design is to consider nonlinear systems in a suitable Observer Canonical Form (OCF) (linear model up to an output injection term) that allows a simplified observer design, by exploiting linear or almost-linear techniques (Bestle and Zeitz 1983; Krener and Respondek 1985; Keller 1987, 1988; Xia and Gao 1989; Hou and Pugh 1999; Noh, Jo, and Seo 2004). Conditions for the existence of a change of coordinates that transforms nonlinear systems in the OCF are studied in Krener and Isidori (1983), Krener and Respondek (1985), Krener and Xiao (2002). An Adaptive Observer Canonical Form (AOCF) is considered in Bastin and Gevers (1988), suitable for the design of adaptive observers (Marino 1990; Marino and Tomei 1992, 1995). Conditions for the existence of an AOCF are studied in Marino (1990). Weaker conditions ensure the existence of other kinds of canonical forms that can be exploited for the observer construction. Triangular canonical forms are considered in Gauthier, Hammouri, and Othman (1992), Gauthier and Kupka (1994), while block triangular canonical form are dealt in Rudolph and Zeitz (1994), Hou, Busawon, and Saif (2000), Lee and Park (2003), Shim and Seo (2003) and Wang and Lynch (2006). State observers for systems that are fully linearisable are developed in Esfandiari and Khalil (1992), while the case of input–output linearisable with stable zero-dynamics is considered in Jo and Seo (2000). The gain of observers designed in all these weaker canonical forms can be made large enough to dominate the system nonlinearities (high-gain observers). These approaches suffer of two main problems, both related with the necessity to design the observer in a transformed coordinate system. First of all, observability conditions, in general, are not sufficient for the existence of the above-listed canonical forms, and therefore the observer construction cannot be applied to all observable systems. Second, even when the existence conditions are satisfied, the nonlinear transformation that performs the change of coordinates, and its inverse, in general cannot be computed in a closed form.

The approach in Ciccarella, Dalla Mora, and Germani (1993), under suitable observability and relative degree conditions, allows for the construction of Luenberger-like observers in the original system coordinates (see also Andrieu and Praly 2006).
When the relative degree condition is not satisfied, conditions on the input can be given to ensure convergence (Dalla Mora, Germani, and Manes 2000), or a further assumption of stability of the zero-dynamics is required (Diao and Yan 2008).

Other approaches allow the state observer construction in the original system coordinates. In many papers, the nonlinear systems is regarded as a linear one forced by nonlinear perturbation terms. In most cases the proposed observer has a classical feedback structure, where the design of the constant gain depends on the Lipschitz constant or on some bounds on the slope of the nonlinear perturbation. Convergence conditions are given in terms of Lyapunov equations, algebraic Riccati equations or LMI (see, e.g. Thau 1973; Raghavan and Hedrick 1994; Rajamani 1998; Kreisselmeier and Engel 2003; Hu 2006; Pertew, Marquez, and Zhao 2006; Zemouche, Boutayeb, and Bara 2008; Zhao, Tao, and Shi 2010). Also the methods of Tsinias (1989, 1990), Arcak and Kokotovic (2001) and Fan and Arcak (2003) belong to this class.

Recently, methods of observer design based on the construction of invariant and attractive manifolds have been proposed (Rapaport and Maloum 2004; Karagiannis, Carnevale, and Astolfi 2008; Karagiannis, Sassano, and Astolfi 2009), and constructive results are provided for some classes of systems.

In this article an original approach is investigated for the construction of state observers, that operates in the original system coordinates, without the need of a canonical form. The goal is to design an observer as a dynamic system forced by an extended measurement vector made up of the system outputs measured at delayed time instants. To this purpose, a suitable observability property is defined, which extends the drift-observability used in Dalla Mora et al. (2000). The observability of nonlinear systems through delayed measurements, that may be seen as a special case of observability for time-delay systems with only output delays, has already been investigated in several papers (see, e.g. Germani, Manes, and Pepe 2001; Xia, Marquez, Zagalak, and Moog 2002; Richard 2003). In this context, we offer additional insights on the relationship between observability with and without measurement delays.

The idea of using artificial delays to increase design flexibility is not new, for example it has been used in the field of control design for linear systems (Seuret, Edwards, Spurgeon, and Fridman 2009) and of state observation of nonlinear systems (Chang, Lee, and Park 1997; Chang and Park 1998). The approach of Chang et al. (1997) is however applicable only to SISO systems in triangular form. The observer presented here is more general, in that it is not limited to systems in triangular form and allows more flexibility in the choice of the observer parameters.

For simplicity, the observability concepts and the observer construction presented in this article are restricted to the case of single-output unforced systems. The aim of this restriction is to keep the notation as simple as possible. However, all the results presented here can be easily extended to multiple-output systems while the extension to the case of systems with external inputs is the subject of future work.

This article is organised as follows. In Section 2, after some preliminaries, the standard observability concept of nonlinear systems is discussed, and the new concepts of delay-observability and delay-drift-observability are introduced and discussed. In Section 3 an observer for drift-observable system is recalled and discussed, while in Section 4 the original observer is presented, its convergence properties are proved and some implementation issues are discussed. A numerical example is reported in Section 5. Conclusions follow in Section 6.

2. Observability of time invariant nonlinear systems

2.1 Preliminaries and notation

We consider autonomous time invariant nonlinear systems \( S \) where the state \( x(t) \in \mathbb{R}^n \) and the output \( y(t) \in \mathbb{R} \) obey equations of the type

\[
\dot{x}(t) = f(x(t)), \quad (1)
\]

\[
y(t) = h(x(t)), \quad (2)
\]

where \( f(x) \) is a \( C^k(\mathbb{R}^n) \) vector field and the function \( h(x) \) is \( C^k(\mathbb{R}^n) \) too, with \( k \) an integer that allows all differentiations needed in this article (in short, \( f \) and \( g \) are smooth functions). The system is assumed to be forward and backward complete over all \( \mathbb{R} \).

The property of observability of a system \( S \) of the type (1)–(2) concerns the possibility of reconstructing the system state \( x(t) \) over a finite time interval by means of the knowledge of the output \( y(t) \) over the same interval. It is known that for nonlinear systems forced by an input function, the observability property may depend on the input applied, and that observability independently of the inputs is a rather strong property, which also denoted uniform observability (Gauthier and Bornard 1981). In this article we restrict the attention to autonomous systems, but the notions of observability that we define here can be suitably extended to systems with a given input or for all inputs in some class.

In order to formalise the observability property of a system, we need to recall the concept of transition
function associated to a stationary system of the type (1)–(2), denoted, \( \varphi(t, x) \), where \( t \in \mathbb{R} \) and \( x \in \mathbb{R}^n \). The transition function allows to write the state trajectory \( x(t) \) that passes at \( x_0 \in \mathbb{R}^n \) at time \( t_0 \), i.e. such that \( x(t_0) = x_0 \), as

\[
x(t) = \varphi(t - t_0, x_0),
\]

for any \( t \geq t_0 \) or \( t < t_0 \), due to the assumed forward and backward completeness. The transition function is such that

\[
\varphi(t - t_0, x_0) = \varphi(t - \bar{t}, \bar{x}(t - t_0, x_0)) \quad \forall t, \bar{t}, t_0 \in \mathbb{R}, \quad \forall x_0 \in \mathbb{R}^n.
\]

and that

\[
\frac{\partial}{\partial t} \varphi(t - t_0, x_0) = f(\varphi(t - t_0, x_0)) \quad \forall t, t_0 \in \mathbb{R}^n, \quad \forall x_0 \in \mathbb{R}^n.
\]

From (3) and (4), for a given \( \delta \in \mathbb{R} \), on a state trajectory \( x(t) = \varphi(t - t_0, x_0) \) it is

\[
x(t + \delta) = \varphi(\delta, x(t)) \quad \forall t \in \mathbb{R}.
\]

The state sensitivity matrix is defined as

\[
S(\delta, x) = \frac{\partial \varphi(\delta, x)}{\partial x},
\]

and is such that \( S(0, x) = L_n \) (identity matrix in \( \mathbb{R}^{n \times n} \)) for all \( x \in \mathbb{R}^n \), and

\[
f(\varphi(\delta, x)) = S(\delta, x) f(x), \quad \text{and} \quad \dot{x}(t + \delta) = S(\delta, x(t)) \dot{x}(t).
\]

It is known that existence and uniqueness of the solution of (1) implies that \( S(\delta, x) \) is nonsingular for all \( \delta \in \mathbb{R} \) and \( x \in \mathbb{R}^n \).

Some observability concepts can be formalised by using the concept of Lie derivatives of a \( C^k \) function \( \sigma(x) \) along a vector field \( f(x) \), denoted \( L_f \sigma(x) \):

\[
L_f \sigma(x) = \frac{d\sigma(x)}{dx} f(x) = \sum_{i=1}^{n} \frac{\partial \sigma(x)}{\partial x_i} f_i(x).
\]

The symbol \( L_f^k \sigma(x) \) means the \( k \)-times repeated iteration of \( L_f \sigma(x) \), i.e. for \( k > 1 \), it is

\[
L_f^k \sigma = L_f (L_f^{k-1} \sigma).
\]

The \( k \)-times repeated Lie derivative applied to the output map \( h(x) \) of system (1)–(2) along \( f(x) \), computed on a state trajectory \( x(t) \), provides the \( k \)-th derivative of the output function, i.e.

\[
L_f^k h(x(t)) = y^{(k)}(t).
\]

The composition of the Lie derivative \( L_f^k h(x) \) with the transition function \( \varphi(\delta, x) \) gives

\[
L_f^k h(\varphi(\delta, x(t))) = y^{(k)}(t + \delta).
\]

Straightforward computations provide the following identity:

\[
\frac{\partial}{\partial \delta} L_f^k h(\varphi(\delta, x)) = L_f^k h(\varphi(\delta, x)) h'(\varphi(\delta, x)).
\]

### 2.2 Observability over time intervals

Given a system of the type (1)–(2), and given a time interval \( T > 0 \), let \( \mathcal{Y}_T \) denote the map that associates to any pair \( (\bar{t}, \bar{x}) \), \( \bar{t} \in \mathbb{R} \) and \( \bar{x} \in \mathbb{R}^n \), the system output in the interval \([\bar{t}, \bar{t} + T]\) computed on the unique state trajectory satisfying \( x(\bar{t}) = \bar{x} \). The map \( \mathcal{Y}_T \) is formally defined as

\[
\mathcal{Y}_T : \mathbb{R}^n \to C^k([\bar{t}, \bar{t} + T]; \mathbb{R}),
\]

\[
(\mathcal{Y}_T(\bar{x}))(t) = h(\varphi(t - \bar{t}, \bar{x})), \quad t \in [\bar{t}, \bar{t} + T].
\]

**Definition 2.1:** For a given time interval \( T > 0 \), a system \( S \) of the type (1)–(2) is said to be \( locally T-observable \) at a state \( \bar{x} \in \mathbb{R}^n \) if there exists a neighbourhood of \( \bar{x} \) where the map \( \mathcal{Y}_T \) is injective. The system \( S \) is said to be \( T-observable \) in an open set \( O \subset \mathbb{R}^n \) if the functional \( \mathcal{Y}_T|_O \) is injective. The system \( S \) is said to be \( globally T-observable \) if the functional \( \mathcal{Y}_T \) is injective in all \( \mathbb{R}^n \).

