Adaptive Observers and Parameter Estimation for a Class of Systems Nonlinear in the Parameters

Ivan Y. Tyukin\textsuperscript{a} Erik Steur\textsuperscript{b} Henk Nijmeijer\textsuperscript{c} Cees van Leeuwen\textsuperscript{b}

\textsuperscript{a}Dept. of Mathematics, University of Leicester, Leicester, LE1 7RH, UK (Tel: +44-116-252-5106; I.Tyukin@le.ac.uk)
\textsuperscript{b}Laboratory for Perceptual Dynamics, KU Leuven, Tiensestraat 102, 3000 Leuven, Belgium (erik.steur@ppw.kuleuven.be, cees.vanleeuwen@ppw.kuleuven.be)
\textsuperscript{c}Dept. of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513 5600 MB, Eindhoven, The Netherlands, (h.nijmeijer@tue.nl)

Abstract

We consider the problem of asymptotic reconstruction of the state and parameter values in systems of ordinary differential equations. A solution to this problem is proposed for a class of systems of which the unknowns are allowed to be nonlinearly parameterized functions of state and time. Going beyond the concept of asymptotic Lyapunov stability, we provide for this class a reconstruction technique based on the notions of weakly attracting sets and non-uniform convergence. Reconstruction of state and parameter values is subjected to persistency of excitation conditions. In absence of nonlinear parametrization the resulting observers reduce to standard estimation schemes. In this respect, the proposed method constitutes a generalization of the conventional canonical adaptive observer design.

Key words: Adaptive observers, nonlinear parametrization, weakly attracting sets

1 Introduction

We consider observer-based methods for state and parameter estimation in nonlinear dynamical systems. These methods are effective as long as the original system has, or can be transformed into, one of the canonical adaptive observer forms [5], [27], [6]. Their common characteristic is linearity in the unknown parameters. For this class of systems, subject to persistency of excitation conditions, reconstruction of state and parameter vectors can be achieved exponentially fast [29], [6], [31].

There are systems, however, in which the unknown parameters enter the model nonlinearly. These systems constitute a remarkably wide class including, for instance, models in chemical kinetics [13], [4], biology and neuroscience [20]. Whereas the problem of state estimation can be solved for a large class of nonlinearly parameterized systems [30], applicability of observer-based parameter reconstruction techniques is often confined to systems with monotone [42], [24] or one-to-one parameterizations [16], [17], [18], [12].

Several authors have recently advanced strategies for overcoming these limitations. In [21], for example, combining interval analysis with multiple shooting methods is proposed in order to tackle the state and parameter reconstruction problem. Another interesting approach is presented in [1]: the original continuous-time model is replaced with a discrete-time approximation. Measured variables can then be considered as known functions of unknown parameters and initial conditions, of which the estimates can be found by using brute force off-line nonlinear optimization routines (see also [40], [2] where optimization techniques with moving horizon are discussed). These optimization-inspired approaches offer obvious advantages, e.g. the availability of a vast library of numerical methods for solving general nonlinear optimization problems. Nevertheless, these methods run into restrictions too. Exhaustive search for a global minimum may become untractable for dimensions higher than 1 or 2. On the other hand, if conventional polynomial-complexity algorithms are used then the possibility arises that the algorithm will converge to a local minimum.

Preprint submitted to Automatica April 9, 2013
In this paper we explore further possibilities of developing adaptive observers for systems which are both linearly and nonlinearly parameterized and thus are not in canonical form. The parametrization is not required to be invertible or monotone. Our approach combines the advantages of the existing schemes, in being capable of ensuring exponentially fast convergence with the flexibility of explorative behavior, a behavior inherent to algorithm for solving genuine nonlinear optimization problems. In particular, we suggest that inferring the values of state and a part of the parameter vector of these systems can be achieved by employing exponentially fast converging estimators. In order to estimate the values of the remaining parameters, a procedure based on explorative search is proposed. This, on the one hand, enables us to deal with a broad range of nonlinearly parameterized uncertainty models. On the other hand, since exploration is restricted to a subset of the unknown parameters, the proposed strategy reduces the overall computational costs incurred, as compared to when full-scale search-based optimization had been invoked.

The resulting observer can be imagined as a system comprising of an exponentially stable part coupled with an explorative one. Systems of this type have previously been used successfully in adaptive control, in particular in the classical universal adaptive stabilization schemes discussed in [38], [19], [32], [33]. Even though in these algorithms, the need for resorting to exploration in the space of parameters is motivated differently from the currently proposed method, clearly they are conceptually related. Here we demonstrate that applicability of these classical ideas can be extended beyond mere stabilization problems for systems with linearly parameterized controllers to the problem of adaptive observer design for systems which are nonlinearly dependent on parameters. We show that, subject to a condition of persistent excitation, it is possible to reconstruct state and parameters of a reasonably broad subclass of these systems. Our analytical constructions invoke the concepts of weakly attracting sets and relaxation times [34], [14], [15], and are using the results of [26] and [43].

The paper is organized as follows. Notational agreements are introduced in Section 2. Section 3 provides the formal statement of the problem, Sections 4, 5 contain main results of the article, Section 6 discusses possible generalizations, Section 7 contains illustrative examples, and Section 8 concludes the paper. Proofs of auxiliary results are presented in the Appendix.

2 Notation

The following notational conventions are used throughout the paper:

- \( \mathbb{R} \) denotes the set of real numbers, \( \mathbb{R}_{>0} = \{ x \in \mathbb{R} \mid x > 0 \} \), and \( \mathbb{R}_{\geq 0} = \{ x \in \mathbb{R} \mid x \geq 0 \} \);
- \( Z \) denotes the set of integers, and \( \mathbb{N} \) stands for the set of positive integers;
- The Euclidian norm of \( x \in \mathbb{R}^n \) is denoted by \( \| x \| \), \( \| x \|^2 = x^T x \), where \( T \) stands for transposition;
- The space of \( n \times n \) matrices with real entries is denoted by \( \mathbb{R}^{n \times n} \); let \( P \in \mathbb{R}^{n \times n} \), then \( P > 0 (P \geq 0) \) indicates that \( P \) is symmetric and positive (semi-)definite; \( I_n \) denotes the \( n \times n \) identity matrix.

By \( L^\infty_{\mathbb{R}}[t_0, T] \), \( t_0 \in \mathbb{R}, T \in \mathbb{R}, T \geq t_0 \) we denote the space of all functions \( f : [t_0, T] \rightarrow \mathbb{R}^n \) such that \( \| f \|_{\infty, [t_0, T]} = \text{ess sup} \{ \| f(t) \|, t \in [t_0, T] \} < \infty \); \( \| f \|_{\infty, [t_0, T]} \) stands for the \( L^\infty_{\mathbb{R}}[t_0, T] \) norm of \( f(t) \); if the function \( f \) is defined on a set larger than \( [t_0, T] \) then notation \( \| f \|_{[t_0, T]} \) applies to the restriction of \( f \) on \( [t_0, T] \);
- \( C^r \) denotes the space of continuous functions that are at least \( r \) times differentiable;
- Let \( A \) be a subset of \( \mathbb{R}^n \), then for all \( x \in \mathbb{R}^n \), we define \( \text{dist}(A, x) = \inf_{q \in A} \| x - q \| \);
- A solution of \( x = f(t, x, \theta, u(t)) \), \( f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \rightarrow \mathbb{R}^n \), \( u : \mathbb{R} \rightarrow \mathbb{R} \) passing through \( x_0 \in \mathbb{R}^n \) at \( t = t_0 \) is denoted by \( x(t, t_0, x_0, \theta, [u]) \). In cases when \( u \) and/or \( \theta, x_0, t_0 \) are clearly determined by the context a more compact notation, \( x(t, t_0, x_0, \theta) \) (or \( x(t, t_0, x_0) \), \( x(t) \) respectively), is used.
- The symbol \( \mathcal{K} \) denotes the class of all strictly increasing continuous functions \( \kappa : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+ \) such that \( \kappa(0) = 0 \); the symbol \( \mathcal{K}_\infty \) denotes the class of all functions \( \kappa \in \mathcal{K} \) such that \( \lim_{s \rightarrow \infty} \kappa(s) = \infty \).
- Let \( \epsilon \in \mathbb{R}_0^+ \), then \( \| x \|_2 \) stands for: \( \| x \| - \epsilon \) if \( \| x \| > \epsilon \), and 0, otherwise.
- Finally, for \( \lambda \in \mathbb{R}^d \) and \( \theta \in \mathbb{R}^m \), the notation \( (\lambda, \theta) \) stands for the col(\( \lambda_1, \ldots, \lambda_p, \theta_1, \ldots, \theta_m \)).