If a system is \( T-observable \) in a given open set \( O \subset \mathbb{R}^n \), then for any \( y \in \mathcal{Y}_T(O) \) the inverse map \( x = \mathcal{Y}_T^{-1}(y) \) is well defined. If the system is globally \( T-observable \), then the inverse map is defined \( \forall y \in \mathcal{Y}_T(\mathbb{R}^n) \). The inverse map provides the system state at the left end-point \( \bar{t} \) of the observation interval. The knowledge of \( x(\bar{t}) \) allows the computation of the state over the whole interval \([\bar{t}, \bar{t} + T]\).

Note that, due to the time-invariance of system (1)–(2), Definition 2.1 provides a notion of observability that is uniform w.r.t. the position of the observation interval \( T \) over the time, i.e. uniform w.r.t. the time \( \bar{t} \). If a system is (locally or globally) \( T-observable \) for any \( T > 0 \), then it can be said \( uniformly observable \) with respect to the length of the observation interval.

**Remark 2.2:** Systems that are not locally \( T-observable \) at some states admit continuous curves of \( strongly indistinguishable \) states (according to the definition of Hermann and Krener (1977)), i.e. curves \( \chi : [-\epsilon, \epsilon] \to \mathbb{R}^n \) such that \( \mathcal{Y}_T(\chi(s))(t) = (\mathcal{Y}_T(\chi(0)))(t) \forall t \in [\bar{t}, \bar{t} + T], \forall s \in [-\epsilon, \epsilon] \).

### 2.3 Observability through output derivatives (drift-observability)

The concept of system observability through output derivatives deals with the possibility to reconstruct the state of a system at a given time \( t \) by means of the
knowledge of some of the output derivatives at the same instant $t$.

For a given integer $k$ let us define the map $\Phi_k: \mathbb{R}^n \rightarrow \mathbb{R}^k$ and its Jacobian as follows:

$$
\Phi_k(x) = \begin{bmatrix}
    h(x) \\
    L_j h(x) \\
    \vdots \\
    L_j^{k-1} h(x)
\end{bmatrix},
$$

$$
Q_k(x) = \frac{d\Phi_k(x)}{dx} = \begin{bmatrix}
    dh(x) \\
    dL_j h(x) \\
    \vdots \\
    dL_j^{k-1} h(x)
\end{bmatrix},
$$

where the symbols $dL_j h(x)$ are used as a shorthand for $dL_j h(x)/dx$. Note that when $k=n$, the map $\Phi_n$ is a square map, and is often denoted as drift-observability map, while its Jacobian $Q_n$ is denoted as drift-observability matrix (Isidori 1995; Dalla Mora et al. 2000). For linear systems, $Q_n$ coincides with the standard observability matrix.

For the system (1)–(2), the map $\Phi_k(x)$, defined in (14), when computed on a system state $x(t) \in \mathcal{O}$ provides the first $k$ output derivatives, from 0 to $k-1$:

$$
Y_k(t) = \begin{bmatrix}
y(t) \\
y^{(1)}(t) \\
\vdots \\
y^{(k-1)}(t)
\end{bmatrix} = \begin{bmatrix}
h(x(t)) \\
L_j h(x(t)) \\
\vdots \\
L_j^{k-1} h(x(t))
\end{bmatrix} = \Phi_k(x(t)).
$$

The (local or global) invertibility of the function $\Phi_n(x)$ w.r.t. $x$ provides the theoretical possibility to reconstruct the state by means of the knowledge of the output derivatives.

**Definition 2.3:** A system $\mathcal{S}$ of the type (1)–(2) is said to be locally drift-observable (i.e. observable through output derivatives) at a point $\bar{x} \in \mathbb{R}^n$ if there exists a neighbourhood of $\bar{x}$ where the map $z = \Phi_n(x)$ is a diffeomorphism, i.e. there exists a $C^\infty$ inverse map $x = \Phi_n^{-1}(z)$. The system $\mathcal{S}$ is said to be drift-observable in an open set $\mathcal{O} \subset \mathbb{R}^n$ if the function $z = \Phi_n(x)$ is a diffeomorphism in all $\mathcal{O}$. The system $\mathcal{S}$ is said to be globally drift-observable if the function $\Phi_n(x)$ is a diffeomorphism in all $\mathbb{R}^n$.

**Remark 2.4:** Although for linear systems the drift-observability property is equivalent to $T$-observability for any $T>0$ (and equivalent to the full-rank condition on the observability matrix), for nonlinear systems drift-observability is stronger than the property of $T$-observability. In particular, it can be proved that drift-observability implies $T$-observability for any $T>0$. This result is evident by considering that the knowledge of a differentiable function over a neighbourhood of a given time $t$ implies the knowledge of all its derivatives at $t$, at least from a theoretical point of view. It follows that if a system is (locally or globally) drift-observable, then the knowledge of the output trajectories over an arbitrarily small neighbourhood of $t$ allows the state reconstruction at time $t$, and this property is exactly the (local or global) $T$-observability for any $T$.

**Remark 2.5:** When a system is drift-observable in an open set $\mathcal{O} \subset \mathbb{R}^n$ (or in all $\mathbb{R}^n$), the drift-observability map $z = \Phi_n(x)$ defines a change of coordinates in $\mathcal{O}$ and its Jacobian $Q_n(x)$ is non-singular in $\mathcal{O}$. Thus, $\rho(Q_n(x)) = n$ in $\mathcal{O}$ is a necessary condition for drift-observability in $\mathcal{O}$, and necessary and sufficient for local drift-observability in $\mathcal{O}$. In general, drift-observability of a system is difficult to check: it must be proved that for all $x_1$ and $x_2$ in $\mathcal{O}$, $x_1 \neq x_2$ implies that $\Phi_n(x_1) \neq \Phi_n(x_2)$. It is easier (but not easy, in general) to prove local drift-observability by testing the non-singularity of the matrix $Q_n(x)$ in open sets of $\mathcal{O}$.

**Remark 2.6:** If the drift-observability matrix $Q_n$ is singular in an open set $U$, any point $x \in U$ admits strongly indistinguishable states in $U$ (as in Remark 2.2, there exists a curve, or a manifold, passing through $x$ made of strongly indistinguishable states).

### 2.4 Observability through delayed output measurements

In this section a new observability concept is introduced, which implies observability over intervals (i.e. $T$-observability).

Let $\mathcal{D}_k = \{\Delta \in \mathbb{R}^n : 0 \leq \Delta_1 < \cdots < \Delta_n\}$ denote the set of all $n$-p.s. of non-negative strictly-increasing time delays. For a given $\delta > 0$, let $\mathcal{D}_k(\delta) = \{\Delta \in \mathcal{D}_k : \Delta_n < \delta\}$ (recall that, by definition, $\Delta_n$ is the largest delay in the array $\Delta$). The notation $h_\Delta(x) = h(\varphi(\Delta, x))$ will be used throughout this article. Recalling that the function $\varphi(\Delta, x)$ operates a backward state transition, i.e., $x(t - \Delta) = \varphi(\Delta, x(t))$, it is $h_\Delta(x(t)) = h(x(t - \Delta))$.

Let the map $H_\Delta: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be defined as

$$
H_\Delta(x) = \begin{bmatrix}
h(\varphi(-\Delta_1, x)) \\
\vdots \\
h(\varphi(-\Delta_n, x))
\end{bmatrix} = \begin{bmatrix}
h_{\Delta_1}(x) \\
\vdots \\
h_{\Delta_n}(x)
\end{bmatrix}.
$$
The map $H_A$ computed over a state trajectory $x(t)$ gives

$$H_A(x(t)) = Y_A(t), \text{ where } Y_A(t) = \begin{bmatrix} y(t - \Delta_1) \\ \vdots \\ y(t - \Delta_n) \end{bmatrix}. \tag{17}$$

The map (16), when invertible, will be called a delay-observability map, because it allows the state reconstruction from the knowledge of a set of delayed sampled outputs. This justifies the following definition.

**Definition 2.7:** A system $S$ of the type (1)–(2) is said to be locally delay-observable at a point $	ilde{x} \in \mathbb{R}^n$ if there exists a set $\Delta$ of $n$ non-negative-strictly-increasing delays and a neighbourhood of $\tilde{x}$ where the delay-observability map $H_0(x)$ defined in (16) is a diffeomorphism, i.e. the inverse map $x = H_A^{-1}(z)$ exists and is $C^\infty$. A system $S$ is said to be delay-observable in an open set $\mathcal{O} \subset \mathbb{R}^n$ if $H_0(x)$ is a diffeomorphism between $\mathcal{O}$ and $H_A(\mathcal{O})$. A system $S$ is said to be globally delay-observable if $H_0(x)$ is a diffeomorphism in all $\mathbb{R}^n$.

Let $Q_A(x)$ denote the Jacobian of the map $H_A(x)$. It takes the following form:

$$Q_A(x) = \frac{dH_A(x)}{dx} = \begin{bmatrix} \frac{dh_{\Delta_1}(x)}{dx} \\ \vdots \\ \frac{dh_{\Delta_n}(x)}{dx} \end{bmatrix} = \begin{bmatrix} \frac{d}{dx} h(\varphi(-\Delta_1, x)) \\ \vdots \\ \frac{d}{dx} h(\varphi(-\Delta_n, x)) \end{bmatrix} = \begin{bmatrix} dh \bigg|_{\varphi(-\Delta_1, x)} S(-\Delta_1, x) \\ \vdots \\ dh \bigg|_{\varphi(-\Delta_n, x)} S(-\Delta_n, x) \end{bmatrix}. \tag{18}$$

A necessary and sufficient condition for the local delay-observability of system (1)–(2) at a point $x$ is that $Q_A(x)$ is nonsingular. Non-singularity of $Q_A(x)$ in an open set $\mathcal{O} \subset \mathbb{R}^n$ is necessary for the local delay-observability in $\mathcal{O}$.

**Proposition 2:** Consider a system $S$ of the type (1)–(2) and a vector $\Delta \in \mathcal{D}_n$ of delays. If the system is delay-observable in $\mathcal{O} \subset \mathbb{R}^n$, for the given $\Delta$, then it is $T$-observable in $\mathcal{O}$ for any $T \geq \Delta_0$. If $\mathcal{O} = \mathbb{R}^n$ the system is globally $T$-observable.

**Proof:** Trivially, the system is $T$-observable when $T \geq \Delta_0$ because for any $t \in \mathbb{R}$ the knowledge of the output function in the interval $[t - \Delta_0, t]$ implies the knowledge of the $n$ outputs $y(t - \Delta_i)$ at all time instants $t - \Delta_i$, $i = 1, \ldots, n$. By assumption of delay-observability for the given $T$, the delay-observability map (17) is invertible and the state $x(t)$ can be computed. \[ \square \]

**Remark 2.8:** It is not difficult to prove that linear systems that are observable (and therefore are drift-observable) are also delay-observable, for some choice of delay vector $\Delta$. In particular, it is sufficient that the delays $\Delta_i$, $i = 1, \ldots, n$ are chosen such that for any pair of distinct eigenvalues $\lambda_a, \lambda_b$ of the system matrix it is $e^{\lambda_a \Delta_i} \neq e^{\lambda_b \Delta_i}$ for any pair $(\Delta_i, \Delta_b)$ of delays in the set $\Delta$.

**Remark 2.9:** Considerations similar to those of Remark 2.6 can be done for systems that are not locally delay-observable in any point of an open set $U$: in this case, the Jacobian $Q_A(x)$ of the delay-observability map $H_A(x)$ is rank-defective in all $U$, and there exist curves of strongly indistinguishable states.

Similar to what happens for drift-observability, where nonsingularity of $Q_A(x)$ in an open set $\mathcal{O}$ is a necessary condition for drift-observability in $\mathcal{O}$, a necessary condition for delay-observability in an open set $\mathcal{O}$ is that there exists a vector $\Delta \in \mathcal{D}_n$ of $n$ delays such that $Q_A(x)$ is nonsingular in $\mathcal{O}$.

The following theorem gives a relationship between necessary conditions of delay-observability and drift-observability.

**Theorem 2.10:** Consider a smooth system $S$ of the type (1)–(2). In all states $x \in \mathbb{R}^n$ where $Q_A(x)$ is nonsingular, there exists $\Delta \in \mathcal{D}_n$ such that $Q_A(x)$ is nonsingular. If $h(x)$ is analytic, then in all open sets of $\mathbb{R}^n$ where $Q_A(x)$ is singular and has constant rank, the matrix $Q_A(x)$ is singular for any $\Delta \in \mathcal{D}_n$.

The proof of Theorem 2.10 is reported in the Appendix.

**Remark 2.11:** Theorem 2.10 does not state an equivalence between necessary conditions of drift- and delay-observability. We have the equivalence of local non-drift-observability and non-delay-observability in open sets when $h(x)$ is analytic, since the first statement of the theorem is equivalent to claim that in all states $x \in \mathbb{R}^n$ where $Q_A(x)$ is singular for any $\Delta \in \mathcal{D}_n$ the drift-observability matrix $Q_A(x)$ is singular. We also have that local drift-observability implies local delay-observability. However, in general, local delay-observability does not imply local drift-observability.

2.5 Observability through delayed output derivatives

Consider a system of the type (1)–(2), together with a vector $\Delta \in \mathcal{D}_m$ of $m$ nonnegative strictly-increasing time delays, and a set $\vec{r} = \{r_1, r_2, \ldots, r_m\}$ of $m \leq n$ integers
such that
\[ \sum_{j=1}^{m} r_j = n. \]  
(19)

Let \( s_i \) denote the pair \((r_i, \Delta_i)\), and let \( \tilde{s} \) denote the set of pairs \( s_i = (r_i, \Delta_i) \), i.e. \( \tilde{s} = \{s_1, \ldots, s_m\} \). For each pair \( s_i = (r_i, \Delta_i) \) let us define the transformation
\[ \tilde{\Phi}_s(x) = \Phi_{r_i}(\phi(-\Delta_i, x)), \]  
(20)
where \( \Phi_{r_i}() \) has been defined in (14). By definition, the function \( \tilde{\Phi}_s(x) \), when computed on a system state \( x(t) \) provides the first \( r_i \) derivatives of the output \( y(t) \), from \( 0 \) to \( r_i - 1 \), computed at time \( t - \Delta_i \), i.e.
\[ \tilde{\Phi}_s(x(t)) = Y_{r_i}(t - \Delta_i), \]  
where
\[ Y_{r_i}(t - \Delta_i) = \begin{bmatrix} y(t - \Delta_i) \\ y^{(1)}(t - \Delta_i) \\ \vdots \\ y^{(r_i-1)}(t - \Delta_i) \end{bmatrix}. \]  
(21)

The functions \( z_i = \tilde{\Phi}_s(x) \), where \( z_i \in \mathbb{R}^r \), for \( i = 1, \ldots, m \), can be used to define a square map \( z = \tilde{\Phi}_z(x) \), with \( z \in \mathbb{R}^m \), as follows:
\[ z = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} = \begin{bmatrix} \tilde{\Phi}_s(x) \\ \vdots \\ \tilde{\Phi}_{s_m}(x) \end{bmatrix}. \]
(22)

The invertibility of the map \( \tilde{\Phi}_z(x) \) allows for the state reconstruction from the knowledge of the delayed output derivatives \( Y_{r_i}(t - \Delta_i), i = 1, \ldots, m \).

For this reason, the square map \( z = \tilde{\Phi}_z(x) \) is denoted here as delay drift-observability map. Note that a delay drift-observability map depends on the choice of the sets \( \tilde{r} \) and \( \Delta \), and therefore is not unique. Then the following definition can be given.

**Definition 2.12:** A system \( S \) of the type (1)–(2) is said to be delay drift-observable in an open set \( O \subset \mathbb{R}^n \) if there exists a set \( \tilde{s} \) of \( m \) pairs \((r_i, \Delta_i)\) satisfying conditions (19), such that the delayed observability map \( \tilde{\Phi}_z(x) \) is a diffeomorphism in \( O \). The system is said to be globally delay drift-observable if \( \tilde{\Phi}_z(x) \) is a diffeomorphism in all \( \mathbb{R}^n \).

Let \( x = \tilde{\Phi}_z^{-1}(z) \) denote the inverse map of \( z = \tilde{\Phi}_z(x) \), well-defined in \( \tilde{\Phi}_z(O) \).

The following proposition is easily proven using arguments similar to those used in the proof of Proposition 2.

**Proposition 3:** A delay drift-observable system \( S \) in an open set \( O \subset \mathbb{R}^n \) is also T-observable in \( O \) for some \( T > 0 \). A globally delay drift-observable system \( S \) is also globally \( T \)-observable for some \( T > 0 \).

**Remark 2.13:** Note that Definition 2.12 of delay drift-observability encompasses both notions of drift-observability and delay-observability (drift-observability, when \( m = 1 \), \( r_i = n \) and \( \Delta_1 = 0 \), and delay-observability when \( m = n \) and \( \tilde{r} = \{1, 1, \ldots, 1\} \)).

A consequence of the definition of the delay drift-observability property in a set \( O \subset \mathbb{R}^n \) is that the Jacobian associated to a delay drift-observability map, defined as
\[ \tilde{Q}_s(x) = \frac{d \tilde{\Phi}_z(x)}{dx}, \]  
(23)
is nonsingular in \( O \). Note that, by definition, the Jacobian \( \tilde{Q}_s(x) \) has the following structure:
\[ \tilde{Q}_s(x) = \begin{bmatrix} \tilde{Q}_{s_1}(x) \\ \vdots \\ \tilde{Q}_{s_m}(x) \end{bmatrix} = \begin{bmatrix} Q_{r_1}(\phi(-\Delta_1, x))S(-\Delta_1, x) \\ \vdots \\ Q_{r_m}(\phi(-\Delta_m, x))S(-\Delta_m, x) \end{bmatrix}, \]  
(24)
where matrices \( Q_{r_i} \) are \( r_i \times n \) matrices defined in (14).

### 3. Observers for drift-observable systems

We summarise in this section some known results concerning drift-observable systems that are needed in the following (see Ciccarella et al. (1993), Dalla Mora et al. (2000) for more details). For a given integer \( k \), let \( (A_k, B_k, C_k) \) denote a Brunowsky triple of dimension \( k \), defined as
\[ A_k = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{k \times k}, \quad B_k = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^k, \]  
(25)
\[ C_k = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^k. \]  
(26)

It is easy to see that the map \( \Phi_k(x) \) and its Jacobian \( Q_k \) defined in (14) are such that
\[ Q_k(x)f(x) = A_k\Phi_k(x) + B_kL_{ij}f(x) \quad \forall k \in \mathbb{N}. \]  
(27)
Drift-observable systems of the type (1)–(2) admit a change of coordinates by means of the drift-observability map \( z = \Phi_n(x) \). The change of coordinates provides the following systems equation, easily proved using Equation (27),

\[
\dot{z}(t) = A_n z(t) + B_n \tilde{d}(z(t)),
\]

(28)

\[
\tilde{y}(t) = C_n z(t),
\]

(29)

where the function \( \tilde{d}(z) \) is defined as

\[
\tilde{d}(z) = L_f^\gamma h(\Phi_n^{-1}(z)).
\]

(30)

The Brunowsky pair \((A_n, C_n)\) of (25)–(26) is observable, so that a gain vector \( K \in \mathbb{C}^n \) can be chosen to arbitrarily assign eigenvalues to the matrix \( A_n - KC_n \) in the complex plane.