3 Preliminaries and Problem formulation

3.1 Adaptive observer canonical form

Throughout the paper we will focus exclusively on the class of systems that are forward-complete:

**Definition 1** Let \( \mathcal{L}_u[t_0, \infty) \) be a subspace of \( L^\infty_{\mathbb{R}}[t_0, \infty) \). A single-input single-output system described by \( \dot{x} = f(t, x, u(t)) \), \( f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n \), \( y = h(t, x) \), \( h : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R} \), \( u \in \mathcal{L}_u[t_0, \infty) \), where \( u \) is the input, and \( y \) is the output, is called forward-complete (with respect to \( \mathcal{L}_u \)) iff for any \( t_0 \in \mathbb{R}, x_0 \in \mathbb{R}^n \), and \( u \in \mathcal{L}_u[t_0, \infty) \) the solution \( x(t, t_0, x_0, [u]) \) exists and is defined for all \( t \geq t_0 \).

Let \( \mathcal{L}_u[t_0, \infty) = L^\infty_{\mathbb{R}}[t_0, \infty) \cap C^0[t_0, \infty) \), and consider a forward-complete single-input single-output system; let \( x \in \mathbb{R}^n \) be its state, \( y : \mathbb{R} \rightarrow \mathbb{R} \) be the measured output, and \( u : \mathbb{R} \rightarrow \mathbb{R} \), \( u \in \mathcal{L}_u \) be the input. We recall that a system is in the adaptive observer canonical form if it is governed by the following set of equations

\[
\begin{align*}
\dot{x} &= Ax + Bo^T(t, \theta) + g(t, y, u) \\
y &= C^T x, \quad x(t_0) = x_0, \quad x_0 \in \mathbb{R}^n,
\end{align*}
\]
where

\[
A = \begin{pmatrix} I_{n-1} \\ a \end{pmatrix}, \quad a \in \mathbb{R}^n, \quad B \in \mathbb{R}^n, \quad C \in \mathbb{R}^n,
\]

\[
B = \text{col}(1, b_1, \ldots, b_{n-1}), \quad C = \text{col}(1, 0, \ldots, 0), \quad \text{the functions } \phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^m, \quad g : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^n, \quad \phi, g \in C^0\text{ are known}^1, \quad \text{and } \theta \in \mathbb{R}^m\text{ is the vector of unknown parameters. The triplet } A, B, C \text{ is supposed to satisfy}
\]

\[
\begin{cases}
P(A + \ell C^T) + (A + \ell C^T)^T P \leq -Q \\
P B = C
\end{cases}
\]

for some \( \ell \in \mathbb{R}^n \) and \( P > 0, Q > 0 \). Although condition (2) may appear restrictive at first glance, it has been shown in [28] that subject to the very natural constraint that the pair \( A, C \) is observable, there always exists a time-varying parameter-dependent coordinate transformation such that in new coordinates the system is still of the form (1) and satisfies condition (2). If requirement (2) holds then the system

\[
\begin{align*}
\dot{x} &= Ax + \ell (C^T x - y) + B \phi^T(t, y) \hat{\theta} + g(t, y, u) \\
\dot{\theta} &= -\gamma (C^T x - y) \phi(t, y), \quad \gamma \in \mathbb{R}_{>0},
\end{align*}
\]

where \( \dot{x}(t) \in \mathbb{R}^n, \dot{\theta}(t) \in \mathbb{R}^m, \) is an adaptive observer for (1) (cf. [27], [5]) provided that the term \( \phi(t, y(t)) \), if viewed as a function of \( t \), is persistently exciting.

**Definition 2** A function \( \beta : \mathbb{R}_{\geq t_0} \to \mathbb{R}^m \) is said to be persistently exciting if there exist \( L, \mu \in \mathbb{R}_{>0} \):

\[
\int_{t_0}^{t + L} \beta(\tau) \beta^T(\tau) d\tau \geq \mu I_m, \quad \forall \ t \geq t_0.
\]

The fact that (3) is an adaptive observer for (1) is based on a well-known result on the exponential stability of the following class of linear time-varying systems:

\[
\dot{e} = A(t)e, \quad A(t) = \begin{pmatrix} A + \ell C^T & B \beta(t) \\
-\beta(t) C^T & 0 \end{pmatrix}.
\]

The result is provided in Theorem 3 below (see e.g. [37] for details of the proof).

---

1. To ensure that solutions of (1) are defined for all \( t \) the functions \( \phi, g \) are often required to be Lipschitz in \( y \). This is not necessary if (1) is known to be forward-complete.

2. In the context of adaptive observer design for (1) the function \( \beta \) in (5) is defined as \( \beta(t) = \phi(t, y(t)) \).

**Theorem 3** Consider system (5). Suppose that condition (2) holds for certain \( \ell \in \mathbb{R}^n, P > 0, Q > 0 \), the function \( \beta(t) \) is persistently exciting, and

\[
\exists \ M \in \mathbb{R}_{>0} : \max \{ \| \beta(t) \|, \| \dot{\beta}(t) \| \} \leq M.
\]

Let \( \Phi(t, t_0), \Phi(t_0, t_0) = I, \) be the fundamental solution matrix of (5). Then there exist \( \rho, D \in \mathbb{R}_{>0} \) such that:

\[
\| \Phi(t_2, t_1) \rho \| \leq D e^{-\rho(t_2 - t_1)} \| \rho \|
\]

\[
\forall \ t_2 \geq t_1 \geq t_0, \quad \rho \in \mathbb{R}^{n+m}.
\]

The parameters \( \rho \) and \( D \) can be expressed explicitly as functions of \( M, \mu, L, \) and \( A, B, C, \ell \) [25]. By letting \( e = (x - \theta, \dot{\theta} - \theta) \) and taking (1), (3) into account one can confirm that the system-observer equations are of form (5). Thus, subject to persistency of excitation of \( \phi^T(t, y(t)), \lim_{t \to \infty} \hat{x}(t, t_0, x_0, \theta_0) - x(t, t_0, x_0, \theta) = 0, \) \( \lim_{t \to \infty} \theta(t, t_0, x_0, \theta_0) = \theta \) along the solutions of (1), (3), and the convergence is exponential. The problem, however, is that if some parameters enter the equations nonlinearly then this creates an obstacle for the explicit use of Theorem 3 and, consequently, observer (3). In the next sections we present and analyze a class of systems nonlinear in the parameters which can be thought of as an immediate generalization of (1).

### 3.2 Systems considered in this article

We begin with the following class of forward-complete single-input-single-output nonlinear systems:

\[
\begin{align*}
\dot{x} &= Ax + B \varphi^T(t, \lambda, y) \theta + g(t, \lambda, y, u) + \xi(t), \\
y &= C^T x, \quad x(t_0) = x_0, \quad x_0 \in \mathbb{R}^n,
\end{align*}
\]

where \( A \in \mathbb{R}^{n \times n}, \) and \( B, C \in \mathbb{R}^n \) are defined as in (1); \( \varphi : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^m, \) \( g : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^n, \) are known continuous functions, \( \lambda = (\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p, \) \( \theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m \) are unknown parameters, and \( u \in \mathcal{L}_u \subset C^1 \), is the input. We assume that the values of \( \lambda, \theta \) belong to the hypercubes \( \Omega_\lambda \subset \mathbb{R}^p, \Omega_\theta \subset \mathbb{R}^m \) with known bounds: \( \theta_j \in [\theta_{j,\min}, \theta_{j,\max}], \lambda_j \in [\lambda_{j,\min}, \lambda_{j,\max}] \), and that \( y(t) \in \mathcal{D}_y, u(t) \in \mathcal{D}_u, \mathcal{D}_u \subset \mathbb{R} \) for \( t \geq t_0 \).

In (7), \( x = \text{col}(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \) is the state vector, \( y \) is the measured output, the input \( u \) is a known function, and \( \xi \in C^0 ; \ \mathbb{R} \to \mathbb{R}^n \) is an unknown yet bounded continuous function:

\[
\| \xi(t) \| \leq \Delta_\xi, \quad \Delta_\xi \in \mathbb{R}_{>0}, \quad \forall \ t
\]

representing some unmodeled dynamics (e.g. noise). The system’s state \( x \) is not measured; only the values of the input \( u(t) \) and the output \( y(t) = x_1(t), t \geq t_0 \) in (7) are...
Table 1
Examples of physical systems in the form of (7), (9)

| Physical system                          | Model                                                                 |
|-----------------------------------------|----------------------------------------------------------------------|
| 1. Reactant dynamics in a reactor [39]  | \( \dot{x} = \frac{F}{v}(x_0 - x) - \kappa_{\text{ref}}e^{\frac{t}{\tau}}(\frac{1}{\tau} + \frac{1}{\tau_{\text{ref}}})x, \) \( y = x \) |
| 2. Magnetic bearings [23]               | \( \ddot{x} + J^{-1}(d) = b\begin{pmatrix} \frac{q_2(t_2,x_2)}{(x+a)g} - \frac{q_1(t_1,x_1)}{(x-a)g} \end{pmatrix} \) \( y = x, |x| < a - \varepsilon, a, \varepsilon \in \mathbb{R}_{>0}, \) \( d, d \) are bounded |
| 3. Evoked potentials in a cell [41]     | \( \begin{align*}
    \dot{x}_1 &= -\frac{\mu}{\tau_m} + \frac{\Delta f}{\tau_m} \tanh\left(\frac{\sigma f}{\tau_f}x_1\right) - \frac{\sigma y}{\tau_f} \\
    \dot{x}_2 &= -\frac{\mu}{\tau_m} + \frac{\Delta f}{\tau_m} x_1 \\
    y &= x_1
\end{align*} \) |

accessible over any time interval \([t_0, t]\) that belongs to the history of the system.

For the time being we suppose that matrix \( A \) and vectors \( B, C \) in (7) satisfy Assumption 3.1 below.

**Assumption 3.1** The triple \( A, B, C \) is known, and there exist (are known) a vector \( \ell \) and matrices \( P, Q > 0 \) such that condition (2) holds.

Note that Assumption 3.1 implies that the vector \( B = \text{col}(1, b_1, \ldots, b_{n-1}) \) in (7) is such that the polynomial \( s^n + b_1 s^{n-1} + \cdots + b_{n-1} \) is Hurwitz. At first, Assumption 3.1 may seem restrictive. In Section 6 we lift this restriction by showing that the results presented for (7) can be generalized to systems

\[ \dot{x} = Ax + \Psi(t, \lambda, y)\theta + g(t, \lambda, y, u) + \xi(t), \]

where the matrix \( A \in \mathbb{R}^{n \times n} \) may be unknown but it is known that the pair \( A, C \) is observable, the function \( \Psi : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^{n \times m}, \Psi \in \mathcal{C}^1, \) is Lipschitz in \( \lambda, \) and \( g_c(t, \lambda, y(t), u(t)), \) \( g(t, \lambda, y(t)), \) \( \Psi(t, \lambda, y(t)), \) \( \Psi(t, \lambda, y(t)) \) are bounded for all \( \lambda \in \Omega_\lambda. \)

With regards to the functions \( \varphi \) and \( g \) in (7) the following additional technical assumptions are made:

**Assumption 3.2** The functions \( \varphi(\cdot, \lambda, \cdot), g(\cdot, \lambda, \cdot, \cdot) \) in (7) are bounded and differentiable in \( \mathbb{R} \times \mathbb{R}_g \) respectively, and Lipschitz in \( \lambda. \) That is, there exist \( D_{\varphi}, D_g, B_{\varphi}, B_g \in \mathbb{R}_{>0} \) such that for all \( t \in \mathbb{R}, \lambda \in \Omega_\lambda, \) \( u \in \Omega_u, \) \( \lambda', \lambda'' \in \Omega_\lambda, \)

\[ \begin{align*}
    ||\varphi(t, \lambda', y) - \varphi(t, \lambda'', y)|| &\leq D_{\varphi}||\lambda' - \lambda''||, \\
    ||g(t, \lambda', y, u) - g(t, \lambda'', y, u)|| &\leq D_g||\lambda' - \lambda''||, \\
    ||\varphi(t, \lambda, y)|| &\leq B_{\varphi}, \\
    ||g(t, \lambda, y, u)|| &\leq B_g.
\end{align*} \]

Furthermore, there exist \( M_{\varphi}, M_g \in \mathbb{R}_{>0} \) such that

\[ \begin{align*}
    \left| \frac{\partial \varphi(t, \lambda, y)}{\partial y} \right| + \left| \frac{\partial g(t, \lambda, y, u)}{\partial u} \right| &\leq M_{\varphi}, \\
    \left| \frac{\partial g(t, \lambda, y, u)}{\partial u} \right| &\leq M_g
\end{align*} \]

for all \( \lambda \in \Omega_\lambda \) along the solutions of (7).

Regarding conditions (10), (11), they are often natural in the context of modelling and identification; they may, however, impose limitations in the framework of controller design. As for condition (12), the first inequality is a version of (6) that is essential for uniform exponential convergence of solutions to the origin of (5) [26]. The second inequality in (12) is a technical condition that will be needed for estimation of the nonlinear parameters of the model. Although this latter condition may look somewhat restrictive, it may be relaxed if the function \( g(t, \lambda, y, u) \) is expressed as \( g(t, \lambda, y, u) = g_1(t, y, u) + g_2(t, \lambda, y, u). \) In this case we would require that (11), (12) hold for \( g_2. \)

A non-exhaustive list of systems that are relevant in engineering applications and are governed by (7) or (9) includes bio- / chemical reactors [7,39], nonlinear saturated magnetic circuits [35], magnetic bearings [23], tyre-road interaction, and dynamics of live cells [41]. A few examples from this list are provided in Table 1. The first model, if described by (7), trivially satisfies Assumption 3.1; it also satisfies Assumption 3.2 if \( T \) is bounded, differentiable, separated away from zero, and \( y, T \) are bounded. In the second model the pair \( A, C \) is observable, and \( \varphi, \bar{\varphi} \) are bounded if \( y, \) magnetic fluxes, expressed by \( \eta_1, \eta_2, \eta_1', \eta_2' \) are bounded. This is achievable via external controls [23], at least for small parameter mismatches and \( d. \) In the third model the pair \( A, C \) is observable, and boundedness of \( y, \bar{y}, \Psi, \bar{\Psi} \) is consistent with the physics of the system. An illustration of the adaptive observer design for this model is provided.
in Section 7.

3.3 Problem formulation

In what follows we will be concerned with the problem of estimating the values of \( x(t, \lambda, \theta, x_0, [u]) \) and \( \theta, \lambda \) of (7) from the values of \( y, u \) available over \([t_0, t]\). Before we proceed with a formal problem statement, several points related to parametrization of (7), (9) need to be discussed. First, note that different definitions of systems (7), (9) may correspond to the same physical model. For example, functions \( \varphi \) and \( g \) and parameters \( \theta, \lambda \) for the first model in Table 1 may also be defined as:

\[
\begin{align*}
\varphi(t, \lambda, y) &= \left(1, y, e^{\lambda_1 t} + \lambda_2 y\right)^T, \quad g = 0, \\
\theta &= \left(E x_0, -E, -k_{cd}E_\theta\right)^T, \quad \lambda = \left(-\frac{E}{R}, \frac{E}{R}, \frac{E}{R_\theta}\right)^T, \\
\varphi(t, \lambda, y) &= \left(1, y, g(t, \lambda, y, u) = -e^{\lambda_1 t} + \lambda_2 y, \right)
\end{align*}
\]

where \( \lambda, \theta \) are chosen then identical outputs \( y(t) \) will be observed for infinitely many combinations of parameters \( \theta, \lambda \). Models of this type are referred to as unidentifiable [11] (see also [9], [10] for further discussion, related definitions, tests and examples). Dealing with unidentifiable models, as will become clear later, imposes technical difficulties. We will therefore assume that parametrizations which are obviously unidentifiable are to be avoided, if possible. Regarding the remaining alternative parametrizations, we assume that preference is given to those in which the dimension of \( \lambda \) is minimal. That is, the parametrization in the first row in Table 1 is preferable to (14).

Second, as far as identifiability is concerned, inferring true values of \( \theta, \lambda \) from mere output observations, \( y(t) \), is not always possible, even if the system is linearly parameterized and no unmodeled dynamics are present. Consider e.g.:

\[
\begin{align*}
\dot{x} &= Ax + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \theta + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \lambda, \quad A = \begin{pmatrix} -a_1 & 1 \\ -a_2 & 0 \end{pmatrix} \\
y &= x_1, \quad a_1, a_2 \in \mathbb{R}_{>0}
\end{align*}
\]