Ciccarella et al. (1993) and Dalla Mora et al. (2000) have been shown that there exists a gain \( K \) such that the system

\[
\hat{x}(t) = f(\hat{x}(t)) + \left[ Q_n(\hat{x}(t)) \right]^{-1} K (y(t) - h(x(t)))
\]

(31)

is a (local/global) exponential observer for a system \( S \) of the type (1)–(2), under the following conditions:

- \( S \) is (locally/globally) drift-observable, and the map \( z = \Phi_n(x) \) and its inverse are uniformly Lipschitz (ULDO and GULDO conditions in Dalla Mora et al. (2000)).
- The function \( \tilde{d}(z) \) in (30) is (locally/globally) uniformly Lipschitz.

It is worth noting that the observer (31) does not need the computation of the function \( \tilde{d}(z) \) defined in (30) for the implementation. The proof of the convergence of the observer (31) (reported in Ciccarella et al. (1993), Dalla Mora et al. (2000)) exploits the Lipschitz assumption on \( \tilde{d}(z) \) and the representation (28)–(29) of the system (1)–(2).

4. An observer for delay-drift-observable systems

In this section an observer is presented for nonlinear systems of the type (1)–(2) which exploits the concept of delay drift-observability. A fundamental role in the observer construction and in the convergence proof is played by the delay drift-observability map \( z = \Phi_n(x) \) defined in (22).

4.1 A representation of delay drift-observable systems

Assume that system (1)–(2) is globally delay drift-observable, i.e. there exists a set \( \tilde{s} \) of \( m \) pairs \((r_i, \Delta_i)\), satisfying conditions (19), such that the map \( z = \Phi_n(x) \) is a global diffeomorphism. Let \( x = \Phi_n^{-1}(z) \) denote the global inverse of the delay drift-observability map. For the given array of \( m \) delays \( \Delta \in \mathbb{D}_m \), let us define a \( \Delta \)-delayed output and a \( \Delta \)-delayed state-output transformation for the system (1)–(2) as

\[
Y_A(t) = \begin{bmatrix} y(t - \Delta_1) \\ \vdots \\ y(t - \Delta_m) \end{bmatrix}, \quad H_A(x) = \begin{bmatrix} h(\varphi(-\Delta_1, x)) \\ \vdots \\ h(\varphi(-\Delta_m, x)) \end{bmatrix}.
\]

(32)

The map \( H_A(x) \) is such that on a state trajectory \( x(t) \) the identity \( H_A(x(t)) = Y_A(t) \) holds. Now consider the system (1)–(2) where the \( \Delta \)-delayed (vector) output is substituted to the original output function

\[
\hat{x}(t) = f(x(t)),
\]

(33)

\[
Y_A(t) = H_A(x(t)).
\]

(34)

The delay drift-observability map \( z = \Phi_n(x) \) can be used as a change of coordinates for system (33)–(34). Notice that \( z \) is partitioned into \( m \) blocks, \( z = [z_1 \cdots z_m]^T \), with \( z_i = \Phi_n(x_i) \in \mathbb{R}^n \). Recall that, from (21), for each block it is

\[
z_i(t) = Y_{r_i}(t - \Delta_i) \quad \text{and} \quad y(t - \Delta_i) = C_{r_i} z_i(t).
\]

(35)

Differentiating \( z_i(t) \), it easily follows that

\[
\dot{z}_i(t) = A_{r_i} z_i(t) + B_{r_i} y(t - \Delta_i),
\]

(36)

\[
y(t - \Delta_i) = C_{r_i} z_i(t).
\]

Let us define the following functions:

\[
d_{r_i}(x) = L_f^\gamma h(\varphi(-\Delta_i, x)), \quad \text{and} \quad \tilde{d}_{r_i}(z) = d_{r_i}(\Phi_n^{-1}(z)).
\]

(37)

which are such that, when \( z(t) = \Phi_n(x(t)) \),

\[
\tilde{d}_{r_i}(z(t)) = d_{r_i}(x(t)) = L_f^\gamma h(x(t - \Delta_i)) = y(t - \Delta_i).
\]

(38)

With these definitions, in analogy with (28)–(29), it is possible to write, for each block \( z_i(t), i = 1, \ldots, m \),

\[
\dot{z}_i(t) = A_{r_i} z_i(t) + B_{r_i} \tilde{d}_{r_i}(z(t)),
\]

(39)

\[
y(t - \Delta_i) = C_{r_i} z_i(t).
\]

(40)

It must be noticed that Equation (39), for \( i = 1, \ldots, m \), is not independent, so that the dynamics of each \( z_i(t) \) depends on the whole \( z(t) \) through the nonlinear terms \( \tilde{d}_{r_i}(z(t)) \).
Defining the vector functions
\[
d_\text{f}(x) = \begin{bmatrix} d_{1}(x) \\ \vdots \\ d_{m}(x) \end{bmatrix}, \quad \tilde{d}_\text{f}(z) = \begin{bmatrix} \tilde{d}_{1}(z) \\ \vdots \\ \tilde{d}_{m}(z) \end{bmatrix},
\]
the \( m \) equations ((39) and (40)) can be written in a matrix form as
\[
\dot{z}(t) = A \tilde{z}(t) + B \tilde{d}_\text{f}(z(t)) \tag{42}
\]
\[
Y_A(t) = Cz(t), \tag{43}
\]
where \( A, B \) and \( C \) are Brunovsky block-diagonal matrices defined as
\[
A = \text{diag}_{i=1}^{m} [A_r], \quad B = \text{diag}_{i=1}^{m} [B_r], \quad C = \text{diag}_{i=1}^{m} [C_r].
\]

System (42)–(43) is the representation of (33)–(34) in \( z \)-coordinates.

A useful relationship between \( \dot{z}(t) \) and \( \dot{x}(t) \) can be obtained by differentiating both terms of the identities \( z(t) = \Phi_t(x(t)) \), yielding
\[
\dot{z}(t) = \frac{d}{dt} \Phi_t(x(t)) = \frac{d}{dt} \Phi_t(\varphi(-\Delta, x(t)))
\]
\[
= \frac{d\Phi_t(\xi)}{d\xi} \bigg|_{\xi = \varphi(-\Delta, x(t))} \frac{\partial\varphi(-\Delta, x)}{\partial x} \bigg|_{x = \varphi(x(t))} \frac{dx(t)}{dt}
\]
\[
= Q_r(\varphi(-\Delta, x(t))) S(-\Delta, x(t)) \dot{x}(t) = \tilde{Q}_r(x(t)) \dot{x}(t), \tag{45}
\]
where \( Q_r(x) \) is the Jacobian of \( \Phi_r(x) \) and
\[
\tilde{Q}_r(x) = Q_r(\varphi(-\Delta, x)) S(-\Delta, x).
\]
Equation (45) can be written for the whole \( \dot{z}(t) \) as
\[
\dot{z}(t) = \tilde{Q}_r(x(t)) \dot{x}(t), \tag{47}
\]
where matrix \( \tilde{Q}_r(x) \) is the Jacobian of \( \tilde{\Phi}_t(x) \), defined in (23).

Another useful relation is obtained by exploiting equation (8), written in the form
\[
S(-\Delta, x)f(x) = \theta(\varphi(-\Delta, x)), \quad \text{and Equation (27), written for } k = \tau, \text{ as } Q_{\tau}(\xi)f(\xi) = A_r \Phi_{\tau}(\xi) + B_r L_{\tau} \theta(\xi), \text{ where } \xi = \varphi(-\Delta, x). \text{ Thus, the following identities are easily obtained}
\]
\[
Q_r(\varphi(-\Delta, x)) S(-\Delta, x)f(x) = A_r \Phi_r(\varphi(-\Delta, x)) + B_r L^r \theta(\varphi(-\Delta, x))
\]
\[
= A_r \Phi_r(\varphi(-\Delta, x)) + B_r d_r(x), \tag{48}
\]
From this, taking into account (24) and the definitions of \( A \) and \( B \), it follows that
\[
\tilde{Q}_r(x)f(x) = A \tilde{Q}_r(x) + B \tilde{d}_r(x). \tag{49}
\]
Substitution of \( x = x(t) \) into (49), taking into account (47), yields
\[
\dot{z}(t) = A \tilde{z}(t) + B \tilde{d}_r(x(t)), \tag{50}
\]
which coincides with (42), because \( d_r(x(t)) = \tilde{d}_r(z(t)) \).

4.2 The proposed observer

In this section it is proved that, for a globally delay-drift-observable systems of the type (1)–(2), the system
\[
\dot{\hat{x}}(t) = f(\hat{x}(t)) + [\tilde{Q}_r(\hat{x}(t))]^{-1} K(Y_A(t) - H_A(\hat{x}(t))) \tag{51}
\]
is an exponential observer, with a suitable choice of the gain matrix \( K \in \mathbb{R}^{n \times m} \) in the block diagonal form \( K = \text{diag}_{i=1}^{m} [K_i] \), where \( K_i \in \mathbb{R}^{r_i} \). Before stating and proving the convergence theorem for the observer (51), some lemmas are needed.

Lemma 4: After the change of coordinates \( \tilde{z} = \tilde{\Phi}_t(\hat{x}) \), system (51) takes the form
\[
\dot{\tilde{z}}(t) = A \tilde{z}(t) + B \tilde{d}_r(\tilde{z}(t)) + K(Y_A(t) - C \tilde{z}(t)), \tag{52}
\]
where \( (A, B, C) \) is the block-diagonal Brunowsky triple defined in (44).

Proof: The equivalence of (52) and (51) can be proved by differentiating both terms of the identity \( \tilde{z}(t) = \tilde{\Phi}_t(\hat{x}(t)) \). In analogy with (47), we get
\[
\dot{\tilde{z}}(t) = \tilde{Q}_r(\hat{x}(t)) \dot{\hat{x}}(t). \tag{53}
\]
Substitution of (51) in (53) gives
\[
\dot{\tilde{z}}(t) = \tilde{Q}_r(\hat{x}(t)) f(\hat{x}(t)) + K(Y_A(t) - H_A(\hat{x}(t))). \tag{54}
\]
Following the same steps that have led to (50), we get
\[
\dot{\tilde{z}}(t) = A \tilde{z}(t) + B \tilde{d}_r(\tilde{z}(t)) + K(Y_A(t) - H_A(\tilde{z}(t))). \tag{55}
\]
Recalling that \( H_A(x) = C \tilde{\Phi}_t(x) \) and that \( d_r(\tilde{\Phi}_t^{-1}(z)) = d_r(z) \), when \( \tilde{z}(t) = \tilde{\Phi}_t(\hat{x}(t)) \) the following hold true:
\[
H_A(\tilde{z}(t)) = C \tilde{z}(t) \quad \text{and} \quad d_r(\tilde{z}(t)) = \tilde{d}_r(\tilde{z}(t)). \tag{56}
\]
Substitution of these identities into (55) gives (52). □

In observer (51) the \( m \) matrices \( K_i \in \mathbb{R}^{r_i} \) must be suitably chosen to assign eigenvalues to the matrices \( A_r - K_i C_r \).

The following proposition will prove useful.
Proposition 5: Consider a Brunowsky pair \((A_r, C_r)\) and a set \(\tilde{\lambda}\) of \(r\) distinct eigenvalues, \(\tilde{\lambda} = \{\lambda_1, \ldots, \lambda_r\}\). The gain \(K(\tilde{\lambda}) \in \mathbb{R}^n\) that assigns such eigenvalues to the matrix \(A_r - K(\tilde{\lambda})C_r\) can be computed as follows:

\[
K(\tilde{\lambda}) = -V^{-1}(\tilde{\lambda}) \begin{bmatrix} \lambda_1^r \\ \vdots \\ \lambda_r^r \end{bmatrix}, \quad \text{where}
\]

\[
V(\tilde{\lambda}) = \begin{bmatrix} \lambda_1^{-1} & \cdots & \lambda_1 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \lambda_r^{-1} & \cdots & \lambda_r & 1 \end{bmatrix},
\]

and

\[
V(\tilde{\lambda})(A_r - K(\tilde{\lambda})C_r)V^{-1}(\tilde{\lambda}) = \Lambda,
\]

where \(\Lambda = \text{diag}(\tilde{\lambda})\). Moreover, the choice \(\lambda_k = -w^k\), i.e. \(\lambda(w) = \{-w, -w^2, \ldots, -w^r\}\), is such that

\[
\lim_{n \to \infty} \|V^{-1}(\tilde{\lambda}(w))\| = 1.
\]

Proof: Direct computations show that the Vandermonde matrix \(V(\tilde{\lambda})\) is such that \(V(\tilde{\lambda})(A_r - K(\tilde{\lambda})C_r) = \Lambda V(\tilde{\lambda})\), from which (58) follows. This proves that matrix \(A_r - K(\tilde{\lambda})C_r\) is similar to \(\Lambda\), and therefore has the same eigenvalues \(\tilde{\lambda}\). The proof of the limit (59) can be found in Ciccarella et al. (1993).

Theorem 4.1: Consider a system \(S\) of the type (1)–(2), and assume that \(S\) is delay drift-observable for given sets of \(m\) integers \(\tau = \{r_1, \ldots, r_m\}\), such that \(\sum_{i=1}^m r_i = n\), and of \(m\) delays \(\Delta \in \mathbb{T}_m\).

Assume that

\((H_1)\) The delay drift-observability map \(z = \tilde{\Phi}_\tau(x)\) and its inverse \(x = \tilde{\Phi}_\tau^{-1}(z)\) associated to the sets \(\tau\) and \(\Delta\) are globally uniformly Lipschitz, i.e.

\[
\|\tilde{\Phi}_\tau(x) - \tilde{\Phi}_\tau(z)\| \leq \gamma_{\tilde{\Phi}_\tau}\|x - z\| \quad \forall x, z \in \mathbb{R}^n,
\]

\((H_2)\) The functions \(\tilde{d}_i(z) = L_{i,z}h(\psi(\Delta, \tilde{\Phi}_\tau^{-1}(z)))\) defined in (39) are uniformly Lipschitz w.r.t. \(z\), i.e.

\[
\|\tilde{d}_i(z) - \tilde{d}_i(\tilde{z})\| \leq \gamma_{\tilde{d}_i}\|z - \tilde{z}\| \quad \forall z, \tilde{z} \in \mathbb{R}^n.
\]

Let the measurements \(y(t)\) be available starting at time \(t_0 = 0\).

Then, for any \(\alpha > 0\) there exist \(m\) gain matrices \(K_i \in \mathbb{R}^n\) and a strictly positive constant \(\mu\) such that the state observation \(\hat{x}(t)\) provided by the following system:

\[
\dot{\hat{x}}(t) = f(\hat{x}(t)), \quad t \in [0, \Delta_m),
\]

\[
\dot{\hat{x}}(t) = f(\hat{x}(t)) + [\tilde{G}(\hat{x}(t))]^{-1} \tilde{K}(Y_\Delta(t) - H_\Delta(\hat{x}(t)));
\]

\(t \geq \Delta_m,
\]

with \(\tilde{K} = \text{diag}_{i=1}^m \{K_i\}\), is such that, for \(t \geq \Delta_m\), the observation error obeys the inequality

\[
\|x(t) - \hat{x}(t)\| \leq \mu e^{-\alpha(t-\Delta_m)}\|\psi(\Delta_m, x(0)) - \psi(\Delta_m, \hat{x}(0))\|
\]

\[\forall x(0), \hat{x}(0) \in \mathbb{R}^n,
\]

and therefore (63)–(64) is an exponential observer for the system \(S\).

Proof: Consider both system and observer in \(z\)-coordinates, Equations (42) and (52). Let \(\varepsilon(t)\) denote the observation error in \(z\)-coordinates:

\[
\varepsilon(t) = z(t) - \tilde{z}(t).
\]

The derivative \(\dot{\varepsilon}(t)\), for \(t \geq \Delta_m\), is computed by subtracting (52) from (42)

\[
\dot{\varepsilon}(t) = z(t) - \tilde{z}(t)
\]

\[
= A(z(t) - \tilde{z}(t)) + B(\tilde{d}_i(z(t)) - \tilde{d}_i(\tilde{z}(t)))
\]

\[
+ \tilde{K}(Cz(t) - \tilde{C}z(t))
\]

\[
(\text{the substitution } Y_\Delta(t) = Cz(t) \text{ has been done in (52)).}
\]

Let us define the function

\[
\delta_i(z, \tilde{z}) = \tilde{d}_i(z) - \tilde{d}_i(\tilde{z}).
\]

Collecting terms in (67), the error dynamics can be written as follows:

\[
\dot{\varepsilon}(t) = (A - \tilde{K}C)\varepsilon(t) + B\delta_i(z(t), \tilde{z}(t)), \quad t \geq \Delta_m.
\]

In order to prove the existence of a gain \(\tilde{K}\) as to guarantee the exponential convergence of the observation error (inequality (65)), let us choose the \(m\) gains \(K_i \in \mathbb{R}^n\) as to assign sets of eigenvalues of the type \(\lambda_i(w) = \{-w, -w^2, \ldots, -w^r\}\) to the matrices \(A_i - K_iC_i\), where \(w > 1\) is a design parameter to be chosen. Let \(\tilde{V}(w) = \text{diag}_{i=1}^m \{V(\lambda_i(w))\}\), where \(V(\cdot)\) is the Vandermonde matrix defined in Proposition 5, and define the variable \(\eta(t) = \tilde{V}(w)\varepsilon(t)\), which obeys the differential equation

\[
\dot{\eta}(t) = \tilde{V}(w)(A - \tilde{K}C)(\tilde{V}(w))^{-1} \eta(t) + \tilde{V}(w)B\delta_i(z(t), \tilde{z}(t)).
\]

It can be easily shown (Proposition 5) that \(\tilde{V}(w)(A - \tilde{K}C)(\tilde{V}(w))^{-1} = \text{diag}_{i=1}^m \{A_i(w)\}\), where \(A_i(w) = \text{diag}_{i=1}^m \{-w^i\}\). Let \(\mathbf{1}_r \in \mathbb{R}^r\) denote a vector whose components are all 1. The vector \(\mathbf{1}_r\) coincides
with the last column of \( V(\tilde{x}_r(w)) \) (see (57)), and therefore \( V(\tilde{x}_r(w))B_n = I_p \). Let \( I_p = \text{diag}^{m}_1 \{ I_p \} \). It follows that \( \bar{V}(w)B = \text{diag}^{m}_1 \{ V(\tilde{x}_r(w))B_n \} = I_p \). Thus, defining \( \Lambda_r(w) = \text{diag}^{m}_1 \{ \Lambda_r(w) \} \), Equation (70) can be written as

\[
\dot{\eta}(t) = \Lambda_r(w) \eta(t) + I_p \delta(z(t), \tilde{z}(t)).
\]

(71)

Let \( v(t) = \| \eta(t) \|^2 \). It is

\[
\dot{v}(t) = 2\eta^T(t) \dot{\eta}(t)
\]

\[
= 2\eta^T(t) \Lambda_r(w) \eta(t) + 2\eta^T(t) I_p \delta(z(t), \tilde{z}(t)).
\]

(72)

Being \( w > 1 \), it is \( \eta^T \Lambda_r(w) < -w \| \eta \|^2 \). Moreover,

\[
2\eta^T I_p \delta_z \leq \| \eta \|^2 + \| I_p^{T} I_p \| \gamma_d \| z - \tilde{z} \|^2.
\]

(73)

and because of the Lipschitz assumption (62) it is

\[
\| \delta(z, \tilde{z}) \| \leq \gamma_d \| z - \tilde{z} \|,
\]

where \( \gamma_d = \max^m_{i=1} \{ \gamma_d \} \), which follows

\[
2\eta^T I_p \delta_z \leq \| \eta \|^2 + \| I_p^{T} I_p \| \gamma_d \| z - \tilde{z} \|^2.
\]

(74)