Let \( x(t, \theta, \lambda, x_0) \) and \( x(t, \theta', \lambda, x'_0) \) be two solutions of (15) corresponding to different parameter values and initial conditions, and let \( e = \text{col}(e_1, e_2) = x(t, \theta, \lambda, x_0) - x(t, \theta', \lambda, x'_0) \). Picking \( e_2(t_0) = 0 \) ensures that \( e_1(t_0) = 0 \) for all \( t \geq t_0 \) if \( \theta = \theta' \) and \( \lambda = \lambda' \). In addition, with nonlinear dependence of (7) on parameters, multiple distinct parameter values may enable the same input-output mapping. A simple example is

\[
\dot{x} = -x + \theta + \sin^2(\lambda + t) + x^2 + 1)^{-1}, \quad y = x.
\]
Since this is desirable from an implementation point of view, we will seek for a recursive procedure taking the values \( y(t), u(t) \) as inputs and producing the estimates \( \hat{x}(t), \hat{\theta}(t), \hat{\lambda}(t) \) of \( x(t, \lambda, \theta, x_0, [u]), \theta, \) and \( \lambda, \) respectively, as outputs. Note that since we allow for non-
identifiable configurations, estimation of parameters \( \theta \) and \( \lambda \) is possible only up to the set \( \mathcal{E}(\lambda, \theta) \). In particular we will be looking for an auxiliary system, i.e., an adaptive observer: \( \hat{q} = \hat{f}(t, y, u, q), f : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^q \rightarrow \mathbb{R}^q, \) \( q(t_0) = q_0, q \in \mathbb{R}^q, \) and functions \( h_\lambda : \mathbb{R}^q \rightarrow \mathbb{R}^m, h_\lambda : \mathbb{R}^q \rightarrow \mathbb{R}^p, h_x : \mathbb{R}^q \rightarrow \mathbb{R}^n \) such that for some given \( q_0, \) known functions \( r_1, r_2 \in \mathbb{K}, \) and all \( t_0 \in \mathbb{R}, \) the following requirements hold for the observer:

\[
\limsup_{t \rightarrow \infty} \| h_\lambda(q(t, q_0)) - x(t, \lambda, \theta, x_0) \| \leq r_1(\Delta_\xi) \tag{21}
\]

\[
\limsup_{t \rightarrow \infty} \text{dist} \left( \left( h_\lambda(q(t, q_0)), h_\theta(q(t, q_0)) \right), \mathcal{E}(\lambda, \theta) \right) \leq r_2(\Delta_\xi),
\]

where \( \Delta_\xi \) is defined in (8).

4 Main result

In this section we show that despite the obvious difficulties related to the presence of unmodelled dynamics, nonlinear parametrisation, and possible non-uniqueness issues, asymptotic reconstruction of state and parameters of (7) is achievable (up to the set \( \mathcal{E}(\lambda, \theta) \)), subject to some persistency of excitation conditions.

4.1 Observer definition

Following the general ideas of [43] with regards to the treatment of uncertain systems with general nonlinear parametrisation, we propose that an asymptotically converging observer for (7) consists of two coupled subsystems, \( S_a \) and \( S_w \) (see Fig. 1). The role of subsystem \( S_a \) is to provide estimates of state and parameters \( \theta \) of (7), and the role of subsystem \( S_w \) is to search the values of parameters \( \lambda. \)

The dynamics of subsystem \( S_a \) is defined as follows:

\[
\begin{align*}
\dot{x} &= A \hat{x} + \ell \left( C^T \hat{x} - y \right) + B \varphi^T(t, \lambda, y) \hat{\theta} + g(t, \hat{\lambda}, y, u) \\
\dot{\theta} &= -\gamma_0(C^T \hat{x} - y) \varphi(t, \lambda, y), \quad \gamma_0 \in \mathbb{R}_{>0} \\
\hat{y} &= C^T \hat{x}, \quad \hat{y}(t_0) \in \mathbb{R}^m, \quad \hat{\theta}(t_0) \in \mathbb{R}^n.
\end{align*}
\tag{22}
\]

The variable \( \hat{\theta} = \text{col}(\hat{\theta}_1, \ldots, \hat{\theta}_m) \) in (22) is an estimate of \( \theta \), and \( \hat{\lambda} = \text{col}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p) \) is the estimate of \( \lambda \). For the time being we suppose that \( \hat{\lambda} \) is a continuous function of \( t \). The matrix \( A \) and vectors \( B, C \) in (22) are identical to those in (7), and the vector \( \ell \) in (22) is chosen such that Assumption 3.1 holds. If the values of \( \lambda \) would be known then substitution \( \hat{\lambda} = \lambda \) reduces system (22) to (3), and conditions for asymptotic reconstruction of state and parameter values of (7) follow from Theorem 3 provided that \( \beta(t) \) is replaced with \( \varphi(t, \hat{\lambda}(t), y(t)) \). The problem, however, is that the values of \( \lambda \) are unknown and therefore a procedure for estimating the values of \( \lambda \) is needed.

Regarding the definition of \( S_w \), we propose that the values of \( \hat{\lambda} \) result from an explorative search in the domain \( \Omega_\lambda \) of the admissible values for \( \lambda \). The exploration can be realized by movements along solutions of a certain class of dynamical systems. Let us, for instance, consider systems governed by the following equations

\[
\dot{s} = f(s), \quad s(t_0) = s_0, \quad \hat{\lambda} = \beta(s) \tag{23}
\]

where \( f : \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_p}, \beta : \mathbb{R}^{n_p} \rightarrow \mathbb{R}^p \) are continuous, and let \( \Omega_\lambda \) be the \( \omega \)-limit set \(^3\) of \( s_0 \). In addition, suppose that the following properties hold:

P1) the functions \( f, \beta \) in (23) are Lipschitz;
P2) the vector \( s_0 \) is such that the solution \( s(t, s_0) \) is bounded for all \( t \geq t_0; \)
P3) the restriction of \( \beta \) on \( \Omega_s \) contains \( \Omega_\lambda; \) for every \( \lambda \in \Omega_\lambda \) there is an \( s \in \Omega_s \) such that \( \beta(s) = \lambda \).

Properties P1 and P2 are technical requirements ensuring that the derivative of \( \lambda \), as a function of \( t \), is bounded and has a bounded growth rate. Property P3, however, is essential. It implies that projection \( \beta(s(t, s_0)) \) of the trajectory \( s(t, s_0) \) onto \( \Omega_\lambda \) is dense and recurring in \( \Omega_\lambda: \)

\[
\forall \lambda \in \Omega_\lambda, \forall \epsilon > 0, \exists s(t, s_0) \in \Omega_s, \exists \epsilon > 0, \exists t' > t \quad \| \lambda - \beta(s(t', s_0)) \| < \epsilon. \tag{24}
\]

Indeed, let \( \lambda' \) be an element from \( \Omega_\lambda \). Then according to P3 there is an \( s' \in \Omega_s; \beta(s') = \lambda' \). Since \( \Omega_\lambda \) is the \( \omega \)-limit set of \( s_0 \), we can conclude that there is a sequence \( \{ t_i \}, i = 1, 2, \ldots, \lim_{i \rightarrow \infty} t_i = \infty, \) such that \( \lim_{i \rightarrow \infty} s(t_i, s_0) = s' \). Finally, using the continuity of \( \beta \) we arrive at \( \lim_{i \rightarrow \infty} \beta(s(t_i, s_0)) = \lambda' \). In other words, for any \( \lambda' \in \Omega_\lambda \) and \( \epsilon > 0 \) there will exist a sequence of time instances \( t_i : \lim_{i \rightarrow \infty} t_i = \infty \) such that \( \| \lambda(t_i) - \lambda' \| < \epsilon \), and hence (24) follows. An example of a very simple system possessing a solution \( s(t, s_0) \) and an output function \( \beta \) satisfying properties P1–P3 for \( \Omega_\lambda = [-1, 1]^2 \) is

\[
\begin{align*}
\dot{s}_1 &= -\sqrt{2} s_2, \quad \dot{s}_2 = \sqrt{2} s_1, \quad \dot{s}_3 = -s_4, \quad \dot{s}_4 = s_3, \quad \beta(s(t, s_0)) = \text{col}(s_1(t), s_3(t)), \quad s_0 = \text{col}(1, 0, 1, 0). \tag{25}
\end{align*}
\]

Phase curves of (25) are shown in Fig. 1, right panel. Projections of the initial segment of the trajectory are shown.

---

\(^3\) Recall that a point \( z \in \mathbb{R}^{n_p} \) is an \( \omega \)-limit point of \( z_0 \in \mathbb{R}^{n_p} \) if there is a sequence \( \{ t_i \}, i = 1, 2, \ldots, \lim_{i \rightarrow \infty} t_i = \infty, \) such that \( \lim_{i \rightarrow \infty} s(t_i, z_0) = z \). The set of all \( \omega \)-limit points of \( z_0 \) is the \( \omega \)-limit set of \( z_0 \).
by thick lines. After evolving beyond the initial segment, curve \( \tilde{\beta}(s(t,s_0)) \) will densely fill the whole square \([\lambda_{1\min}, \lambda_{1\max}] \times [\lambda_{2\min}, \lambda_{2\max}] = [-1,1]^2\), cf. [36].