From the identity \( I_p^{T} I_p = \text{diag}^{m}_1 \{ r_i \} \) it follows \( \| I_p^{T} I_p \| = \bar{r} \), where \( \bar{r} = \max^m_{i=1} \{ r_i \} \). Using (71) and (74) the following inequality is easily computed:

\[
\dot{v}(t) \leq (-2w + 1) v(t) + 2\bar{r} \gamma_d \| \varepsilon(t) \|^2.
\]

(75)

Being \( \eta(t) = \bar{V}(w) \varepsilon(t) \), it is \( \| \varepsilon(t) \|^2 \leq \| (\bar{V}(w))^{-1} \|^2 \times \| \eta(t) \|^2 \), and therefore

\[
\dot{v}(t) \leq \left( -2w + 1 + 2\bar{r} \gamma_d \| (\bar{V}(w))^{-1} \|^2 \right) v(t), \quad t \geq \Delta_m.
\]

(76)

Since \( \| (\bar{V}(w))^{-1} \| = \text{diag}^{m}_1 \{ \| V^{-1}(\tilde{x}_r(w)) \| \} \), it follows that \( \lim_{w \to \infty} \| (\bar{V}(w))^{-1} \| = 1 \) (Proposition 5). Therefore, for any \( \alpha > 0 \), there exists a finite \( w > 1 \) such that

\[
-2w + 1 + 2\bar{r} \gamma_d \| (\bar{V}(w))^{-1} \|^2 \leq -2\alpha, \quad \forall w > \tilde{w}.
\]

(77)

Thus, for \( w \geq \tilde{w} \)

\[
\dot{v}(t) \leq -2\alpha v(t), \quad \forall t \geq \Delta_m,
\]

(78)

from which

\[
v(t) \leq e^{-2\alpha (t-\Delta_m)} v(\Delta_m), \quad \forall t \geq \Delta_m.
\]

(79)

Being \( v(t) = \| \eta(t) \|^2 \) and \( \eta(t) = \bar{V}(w) \varepsilon(t) \), it follows that

\[
\| \eta(t) \| \leq e^{-\sigma (t-\Delta_m)} \| \eta(\Delta_m) \|
\]

(80)

\[
\| \varepsilon(t) \| \leq e^{-\sigma (t-\Delta_m)} \| \bar{V}(w) \| \| (\bar{V}(w))^{-1} \| \| \varepsilon(\Delta_m) \|.
\]

Moreover, by the Lipschitz assumption \( H_1 \),

\[
\| x(t) - \hat{x}(t) \| = \| \tilde{\Phi}_\gamma^{-1}(z(t)) - \Phi_\gamma^{-1}(\tilde{z}(t)) \|
\]

\[
\leq \gamma_{\Phi} \| z(t) - \tilde{z}(t) \| = \gamma_{\Phi} \| \varepsilon(t) \|,
\]

(82)

\[
\| \varepsilon(\Delta_m) \| \leq \| \tilde{\Phi}_\gamma(x(\Delta_m)) - \Phi_\gamma(\tilde{x}(\Delta_m)) \|
\]

\[
\leq \gamma_{\Phi} \| x(\Delta_m) - \tilde{x}(\Delta_m) \|,
\]

and from these we have

\[
\| x(t) - \hat{x}(t) \| \leq e^{-\sigma (t-\Delta_m)} \gamma_{\Phi} \gamma_{\Phi}^{-1} \| \bar{V}(w) \| \| (\bar{V}(w))^{-1} \| \times \| x(\Delta_m) - \tilde{x}(\Delta_m) \|.
\]

(83)

Thus, inequality (65) follows with

\[
\mu = \gamma_{\Phi} \gamma_{\Phi}^{-1} \| \bar{V}(w) \| \| (\bar{V}(w))^{-1} \|,
\]

(84)

also by noting that \( x(\Delta_m) = \varphi(\Delta_m, x(0)) \) and \( \tilde{x}(\Delta_m) = \varphi(\Delta_m, \tilde{x}(0)) \).

Remark 4.2: The observer based on delayed output measurements (63)–(64) is actually a class of observers, each one based on a specific choice of \( m \) pairs \( s_i = (r_i, \Delta_i) \), as to satisfy (19). In the particular case \( m = 1, \ r_i = n \) and \( \Delta_i = 0, \ i.e. \ s_i = (n, 0) \), the observer becomes the nonlinear observer based on the drift-observability map described in Ciccarella et al. (1993) With the choice \( m = n \) and \( r_i = 1, \ i.e. \ s_i = (1, \Delta_i) \), the equivalence \( H_\Delta(x) = \tilde{\Phi}_\gamma(x) \) is true so that \( n \) delayed measurements are used as forcing terms in the observer. Another interesting case is when \( m = 1 \) and \( \Delta_i > 0 \), since in this case only one delayed measurement is used. The observer can thus work on nonlinear systems with delayed output measurements, as the one presented in Germani, Manes, and Pepe (2002).

The conditions that ensure the convergence of the observer (63)–(64) can be summarised in the delay drift-observability and the global uniform Lipschitz properties of \( \tilde{\Phi}_\gamma(x), \tilde{\Phi}_\gamma^{-1}(z), L_i^{\gamma}(\varphi(-\Delta, \tilde{\Phi}_\gamma^{-1}(z))) \). Even though the latter requirement is difficult to verify and particularly restrictive in practical cases, the global Lipschitz property can be considerably weakened for systems where the state \( x(t) \) is confined into a bounded set of \( \mathbb{R}^n \). In these cases the following property can be established.

**Theorem 4.3:** Consider a system \( S \) of the type (1)–(2), and assume that there exist two compact sets \( \Omega_a \subset \Omega_b \subset \mathbb{R}^n \) such that for any \( x(0) \in \Omega_b \) it is \( x(t) \in \Omega_a \) \( \forall t \geq 0 \). Let \( S \) be delay-drift-observable for a given set of \( m \) integers \( \bar{r} = \{ r_1, \ldots, r_m \} \), such that \( \sum_{i=1}^{m} r_i = n \), and for a given vector \( \delta \in \mathcal{D}_m \) of \( m \) delays.
Moreover, assume that

(H1) The delay drift-observability map \( z = \tilde{\Phi}_t(x) \) and its inverse \( x = \tilde{\Phi}_t^{-1}(z) \) associated to the sets \( \tilde{\Delta} \) and \( \Delta \) are locally Lipschitz.

(H2) The functions \( \tilde{d}_i(z) = L_i f(\varphi(-\Delta_i, \tilde{\Phi}_t^{-1}(z))) \) defined in (39) are locally Lipschitz.

Let the measurements \( y(t) \) be available starting at time \( t_0 = 0 \). Then, for any \( \alpha > 0 \) there exist \( m \) gain matrices \( K_i \in \mathbb{R}^{n \times n} \) and strictly positive constants \( \mu \) and \( \delta \) such that the state observation \( \hat{x}(t) \) provided by (63)–(64) with \( K = \text{diag} \{ K_i \} \) is such that, for \( t \geq \Delta_m \) the observation error obeys the inequality

\[
\| x(t) - \hat{x}(t) \| \leq \mu e^{-\alpha t}\| x(0) - \hat{x}(0) \|,
\]

\[\forall x(0), \hat{x}(0) \in \mathbb{R}^n, \text{ such that } \| \hat{x}(0) - x(0) \| \leq \delta. \tag{85}\]

Theorem 4.3 can be proved by extending Theorem 4.1, following the same procedure used in the proof of Theorem 3.8 in Dalla Mora et al. (2000) to extend the result of Theorem 3.1.

4.3 Implementation issues

This section deals with the issue of the computation of the terms \( \tilde{Q}_i(\hat{x}(t)) \) and \( H_\Delta(\hat{x}(t)) \) in the observer equation (51). Defining the variables \( x_\Delta^i(t) = \varphi(-\Delta_i, \hat{x}(t)) \), which denote the backward evolution from \( \hat{x}(t) \) for a time \( \Delta_i \) along the vector field \( f(x) \), the following formulas hold:

\[
H_\Delta(\hat{x}(t)) = \begin{bmatrix}
    h(x_\Delta^1(t)) \\
    \vdots \\
    h(x_\Delta^m(t))
\end{bmatrix}
\tag{86}
\]

\[
\tilde{Q}_i(\hat{x}(t)) = \begin{bmatrix}
    Q_{i,1}(x_\Delta^1(t))S(-\Delta_i, \hat{x}(t)) \\
    \vdots \\
    Q_{i,m}(x_\Delta^m(t))S(-\Delta_i, \hat{x}(t))
\end{bmatrix}
\tag{87}
\]

Thus, the main problem in the implementation of the observer (63)–(64) is the computation of the variables \( x_\Delta^i(t) \) and of the sensitivity matrices \( S(-\Delta_i, \hat{x}(t)) \), \( i = 1, \ldots, m \), needed for the evaluation of the Jacobian \( \tilde{Q}_i(\hat{x}(t)) \) and of the vector \( H_\Delta(\hat{x}(t)) \).

A method to compute \( x_\Delta^i(t) \) is to use a truncated Taylor series with initial point \( \hat{x}(t) \). To this aim, consider the Taylor expansion of \( \varphi(x, x) \):

\[
\varphi(x, x) = \varphi(0, x) + \sum_{k=1}^{\infty} \frac{x^k}{k!} \frac{\partial^k \varphi(x, x)}{\partial x^k} \bigg|_{x=0}.
\tag{88}
\]

From the identity (5) we have

\[
\frac{\partial \varphi(x, x)}{\partial t} = f(\varphi(x, x)),
\tag{89}
\]

so that, iterating derivatives, the following equalities are easily computed:

\[
\frac{\partial^k \varphi(x, x)}{\partial t^k} = L_j^{k-1} f(\varphi(x, x)).
\tag{90}
\]

Recalling that \( \varphi(0, x) = x \), the Taylor expansion of \( \varphi(x, x) \) can be written as

\[
\varphi(x, x) = x + \sum_{k=1}^{\infty} \frac{x^k}{k!} L_j^{k-1} f(\varphi(x, x)).
\tag{91}
\]

The variables \( x_\Delta^i(t) = \varphi(-\Delta_i, \hat{x}(t)) \) can therefore be rewritten as

\[
x_\Delta^i(t) = \hat{x}(t) + \sum_{k=1}^{\infty} \frac{(-\Delta_i)^k}{k!} L_j^{k-1} f(\hat{x}(t)).
\tag{92}
\]

while the polynomial approximation of degree \( v \) takes the form

\[
\hat{x}(t) = \hat{x}(t) + \sum_{k=1}^{v} \frac{(-\Delta_i)^k}{k!} L_j^{k-1} f(\hat{x}(t)).
\tag{93}
\]

The remainder in the Lagrange form provides the following error estimate:

\[
\| x_\Delta^i(t) - \hat{x}(t) \| \leq \frac{(\Delta_i)^{v+1}}{(v+1)!} \| L_j^v f(\varphi(t, \hat{x}(t))) \|.
\tag{94}
\]

Differentiating (91) with respect to \( x \) provides the Taylor expansion of the sensitivity matrix \( S(x, x) \):

\[
S(x, x) = I_n + \sum_{k=1}^{\infty} \frac{x^k}{k!} \left( \frac{d}{dx} L_j^{k-1} f(x) \right)
\tag{95}
\]

from which the polynomial approximations of degree \( v \) of the sensitivities \( S(-\Delta_i, \hat{x}(t)) \) can be computed as

\[
\tilde{S}(-\Delta_i, \hat{x}(t)) = I_n + \sum_{k=1}^{v} \frac{(-\Delta_i)^k}{k!} \left( \frac{d}{dx} L_j^{k-1} f(x) \right) \bigg|_{x=\hat{x}(t)}.
\tag{96}
\]

Thus, the two terms \( \tilde{Q}_i(\hat{x}(t)) \) and \( H_\Delta(\hat{x}(t)) \) in the observer equation (51) can be approximated using formulas (86)–(87), where the terms \( x_\Delta^i(t) \) are replaced by \( \tilde{x}(t) \) given by (93), and \( \tilde{S}(-\Delta_i, \hat{x}(t)) \) are replaced by \( \tilde{S}(-\Delta_i, \hat{x}(t)) \) given by (96).

4.4 Motivations for delay-drift-observers

Even though drift-observers are a particular design within the more general class of delay-drift-observers, one may question the practical utility of a more general and complicated design in the light of a somewhat
Theorem 2.10). In our view, the use of delay-drift-observer can be motivated by a number of reasons:

- A delay-drift-observer design does not require the output function \( h(x) \) to be smooth. Indeed, it is sufficient for \( h(x) \) to be a \( C^1(\mathbb{R}^2) \) function. This aspect may be useful in many cases, for instance when \( h(x) \) is artificially constructed as a piecewise smooth function that approximate a sensor characteristic obtained by experiments.

- The additional complexity related to the polynomial approximation of \( \phi(-\Delta, x) \) and \( S(-\Delta, x) \) discussed in Section 4.3 is balanced by the fact that it is no more necessary to compute the differentials \( dL^i_j h(x) \), for \( i=0, \ldots, n-1 \), that make up the drift-observability matrix \( Q_h(x) \). For a system with many state variables, i.e. when \( n \) is large, the analytical expression of \( Q_h(x) \) may become too complex or large even using tools for symbolic computations. The implementation of a delay drift-observer, on the other hand, only requires the computation of \( dL^i_j f(x)/dx \) up to an integer \( v \) that depends on the size of the chosen delays, and not on the size of the system. It is therefore possible to use small delays for which a value of \( v \) as small as 1 or 2 is sufficient.

- The extension of the delay-drift-approach to non-autonomous systems would allow to overcome the limitations imposed by a low relative degree \( r \) (Isidori 1995) to the construction of drift-observers (see Ciccarella et al. 1993; Dalla Mora et al. 2000). It is known that when \( r<n \) the drift-observability map does not allow the computation of the first \( n \) output derivatives as a function of the state, in that also the input on some of its time-derivatives are involved in the computation of output derivatives of order \( k \geq r \) (and input derivatives are not available, in general). On the other hand, by choosing a set of integers \( \bar{r} = \{r_1, \ldots, r_m\} \) such that \( \sum_{i=1}^m r_i = n \) and \( r_i \leq r \) for all \( i=1, \ldots, m \), it is possible to write a map from the state to delayed output derivatives, without requiring the knowledge of input derivatives. As a consequence, if such a map is invertible, a delay-drift-observer can be designed without making use of input derivatives, thus overcoming the problem of low relative degree.

\[
\lambda(\zeta) = \begin{cases} 
(a - b\zeta^2)(\zeta - \bar{\zeta}) + \bar{a} - \frac{b\bar{\zeta}^2}{3}, & \zeta \geq \bar{\zeta} \\
\bar{a}\zeta - \frac{b}{3}\bar{\zeta}^3, & |\zeta| < \bar{\zeta} \\
(a - b\bar{\zeta}^2)(\bar{\zeta} + \bar{\zeta}) - \frac{b\bar{\zeta}^2}{3}, & \zeta \leq -\bar{\zeta}.
\end{cases}
\]

(97)

The function \( \lambda(\zeta) \) is plotted in Figure 1 (note the discontinuity of the second derivative). Let \( \lambda(\zeta) \) be the measurement function of a simple three-species Volterra-like system

\[
\begin{align*}
\dot{x}_1(t) &= -x_1(t) + \beta_1 x_1(t)x_2(t) + \beta_2 x_1(t)x_3(t) \\
\dot{x}_2(t) &= x_2(t) - \mu_2 x_1(t)x_2(t) \\
\dot{x}_3(t) &= x_3(t) - \mu_3 x_1(t)x_3(t) \\
y(t) &= h(x) = \lambda(x_2(t) + x_3(t)).
\end{align*}
\]

(98)
For this system the drift-observability map $\Phi_d(x)$ is not defined everywhere, because $\frac{dL^2}{h(x)}$ does not exist in the set $|x_2 + x_3| = \xi$. As a consequence, the observer (31) cannot be implemented. Since system (98) satisfies the hypothesis of Theorem 4.3, we may design an observer of the type (51), by choosing $m = 3$ and $\tilde{r} = \{1, 1, 1\}$ with $\Delta \in \mathcal{D}_3$. Choosing $\Delta_1 = 0$, it is $\tilde{s} = \{(1,0), (1,\Delta_2), (1,\Delta_3)\}$, and the the map $\Phi_d(x)$ is as follows:

$$
\tilde{\Phi}_d(x) = \begin{bmatrix}
  h(x) \\
  h(\phi(-\Delta_2, x)) \\
  h(\phi(-\Delta_3, x))
\end{bmatrix}, \quad \text{so that}
$$

$$
\tilde{\Phi}_d(x(t)) = \begin{bmatrix}
  y(t) \\
  y(t - \Delta_2) \\
  y(t - \Delta_3)
\end{bmatrix}.
$$

(99)

The delay-observability $\tilde{Q}_d(x)$ takes the following form:

$$
\tilde{Q}_d(x) = \begin{bmatrix}
  \frac{dh(x)}{dx} \\
  \frac{dh(\phi(-\Delta_2, x))}{dx} S(-\Delta_2, x) \\
  \frac{dh(\phi(-\Delta_3, x))}{dx} S(-\Delta_3, x)
\end{bmatrix},
$$

where, recalling that $h(x) = \lambda(x_2 + x_3)$, it is

$$
\frac{dh}{dx} = \frac{d\lambda}{d\xi} \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}, \quad \text{with}
$$

$$
\frac{d\lambda}{d\xi} = \begin{cases} 
  a - b\xi^2, & |\xi| < \xi \\
  a - b\xi^2, & |\xi| \geq \xi,
\end{cases}
$$

(100)

is continuous but not differentiable in the set $|x_2 + x_3| = \xi$. The simulation reported in Figure 2 refers to the system (98) with $\beta_1 = 1$, $\beta_2 = 0.3$, $\mu_2 = 0.8$, $\mu_3 = 0.4$. The terms $x^*(t) = \varphi(-\Delta, \tilde{x}(t))$ and $S(-\Delta, \tilde{x}(t))$ in the observer equations have been computed by the following Taylor polynomials:

$$
\tilde{x}(t) = \tilde{x}(t) - \Delta f(\tilde{x}(t)) + \frac{\Delta^2}{2} L_f(\tilde{x}(t)) - \frac{\Delta^3}{3!} L^2_f(\tilde{x}(t))
$$

$$
\tilde{S}(-\Delta, \tilde{x}(t)) = I_3 - \Delta \left( \frac{dL^2(f(x))}{dx} \right)_{x = \tilde{x}(t)} + \frac{\Delta^2}{2} \left( \frac{dL^2(f(x))}{dx} \right)_{x = \tilde{x}(t)}
$$

(101)

The plots in Figure 2 include two distinct sets of delays, $\Delta_2 = 0.1$, $\Delta_3 = 0.2$ and $\Delta_2 = 0.2$, $\Delta_3 = 0.4$. The observer gain $\mathcal{R}$ has been chosen so to assign eigenvalues $(-2, -2.1, -2.2)$ to the matrix $A - \mathcal{R}C$, with initial conditions $x(0) = [0.7 \ 0.5 \ 0.6]^T$ and
\( \dot{x}(0) = [2 \ 1.5 \ 2.5]^T \). It can be noticed that, at the same gain, the convergence of the observer is slower when the delay is larger. The choice of the parameters of \( \lambda(\gamma) \) is the same as in Figure 1, in particular \( \zeta = 3 \). From Figure 2 it can be noticed that the transient change in the dynamics of the observers (e.g. at \( t = 2 \)), corresponding to the points where \( x_2 + x_3 \) crosses the \( \zeta \) threshold.

6. Conclusions

This article presents an original approach for the construction of state observers for dynamic systems, which exploits the concepts of delay-observability and of delay-drift-observability of systems. These concepts formalise the theoretical possibility of state reconstruction from the knowledge of the output and of some of its derivative at a finite number of time instants, and is strictly related with the standard observability property (observability over finite time intervals) and with the drift-observability property. The observer construction is straightforward and does not require any coordinate transformation of the system into a canonical form. A numerical example is reported which demonstrates that the proposed observer can be constructed and performs well even in cases where the output function is not smooth, so that the drift-observability condition cannot be fulfilled and observers based on the drift-observability property cannot be constructed.

References


Appendix A: Proofs of Theorem 2.10

In this section, the symbol $B_{\epsilon}(x)$ denotes the open ball of radius $\epsilon$ centred at $x$, and $\rho(A)$ the rank of matrix $A$. Before proving Theorem 2.10, some preliminary results and formulas must be given.