The problem with using (23) directly as an estimator for \( \lambda \) is that exploration of the domain \( \Omega_\lambda \) will continue indefinitely. For the purposes of observer design we need to ensure that exploration of \( \lambda \) stops once a sufficiently small neighborhood of the set \( \mathcal{E}(\lambda, \theta) \) has been reached. To enable this, the explorative subsystem must be supplied with an error measure. A function of \( \|y(t) - \hat{y}(t)\|_\epsilon \) is one of the possible candidates for such a measure. Thus we replace the earlier definition (23) for \( \hat{\lambda}(t) \) with the following:

\[
\dot{s} = \gamma \sigma(\|y(t) - \hat{y}(t)\|_\epsilon) f(s), \quad \epsilon \in \mathbb{R}_{\geq 0}, \quad \gamma \in \mathbb{R}_{>0}, \quad \hat{\lambda} = \beta(s), \quad s(t_0) = s_0, \tag{26}
\]

where \( \sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \) is a bounded Lipschitz function:

\[
\exists D_\sigma, \quad M_\sigma \in \mathbb{R}_{\geq 0} : \quad \sigma(v) \leq M_\sigma, \quad \sigma(v) \leq D_\sigma v \quad \forall \, v \geq 0 \tag{27}
\]

such that \( \sigma(v) > 0 \) for \( v > 0 \), and \( \sigma(0) = 0 \).

For the sake of simplicity and without loss of generality, instead of dealing with general systems (26), we will focus on a specific system of equations:

\[
S_w : \begin{cases}
\dot{s}_{2j-1} = \gamma \sigma(\|y - \hat{y}\|_\epsilon) \cdot \omega_j \cdot (s_{2j-1} - s_{2j}) \\
- s_{2j-1} (s_{2j-1}^2 + s_{2j}^2) \\
\dot{s}_{2j} = \gamma \sigma(\|y - \hat{y}\|_\epsilon) \cdot \omega_j \cdot (s_{2j-1} + s_{2j}) \\
- s_{2j} (s_{2j-1}^2 + s_{2j}^2) \\
\dot{\lambda}_j = \beta_j(s), \quad j = \{1, \ldots, p\}, \\
\beta_j(s) = \lambda_{j,\min} + \frac{\lambda_{j,\max} - \lambda_{j,\min}}{2} (s_{2j-1} + 1) \\
s_0 = s(t_0) : \quad s_{2j-1}(t_0) + s_{2j}(t_0) = 1, \tag{28}
\end{cases}
\]

where \( \sigma \) is a function satisfying (27). Parameters \( \omega_j \in \mathbb{R}_{\geq 0} \) in (28) are supposed to be rationally-independent:

\[
\sum_{j=1}^p \omega_j k_j \neq 0, \quad \forall \, k_j \in \mathbb{Z}. \tag{30}
\]

Equations (28)–(30) are straightforward generalizations from the example system of which the phase curves are shown in Fig. 1. If the term \( \gamma \sigma(\|y - \hat{y}\|_\epsilon) \) in the right-and side of (28) is substituted with 1, these equations satisfy the requirements P1–P3. Indeed, we can immediately see that in this case variables \((s_{2j-1}(t,s_0), s_{2j}(t,s_0))\) can be expressed as \((s_{2j-1}(t,s_0), s_{2j}(t,s_0)) = (\cos(\omega_j(t - t_0) + a_j), \sin(\omega_j(t - t_0) + a_j)) \), \( a_j \in \mathbb{R} \). Thus properties P1, P2 hold. Trajectories \( s_1(t,s_0), s_2(t,s_0), \ldots, s_{2p-1}(t,s_0) \) evolve on a corresponding \( p \)-dimensional invariant torus. Since \( \omega_j \) are rationally-independent these trajectories densely fill the torus (cf. [3], [36], [22]) or, alternatively, they densely fill the hypercubed \([-1,1]^p\). This implies that \( \Omega_\lambda \), the omega-limit set of \( s_0 \), is \( \Omega_\lambda = \{(s_1, s_2, \ldots, s_{2p}) \in \mathbb{R}^p : (s_1, s_3, s_5, \ldots, s_{2p-1}) \in [-1,1]^p, \quad s_{2j} = \pm \sqrt{1 - s_{2j-1}^2}, \quad j = 1, \ldots, p\} \). Noticing that the image of \( \Omega_\lambda \) under transformation \( \beta \) coincides with \( \Omega_\lambda \) we conclude that P3 holds.

Concerning the structure of \( S_w \), no additional model-dependent constraints are imposed on (28) (or, in general, on (26)), apart from the general requirements P1–P3. Model-specific nonlinearities are accounted for in the “converging” part, \( S_w \), of the observer producing the estimates for \( \theta \) and \( x \). The information about the values of \( \lambda \) is transferred to the exploratory part, \( S_w \), by means of \( \|y - \hat{y}\|_\epsilon \). The latter variable modulates the speed of exploration in \( \Omega_\lambda \) along a search trajectory. The search trajectory itself doesn’t need to be dependent on the properties of \( g, \varphi \), and neither is the structure of \( S_w \). On the one hand, independence of the choice of \( S_w \) on \( \varphi, \, g \) may be viewed as an advantage of the approach. On the other hand, such advantage comes at a price. According to (24), small neighborhoods of sets to which the solutions of the combined system (22), (28) converge are not
necessarily forward invariant. Hence the sets themselves are not guaranteed to be asymptotically stable. Nevertheless, albeit in a weaker sense, they are still attracting. We illustrate this point in more detail in Section 7.

4.2 Asymptotic properties of the observer

Let us now proceed with specifying those properties of (7) that can be useful for state and parameter reconstruction. First we recall that (7) is a generalization of the standard canonic observer form (1). According to Theorem 3, one of the conditions for (3) to be an adaptive observer for (1) is persistency of excitation of the regressor $\phi(t, y(t))$. It is therefore natural to expect that some kind of persistency of excitation conditions might be needed for reconstruction of parameters $\theta, \lambda$ in (7) too. Two versions of these conditions will be considered, namely the notions of uniform persistency of excitation [26] and nonlinear persistency of excitation [8]:

**Definition 4** A function $\alpha : \mathbb{R}_{\geq 0} \times \Omega_{\lambda} \to \mathbb{R}^p$ is said to be $\lambda$-Uniformly Persistently Exciting ($\lambda$-UPE with $T, \mu$), denoted by $\alpha(t, \lambda) \in \lambda$-UPE$(T, \mu)$, if there exist $T, \mu \in \mathbb{R}_{>0}$:

$$
\int_t^{t+T} \alpha(\tau, \lambda) \alpha^T(\tau, \lambda) d\tau \geq \mu I_p, \quad \forall t \geq t_0, \lambda \in \Omega_{\lambda}.
$$

(31)

In contrast to the conventional definitions of persistency of excitation (cf. Definition 2), uniform persistency of excitation requires existence of $\mu, T \in \mathbb{R}_{>0}$ in (31) that are independent on $\lambda$ for all $\lambda \in \Omega_{\lambda}$. This is a stronger restriction; we will, however, require that the uniform persistency of excitation condition holds for $\varphi(t, \lambda, y(t))$ (as a function of $t$) in (7).

Since parametrization of (7) is allowed to be nonlinear, it is natural to expect that reconstruction of model parameters might require a nonlinear version of standard persistency of excitation. Here we employ the following generalization of the standard notion (cf. [9])

**Definition 5** Let $E$ be a set-valued map defined on $\mathbb{D} \subset \mathbb{R}^d$ and associating to a subset of $\mathbb{D}$ to every $p \in \mathbb{D}$. A function $\alpha : \mathbb{R} \times \mathbb{D} \times \mathbb{D} \to \mathbb{R}^k$ is said to be weakly Nonlinear Persistently Exciting in $p$ wrt $E$ (wNPE with $L, \beta, E$), denoted by $\alpha(t, p, p') \in \text{wNPE}(L, \beta, E)$, if there exist $L \in \mathbb{R}_{>0}, t_1 \geq t_0, \beta \in K_{\infty}$:

$$
\forall t \geq t_1, p, p' \in \mathbb{D} \exists \xi' \in [t, t+L) : \|\alpha(t', p, p')\| \geq \beta(\text{dist}(E(\xi), p')).
$$

(32)

If the set $E(p)$ contains just one element, $p$, then the inequality in (32) reduces to $\|\alpha(t', p, p')\| \geq \beta(\|p-p'\|)$. Taking the above notions into account we formulate the main technical assumption on the nonlinearities in (7):

**Assumption 4.1** The functions $\varphi, g$ in the right-hand side of (7) are such that

A1) the restriction of the function $\alpha_1 : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^m$, $\alpha_1(t, \lambda) = \varphi(t, \lambda, y(t))$ on $\mathbb{R}_{\geq 0} \times \Omega_{\lambda}$ is $\lambda$-UPE with $T, \mu$;

A2) the function $\alpha_2 : \mathbb{R} \times \mathbb{R}^{p+m} \times \mathbb{R}^{p+m} \to \mathbb{R}$, $\alpha_2(\lambda, \theta, X, Y) = \eta(\lambda, \theta, X, Y)$, where $\eta(\cdot)$ is defined in (19), is weakly nonlinearly persistently exciting in $(\lambda, \theta)$ w.r.t. the map $E(\lambda, \theta)$ determined by (18).

Remark 6 Checking that condition A1 holds is straightforward if e.g. $\varphi(t, \lambda, y(t))$ is periodic in $t$. Regarding condition A2, we note that, according to (19), $\eta(\lambda, \theta, X, Y)$ can be expressed as $\eta(t, \lambda, \theta, X, Y) = r(t, \lambda, \theta) - r(t, \lambda, \theta, X, Y) = \varphi(t, \lambda, y(t))^T(\theta - \theta')$, where

$$
r(t, \lambda, \theta, X, Y) = \varphi(t, \lambda, y(t))^T \theta + g_1(t, \lambda, y(t), u(t)) + \tilde{C}^T \int_t^{t+T} e^{\lambda(t-r)} \Omega(\lambda, \theta, y(t), u(r)) dr.
$$

If $\varphi, g$ are differentiable then $r(t, \lambda, \theta, X, Y) \leq m(t, \lambda, \theta, X') \leq \mathbb{R}(t, \lambda, \theta, X') (X - X')$, where $\mathbb{R}(t, \lambda, \theta, X') = \int_t^T \frac{\partial}{\partial \tau} r(t, s(\xi, \lambda, \theta)) d\xi, s(\xi, \lambda, \theta) = \lambda + (1 - \xi)X$. Hence

$$
\eta(t, \lambda, \theta, X, Y) = \varphi(t, \lambda, y(t))^T \Omega(t, \lambda, \theta, X, Y) \leq \mathbb{R}(t, \lambda, \theta, X') (\theta - \theta', X - X').
$$

(33)

It is therefore clear that if $(\varphi(t, \lambda, y(t))^T, \mathbb{R}(t, \lambda, \theta, X))$ is $\lambda, \lambda'$-uniformly persistently exciting, $\lambda, \lambda' \in \Omega_{\lambda}$, $\theta \in \Omega_{\theta}$, then the system is uniquely identifiable, it satisfies condition A2, and $E_0(\lambda, \theta)$ coincides with $E(\lambda, \theta)$.

We are now ready to state the main result:

**Theorem 7** Consider system (7) together with the observer defined by (22), (28)–(30). Suppose that Assumptions 3.1, 3.2, and 4.1 hold. Then there exist a constant $\gamma \in \mathbb{R}_{>0}$ and functions $r_1, r_2 \in \mathbb{K}$ such that if $\gamma, \varepsilon > r_1(\Delta_{\xi})$, then

$$
\limsup_{t \to \infty} \text{dist} \left( \begin{pmatrix} \lambda(t) \\ \theta(t) \end{pmatrix}, E(\lambda, \theta) \right) \leq r_2(\varepsilon).
$$

(34)

If, in addition, $E(\lambda, \theta)$ coincides with $E_0(\lambda, \theta)$ then there is an $r_3 \in \mathbb{K}$:

$$
\limsup_{t \to \infty} \| \hat{x}(t) - x(t) \| \leq r_3(\varepsilon).
$$

(35)

The proof of Theorem 3 is presented in the next section.
Let us briefly comment on assumptions made in the statement of the theorem. Assumptions 3.1, 3.2 are standard; condition A1 in Assumption 4.1 is the conventional requirement ensuring exponential convergence of $x(t), \hat{\theta}(t)$ to $x(t)$ and $\theta$ provided that the value of $\lambda$ and hence the values of $\varphi(t, \lambda, y(t))$ are known (cf. Theorem 3); A2 ensures that the distance from $(\lambda, \theta)$ to the set $\mathcal{E}(\lambda, \theta)$ is inferable from values of $y(t) = \hat{y}(t)$ over $[t_0, \infty)$ (cf. Lemma 12). Note that nonlinear dependence of $\eta$ on $\lambda, \lambda'$ can impose certain technical and computational difficulties whilst checking that this condition holds. Finally, observe that state estimation in the proposed scheme requires that $\mathcal{E}_0(\lambda, \theta) = \mathcal{E}(\lambda, \theta)$. This is not a mere technical condition, since as system (15) illustrates, violation of this assumption may prevent reconstruction of the state from observed output data.

The value of $\gamma$ and the functions $r_1, r_2, r_3$ could in principle be given explicitly. However, due to dependence of these functions on $A, B, C, \mu, \beta$ specified in Assumptions 3.1, 3.2 and Definitions 4, 5 explicit expressions for $\gamma$ and $r_1, r_2, r_3$ are too lengthy and thus are removed from the theorem's statement. They are, nevertheless, provided in the proof (see e.g., (60), (70)). A procedure for finding the values of $\gamma$ and $\varepsilon$ is discussed in Section 7.

The value of $\varepsilon$, viz. the accuracy of the estimation, is determined by the bound $\Delta_\varepsilon$ on the amplitude of perturbation $\xi(t)$. This dependency is established through the function $\gamma(\Delta_\varepsilon)$ determining a lower bound for parameter $\varepsilon$ in (28). If no perturbation $\xi(t)$ is present in the right-hand side of (7) then the value of $\varepsilon$ can be chosen arbitrarily small. Note that the convergence itself is asymptotic and not necessarily exponential. This is the price for the presence of unknown parameters $\lambda$ in (7).

Remark 8 The estimate $\hat{\lambda}(t)$ is guaranteed to converge to a single element of $\Omega_\lambda$ (see (61)); estimates $\hat{\theta}(t)$ may oscillate due the influence of $\xi(t)$. These oscillations are bounded, and will eventually be confined to the $2r_2(\varepsilon)$-neighborhood of $\mathcal{E}(\lambda, \theta)$. Hence, for $t_1$ sufficiently large and all $t \geq t_1 \geq t_0$, the $2r_2(\varepsilon)$-neighborhood of $(\lambda(t), \theta(t))$ will always contain an element of $\mathcal{E}(\lambda, \theta)$. This element may not necessarily be from $\Omega_\lambda \times \Omega_\theta$. If the elements of $\mathcal{E}(\lambda, \theta)$ are separated by distances exceeding $3r_2(\varepsilon)$ then the estimates are guaranteed to converge to the $r_2(\varepsilon)$-vicinity of just one element. This point in $\mathcal{E}(\lambda, \theta)$ will depend on $\xi(t), x_0$, and on the initial state of the observer.

Remark 9 The function $\beta$ in Definition 5 can be allowed to depend on $(\lambda, \theta)$. In view of Remark 6, this relaxes the requirement that $(\varphi(t, \lambda', y(t)) T, R(t, \lambda, \lambda', \theta))$ in (33) is $\lambda, \lambda', \theta$-UPE to the condition that the function $(\varphi(t, \lambda', y) T, R(t, \lambda, \lambda', \theta))$ is $\lambda'$-UPE. Note that this will make $r_2$ in (34) dependent on $(\lambda, \theta)$. Finally, note that A2 need not hold for all $(\lambda, \theta) \in \mathbb{R}^{\rho+m}$ and can be restricted to the union of $\Omega_{\lambda} \times \Omega_{\theta}$ and the domain to which $(\hat{\lambda}(t), \hat{\theta}(t))$ belong for $t \geq t_0$.

5 Proof of Theorem 7

According to Assumption 3.2 and (8), functions $\varphi, g$ and $\xi$ are continuous and are bounded in $\mathbb{R} \times \Omega_{\lambda} \times \mathcal{D}_y, \mathbb{R} \times \Omega_{\lambda} \times \mathcal{D}_q \times \mathcal{D}_u$ and $\mathbb{R}$ respectively. Therefore solutions of the combined system, (7), (22), (28)–(30) exist and are defined for all $t \geq t_0$. Let us denote $e = (e_1, e_2)$, $e_1 := \hat{x} - x$, $e_2 := \hat{\theta} - \theta$, $\alpha(t, \hat{\lambda}) = \varphi(t, \hat{\lambda}, y(t))$. Then according to (7) and (22) the following holds along the solutions of (7), (22), (28)–(30)

\[
\begin{align*}
\begin{pmatrix} \dot{e}_1 \\ \dot{e}_2 \end{pmatrix} &= \begin{pmatrix} A + \ell C^T & B \alpha(t, \hat{\lambda}) C^T \\ -\gamma_0 \alpha(t, \hat{\lambda}) C^T & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} + \begin{pmatrix} v(t, \hat{\lambda}, \lambda, y, u) \\ 0 \end{pmatrix}
\end{align*}
\]

(36)

where the function $v$: $v(t, \hat{\lambda}, \lambda, y, u) = B(\varphi(t, \hat{\lambda}, y) - \varphi(t, \lambda, y)) \theta + g(t, \hat{\lambda}, y, u) - g(t, \lambda, y, u) - \xi(t)$.

is continuous and bounded for all $y \in \mathcal{D}_y, u \in \mathcal{D}_u, \lambda, \hat{\lambda} \in \Omega_{\lambda}$ and $t \geq t_0$.