**Lemma 6:** Consider a smooth system (1)-(2) and a state $x' \in \mathbb{R}^n$. Assume that for some $\delta > 0$ the Jacobian $Q_{x}(x')$ of the delay-observability map $H_{\delta}(x')$ is singular for any $\Delta \in D_\delta(\delta)$.

Then, for any $\epsilon > 0$, there exists $\delta_\epsilon < \delta$ such that $\varphi(-\tau, x') \in B_{\epsilon}(x') \forall \tau \in [0, \delta_\epsilon]$ such that for any $\Delta \in D_\delta(\delta_\epsilon)$ it is $|Q_{\Delta}(\varphi(-\gamma\Delta, x'))| = 0$ for some $\gamma \in [0, 1]$.

**(A1)**

**Proof:** The notation $L^k_{\Delta}h_{\lambda}(x) = L^k_{\Delta}h_{\lambda}(\varphi(-\Delta, x))$, $i = 1, \ldots, n$, and $k = 0, 1, \ldots$, will be used throughout this proof. By smoothness assumption, using (12), straightforward computations yield

$$\frac{d}{d\Delta} \left( \frac{dL^k_{\Delta}h_{\lambda}(x)}{dx} \right) = \frac{d}{dx} \left( \frac{dL^k_{\Delta}h_{\lambda}(x)}{d\Delta} \right)$$

$$= -\frac{dL^k_{\Delta+1}h_{\lambda}(x)}{dx}, \quad 1, \ldots, n, \quad k = 0, 1, \ldots$$

**(A2)**

Iterated application of (A2) gives

$$\frac{d^k}{d\Delta^k} \left( \frac{dh_{\lambda}(x)}{dx} \right) = (-1)^k \frac{dL^k_{\Delta}h_{\lambda}(x)}{dx}, \quad 1, \ldots, n, \quad k = 0, 1, \ldots$$

**(A3)**

For the given $\epsilon$, consider $\delta_\epsilon < \delta$ such that $\varphi(-\tau, x') \in B_{\epsilon}(x') \forall \tau \in [0, \delta_\epsilon]$ such that for any $\Delta \in D_\delta(\delta_\epsilon)$ it is continuous w.r.t. $\tau$ and $\varphi(0, x') = x$. Consider any $\Delta \in D_\delta(\delta)$. By assumption, the rows of the Jacobian $Q_{\Delta}(x')$ are not linearly independent, and therefore there exists an
index \( j \in \{1, \ldots, n\} \) such that
\[
\frac{dh_A}{dx} \bigg|_{x^*} = \sum_{i \neq j} a_i(\Delta) \frac{dh_A}{dx} \bigg|_{x^*},
\]
where the \( n-1 \) coefficients \( a_i \) are smooth functions of \([\Delta_1 \ldots \Delta_n]\).

Repeatedly differentiating with respect to \( \Delta_j \) yields
\[
\frac{d^k h_A}{d\Delta_j^k} \bigg|_{x^*} = \sum_{i \neq j} \frac{\hat{a}^k_i}{\hat{a}^{k-1}_j} \frac{dh_A}{dx} \bigg|_{x^*}, \quad k = 0, 1, \ldots, n-1,
\]
as shown in Equation (A3),
\[
\frac{dL_j^k h_A}{dx} \bigg|_{x^*} = (-1)^k \sum_{i \neq j} \frac{\hat{a}^k_i}{\hat{a}^{k-1}_j} \frac{dh_A}{dx} \bigg|_{x^*}, \quad k = 0, 1, \ldots, n-1.
\]

(A6)

Note that the first \( n \) terms \( \frac{dL_j^k h_A}{dx} \bigg|_{x^*} \) are the rows of the delay-observability matrix \( Q_n(x) \) evaluated at \( \varphi(-\Delta_n, x^*) \), i.e.
\[
\text{row}_{j=1,n} \left[ \frac{dL_j^k h_A}{dx} \right] = Q_n(\varphi(-\Delta_n, x^*)),
\]
with obvious meaning of the symbol \( \text{row}_{j=1,n} \). It follows that (A3) can be written as matrix form as
\[
Q_n(\varphi(-\Delta_n, x^*)) = \text{col}_{i \neq j} \left[ \begin{array}{c} a_i \\ \frac{\partial a_i}{\partial \Delta_j} \\ \vdots \\ (-1)^{n-1} \frac{\partial^{n-1} a_i}{\partial \Delta_j^{n-1}} \end{array} \right] \text{row}_{i \neq j} \left[ \frac{dh_A}{dx} \right] \bigg|_{x^*},
\]
where the symbol \( \text{col}_{i \neq j} \) denotes a matrix with \( n-1 \) columns \( c_j \) (rows \( r_j \)), with \( i \neq j \). By (A8), the drift-observability matrix \( Q_n(\varphi(-\Delta_n, x^*)) \) can be written as a product of two matrices with at most rank \( n-1 \), and therefore is singular. Recalling that \( \varphi(-\Delta_n, x^*) \in \mathcal{B}_n(x^*) \), the lemma is proved with \( \gamma = \Delta_n \).

Corollary 7: Consider a smooth system (1)–(2) and a state \( x^* \in \mathbb{R}^n \). Assume that for some \( \delta > 0 \) the delay-observability matrix \( Q_n(x^*) \) is singular for any choice of \( \Delta \in \mathcal{D}_\delta(\delta) \). Then, for any \( \epsilon > 0 \), there exists a state \( \tilde{x} \in \mathcal{B}_n(x^*) \) where the drift-observability matrix \( Q_n(\tilde{x}) \) is singular.

Proof: It follows directly from Lemma 6, setting \( \tilde{x} = \varphi(-\gamma \Delta_n, x^*) \).

The following theorem directly follows from Lemma 6 and Corollary 7.

Theorem A.1: Given a smooth system \( S \) of the type (1)–(2), at any state \( x \in \mathbb{R}^n \) where \( Q_n(x) \) is singular, for any choice of \( \Delta \in \mathcal{D}_\delta(\delta) \), for some \( \delta > 0 \), the drift-observability matrix \( Q_n(x) \) is singular.

Proof: From Corollary 7, \( x \) is an accumulation point of a sequence \( x_k \) such that \( Q_n(x_k) \to 0 \). By smoothness assumption, it follows that \( Q_n(x) = 0 \).

Theorem A.2: Given a smooth system \( S \) of the type (1)–(2), at any state \( x \in \mathbb{R}^n \) where \( Q_n(x) \) is nonsingular, for any \( \delta > 0 \) there exists \( \Delta \in \mathcal{D}_\delta(\delta) \) such that \( Q_n(x) \) is nonsingular.

Proof: From Theorem A.1, the thesis follows by contradiction.

When \( h(x) \) is analytic, the Taylor series expansion of \( h(x(t - \Delta)) \) can be written as
\[
h(x(t - \Delta)) = h(x(t)) + \sum_{k=1}^{\infty} \frac{L_j^k h(x(t)) (-\Delta)^k}{k!},
\]
or, using the shorthand \( h_{\Delta}(x) = h(\varphi(-\Delta, x)) \) we have
\[
h_{\Delta}(x) = h(\varphi(-\Delta, x)) = \sum_{k=0}^{\infty} L_j^k h(x) (-\Delta)^k.
\]
The delay-observability map can be written as
\[
H_{\Delta}(x) = \left[ \begin{array}{c} h_{\Delta}(x) \\ h_{\Delta}(x) \\ \vdots \\ h_{\Delta}(x) \end{array} \right] = M_\Delta \left[ \begin{array}{c} h(x) \\ L_j^1 h(x) \\ \vdots \\ L_j^k h(x) \end{array} \right],
\]
where
\[
M_\Delta = \left[ \begin{array}{ccccc} 1 & -\Delta_1 & \frac{1}{2!}(-\Delta_1)^2 & \frac{1}{3!}(-\Delta_1)^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -\Delta_n & \frac{1}{2!}(-\Delta_n)^2 & \frac{1}{3!}(-\Delta_n)^3 & \cdots \end{array} \right].
\]

When \( \Delta \in \mathcal{D}_\delta \), so that \( \Delta \neq \Delta \), the rank of \( M_\Delta \) is \( n \). Using \( dL_j^k h \) as a shorthand for \( dL_j^k h/\!dx \), and
\[
dh_{\Delta}(x) = dh(\varphi(-\Delta, x)) \bigg|_{x^*},
\]
where
\[
Q_{\Delta}(x) = \left[ \begin{array}{c} dh_{\Delta}(x) \\ dh_{\Delta}(x) \\ \vdots \\ dh_{\Delta}(x) \end{array} \right],
\]
\[
Q_{\infty}(x) = \left[ \begin{array}{c} dh_{\Delta}(x) \\ dh_{\Delta}(x) \\ \vdots \\ dh_{\Delta}(x) \end{array} \right].
\]

Then, the following identity can be written as
\[
Q_{\Delta}(x) = M_\Delta Q_{\infty}(x).
\]

Now, the proof of Theorem 2.10 can be given (the statement of the theorem is reported below, for convenience).

Theorem 2.10: Consider a smooth system \( S \) of the type (1)–(2). In all states \( x \in \mathbb{R}^n \) where \( Q_n(x) \) is nonsingular, there exists \( \Delta \in \mathcal{D}_\delta \) such that \( Q_n(x) \) is nonsingular. If \( h(x) \) is analytic, then in all open sets of \( \mathbb{R}^n \) where \( Q_n(x) \) is singular and has constant rank, the matrix \( Q_{\Delta}(x) \) is singular for any \( \Delta \in \mathcal{D}_\delta \).

Proof: The first statement follows directly from Theorem A.2. The second statement is proved by recalling that if \( \rho(Q_n(x)) \) is constant in an open set, then \( \rho(Q_{\Delta}(x)) = \rho(Q_n(x)) \) for any \( k > n \), so that \( \rho(Q_{\Delta}(x)) = \rho(Q_{\infty}(x)) \) (this happens because in this case the observability co-distribution has constant dimension (see e.g. Isidori 1995). Thus, in open sets where \( Q_n(x) \) is singular and has constant rank, it is \( \rho(Q_{\infty}(x)) < n \). As a consequence of formula (A14), \( Q_{\Delta}(x) \) is also singular.