The proof of the theorem is split into three parts. In the first part we consider systems

\[
\begin{align*}
\begin{pmatrix} \dot{e}_1 \\ \dot{e}_2 \end{pmatrix} &= \begin{pmatrix} A + \ell C^T & B \alpha(t, \hat{\lambda}) C^T \\ -\gamma_0 \alpha(t, \hat{\lambda}) C^T & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}
\end{align*}
\]

(38)

where $\gamma_0 \in \mathbb{R}_{>0}$ and $\hat{\lambda} : \mathbb{R} \to \Omega_{\lambda}$ is a continuous, differentiable and bounded function. Let $\Phi(t)$ be a fundamental solution matrix of (38), and denote $\Phi(t, t_0) = \Phi(t)\Phi(t_0)^{-1}$.

(39)

Since $\Phi(t_0, t_0)$ is the identity matrix we say that $\Phi(t, t_0)$ is the normalized solution matrix of (38). We show that if Assumptions 3.1, 3.2 and condition A1 of Assumption 4.1 hold then there are positive numbers $M_\lambda, \rho, D_\rho$ such that the fundamental (normalized) matrix of solutions, $\Phi(t, t_0)$, of (38) with $\lambda$, satisfies the inequality $\|\Phi(t, t_0)p\| \leq D_\rho e^{-\rho(t-t_0)}\|p\|$, $p \in \mathbb{R}^{\rho+m}$.

Using this result, in the second part of the proof we demonstrate existence of $\gamma$ and $r \in \mathcal{K}$ such that setting $\gamma \in (0, \bar{\gamma}]$ and $\varepsilon > r(\Delta_\varepsilon), r \in \mathcal{K}$ is some function which will be specified later, ensures that the estimate $\hat{\lambda}(t)$ converges to a $\lambda^*$ from $\Omega_{\lambda}$: $\lim_{t \to \infty} \hat{\lambda}(t) = \lambda^*$. 


In the third part of the proof we show that condition A2 of Assumption 4.1 guarantees that (35) holds and that, subject to the condition that $E_0(\lambda, \theta)$ coincides with $E(\lambda, \theta)$, property (34) holds too.

Part 1 of the proof is contained in the following.

**Lemma 10.** Consider system (38), and suppose that

1. the matrices $A$, $B$, and $C$ satisfy Assumption 3.1
2. the function $\alpha(x)$ in the right-hand side of (38) is Lipschitz in $\lambda$,
3. for each $t \in \mathbb{R}$, the function $\alpha(t, \lambda)$ is Lipschitz in $\Omega_\lambda$,
4. the function $\alpha$ and its partial derivatives wrt $t, \lambda$ are bounded; that is there is a constant $\lambda$ such that

$$\max \{ \| \alpha(t, \lambda) \|, \| \alpha(t, \lambda) \|, \| \alpha(t, \lambda) \| \} \leq M \forall \lambda \in \Omega_\lambda, \forall t \in \mathbb{R}.$$

Then there exist positive numbers $\rho$, and $D_p$ such that

$$\| \hat{\lambda}(t) \| \leq M_\lambda \leq \mu_\rho \{ (2DMT^2) \}, \quad r \in [0,1]$$

(40)

the following holds

$$\| \Phi(t, s) \| \leq e^{-\rho(t-s)}D_p\| p \|, \quad p \in \mathbb{R}^{n+m},$$

(42)

where $t_2 \geq t_1 \geq t_0$, and $\Phi(t, s)$ is the fundamental (normalized) solution matrix of (38).

The proof of Lemma 10 and other auxiliary results are provided in Appendix.

Part 2. Consider now the interconnection of (7) with the observer (22), (28)–(30). The dynamics of the combined system in the coordinates $e, \hat{\lambda}$ is described by (36), (28)–(30). Recall that $e(t), \hat{\lambda}(t)$ are defined for all $t \geq t_0$. With respect to $e(t)$, the following holds

$$e(t) = \Phi(t, t_0)e(t_0) + \int_{t_0}^t \Phi(t, \tau) \left( \frac{\nu(\tau, \hat{\lambda}(\tau), \theta(\tau), \lambda(\tau))}{0} \right) d\tau.$$  

(43)

where $\Phi(t, t_0)$ is defined as in (39).

The variable $\hat{\lambda}$ in the combined system, as a function of $t$, is bounded, continuous, and differentiable with bounded derivative. Moreover, for any $A, B, C, \theta \in \Omega_\lambda$, $\hat{\lambda} \in \Omega_\lambda$ and for any choice of $\ell, \gamma_0$ in (22) we have that $|\dot{\lambda}(t)| \leq \gamma M_\sigma \max |\omega_i|^{\lambda_{i,max} - \lambda_{i,min}}$, where $M_\sigma, \omega_i$, and $\gamma$ are defined in (27), (30), and (28). Thus $\forall t \geq t_0$ we have

$$\hat{\lambda}(t) \in \Omega_\lambda, \| \hat{\lambda}(t) \| \leq \sqrt{\gamma M_\sigma \max |\omega_i|^{\lambda_{i,max} - \lambda_{i,min}}}$$

(44)

where $p$ is the dimension of $A$. According to Assumption 3.2, function $\alpha$ is bounded, differentiable, and Lipschitz in the second argument. In particular, $\| \alpha(t, \lambda) \| \leq B_\alpha, \| \partial_\lambda \alpha(t, \lambda) \| \leq D_\alpha \forall t \geq t_0, \lambda \in \Omega_\lambda$. Hence there is an $M = \max \{ B_\alpha, M_\sigma, D_\sigma \}$ such that condition C4 of Lemma 10 holds. Moreover, according to A1 in Assumption 4.1, $\alpha(t, \lambda) \in \Lambda(\mathbb{R}, T, \mu)$. This, together with (44), implies that requirements C1–C4 of Lemma 10 are satisfied.

Let $\gamma \in (0, \gamma_*^*], \gamma_*$ is defined as:

$$\gamma_* = \frac{\mu_r}{2D_\sigma \max \{ B_\alpha, M_\sigma, D_\sigma \} T^2}, \quad \lambda \in \Omega_\lambda.$$ 

(45)

This and (44) imply that the requirement (40) of the lemma is satisfied. Hence if $\gamma \in (0, \gamma_*^*]$ then according to Lemma 10 the fundamental matrix $\Phi(t, t_0)$ in (43) satisfies (42). This ensures existence of $\rho, D_p \in \mathbb{R}_{>0}$ such that the following holds along the solutions of (36), (28)–(30) for all $t \geq t_0$:

$$\| e(t) \| \leq e^{-\rho(t-t_0)}D_p \| e(t_0) \| + D_p \int_{t_0}^t e^{-\rho(t-\tau)} \| \nu(\tau, \lambda(\tau), \theta(\tau)) \| d\tau.$$ 

The functions $\nu, \theta$ in the definition of the function $v$, (37), are Lipschitz with respect to $\lambda$. Furthermore, according to (8) the following holds: $\| v(t) \| \leq \Delta_\lambda$. Therefore, in accordance to (37), (8), and Assumption 3.2

$$\| v(t, \hat{\lambda}(t), \theta(t)) \| \leq D_v \| \hat{\lambda}(t) \| - \lambda \| \hat{\lambda}(t) \| \in \Omega_\lambda,$$ 

(46)

$$D_v = \frac{D_p \| B \| \| \theta \| + D_\sigma}{\rho},$$ 

and hence

$$\| e(t) \| \leq e^{-\rho(t-t_0)}D_p \| e(t_0) \|$$

(47)

Let $j \in \{1, \ldots, p\}$, and consider the solution of

$$q_{2j-1} = \omega_j \cdot (q_{2j-1} - q_{2j-1}(q_{2j-1} + q_{2j-1})), \quad q_{2j} = \omega_j \cdot (q_{2j-1} + q_{2j-1} - q_{2j} + q_{2j}(q_{2j-1} + q_{2j-1})), $$

(48)

satisfying the initial condition $q_{2j-1}(t_0) = s_{2j-1}(t_0), q_{2j}(t_0) = s_{2j}(t_0)$; the parameters $\omega_j$ and values of $s_{2j-1}(t_0), s_{2j}(t_0)$ are supposed to coincide with those defined in (28), (29). The solution of (28) satisfying initial condition (29) is obviously unique, and can be expressed as a function $q : R \rightarrow R^p$:

$$q_{2j-1} = \cos(\omega_j t + \phi_j), \quad q_{2j} = \sin(\omega_j t + \phi_j), \quad \omega_j \in \mathbb{R}, j \in \{1, \ldots, p\}. $$

Parameters $\phi_j$ determined in accordance with $\cos(\omega_j t_0 + \phi_j) = s_{2j-1}(t_0), \sin(\omega_j t_0 + \phi_j) = s_{2j}(t_0)$. Given that $\omega_j$ in (48) are rationally independent we conclude that the $q$-limit set of $(q_1(t, s_0), q_2(t, s_0), \ldots, q_{2p-1}(t, s_0))$ is the hypercube
Consider the function $\beta : \mathbb{R}^{2p} \rightarrow \mathbb{R}^p$:

$$
\beta_j(q) = \lambda_{j,\min} + \frac{\lambda_{j,\max} - \lambda_{j,\min}}{2} (q_{j-1}(t) + 1),
$$

and define $\hat{\lambda}(t) = \beta(q(t, s_0))$. System (48), (49) satisfies conditions P1–P3, and hence we can conclude that trajectory $\hat{\lambda}(t)$ satisfies the recurrence property (24):

$$
\forall \lambda \in \Omega_\lambda, \Delta_\lambda \in \mathbb{R}_{>0}, t \geq t_0
\Rightarrow \exists t' > t : \|\lambda - \hat{\lambda}(t')\| < \Delta_\lambda.
$$

Noticing that $s(t, s_0) = q(T(t), s_0)$, where $T(t) = t_0 + \gamma \int_{t_0}^{t} \sigma(\|y(\tau) - \hat{y}(\tau)\|_\varepsilon) d\tau$, we can conclude that for all $t \geq t_0$ the function $\hat{\lambda}(t)$ defined in (28) can be expressed in terms of $\lambda(T(t))$ as

$$
\hat{\lambda}(t) = \lambda \left( t_0 + \gamma \int_{t_0}^{t} \sigma(\|y(\tau) - \hat{y}(\tau)\|_\varepsilon) d\tau \right).
$$

Denoting $h(t) = t' - t_0 - \gamma \int_{t_0}^{t} \sigma(\|y(\tau) - \hat{y}(\tau)\|_\varepsilon) d\tau$, where the value of $t'$ is chosen such that (50) holds, we arrive at the following estimate:

$$
\|\lambda - \hat{\lambda}(t)\| \leq \|\lambda - \hat{\lambda}(t')\| + \|\hat{\lambda}(t') - \hat{\lambda}(t)\|
= \|\lambda - \hat{\lambda}(t')\| + \|\hat{\lambda}(t') - \hat{\lambda}(t' - h(t))\|.
$$

The function $\hat{\lambda}(t)$ is Lipschitz with respect to $t$:

$$
\|\hat{\lambda}(t_1) - \hat{\lambda}(t_2)\| \leq \sqrt{\max_i \|\omega_i(\lambda_{i,\max} - \lambda_{i,\min})\| |t_1 - t_2|}, \quad t_1, t_2 \in \mathbb{R}_{\geq t_0}.
$$

Thus (51), (52) imply that

$$
\|\lambda - \hat{\lambda}(t)\| \leq \|\lambda - \hat{\lambda}(t')\| + \|\hat{\lambda}(t') - \hat{\lambda}(t)\| \leq \Delta_\lambda.
$$

Taking (47) and (53) into account we can conclude that the dynamics of the combined system (7), (22), (28)–(30) obeys the following set of constraints:

$$
\|e(t)\| \leq e^{-\rho(t-t_0)}D_\rho\|e(t_0)\| + \frac{D_\rho D_\varepsilon D_\lambda}{\rho} \|\tau(\|\hat{\lambda}(t')\|_\varepsilon)\|_{\infty, [t_0, t]}
+ \frac{D_\rho D_\varepsilon D_\Delta}{\rho} \|\hat{\lambda}(t')\|_{\infty, [t_0, t]}
+ \frac{D_\rho D_\varepsilon D_\Delta}{\rho} \|\hat{\lambda}(t') - \hat{\lambda}(t)\| \leq \Delta_\lambda,
$$

$$
h(t) = h(t_0) - \gamma \int_{t_0}^{t} \sigma(\|C^T e_1(\tau)\|_\varepsilon) d\tau.
$$

To proceed further we need an auxiliary result below.

**Lemma 11** Consider a system of which the dynamics satisfy the following inequalities:

$$
\|\dot{x}(t)\| \leq g(t - t_0)\|x(t_0)\| + c\|h(\tau)\|_{\infty, [t_0, t]} + \Delta,
$$

$$
- \int_{t_0}^{t} \gamma_0(\|x(\tau) + \dot{d}(\tau)\|_\varepsilon) d\tau \leq h(t) - h(t_0) \leq 0, \quad \forall t \geq t_0,
$$

where $x : \mathbb{R} \rightarrow \mathbb{R}^n$, $h : \mathbb{R} \rightarrow \mathbb{R}$ are trajectories corresponding to the evolution of the system's state, $d : \mathbb{R} \rightarrow \mathbb{R}^n$, $\|d(\tau)\|_{\infty, [t_0, \infty)} \leq \Delta_d$ is a continuous and bounded function on $[t_0, \infty)$, $\gamma$ is a strictly monotonically decreasing function with, $\gamma(0) \geq 1$, $\lim_{s \rightarrow \infty} \gamma(s) = 0$; $c, \Delta \in \mathbb{R}_{>0}$, and $\gamma_0 : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$:

$$
|\gamma_0(s)| \leq D_\gamma |s|, \quad D_\gamma \in \mathbb{R}_{>0}.
$$

Then trajectories $x(t)$, $h(t)$ in (55) are bounded in forward time, for $t \geq t_0$, provided that the following conditions hold for some $d \in (0, 1)$, $\kappa \in (1, \infty)$:

$$
\varepsilon \geq \Delta \left( 1 + \frac{\rho D_\rho}{\rho \kappa - \kappa - d} \right) + \Delta_d,
$$

$$
D_\gamma \leq \frac{\kappa - 1}{\kappa} \left[ \left( \frac{d}{\rho} \right)^{\kappa - 1} \right] \times
\frac{h(t_0) - \delta_0(\|x(t_0)\| + |h(t_0)|c(1 + \kappa \rho(0)/(1 - d)))}{\|x(t_0)\| + |h(t_0)|c(1 + \kappa \rho(0)/(1 - d))}.
$$

The proof of Lemma 11 is provided in the Appendix. Notice that $h(t)$ in (54) satisfies $-\gamma D_\sigma \int_{t_0}^{t} \sigma(\|e(\tau)\|_\varepsilon) d\tau \leq h(t) - h(t_0) \leq 0$. Indeed, $\|C^T e_1\|_\varepsilon \leq \|e\|_\varepsilon$ by virtue of definition of $\| \cdot \|_\varepsilon$ and $C$, and the function $\sigma$ in (54) is Lipschitz (see (27)). Thus (54) is of the form (55) where

$$
c = D_\rho D_\varepsilon D_\lambda / \rho, \quad \Delta = D_\rho D_\varepsilon D_\lambda / \rho + D_\rho D_\omega / \rho, \quad e(s) = D_\rho e^{-\rho s}.
$$

Notice also that because (50) holds, the value of $t'$ in (53) can be chosen arbitrarily large. This implies that the value of $h(t_0)$ in (54) may be chosen arbitrarily large too. Having this in mind, and invoking Lemmas 10, 11 we can conclude that choosing $\varepsilon, \gamma$ in (28) as

$$
\varepsilon \geq \gamma_0(\Delta), \quad \gamma_0(\Delta) = \Delta \left( 1 + \frac{\rho D_\rho}{\rho \kappa - \kappa - d} \right),
$$

$$
0 < \varepsilon < \gamma = \min\{\gamma^*, D_\gamma\},
$$

$$
D_\gamma = \frac{\kappa - 1}{\kappa} \left[ \ln \left( \frac{D_\rho}{\rho} \right) \right]^{-1} \frac{\rho}{c(1 + \kappa D_\rho/(1 - d))},
$$

where $\gamma^*$ is defined as in (45), ensures that the function $h(t)$ in (54) is bounded. Given that $h(t)$ by construction is monotone and bounded, the Bolzano-Weierstrass theorem implies that $h(t)$ converges to a limit, and hence

$$
\lim_{t \rightarrow \infty} \int_{t_0}^{t} \sigma(\|C^T e_1(\tau)\|_\varepsilon) d\tau = \bar{h}, \quad \bar{h} \in \mathbb{R},
$$

$$
\lim_{t \rightarrow \infty} \hat{\lambda}(t) = \lambda^*, \quad \lambda^* \in \Omega_\lambda.
$$

Noticing that $\sigma(\|C^T e(\tau)\|_\varepsilon)$ is uniformly continuous and
using Barbalat’s lemma we conclude that
\[
\lim_{t \to \infty} \sup_{\tau \in [T, \infty)} \| C^T e_1(\tau) \| = \varepsilon. \tag{62}
\]

Part 3. Consider the equation for \( \dot{e}_1 \) in (36), and rewrite it as:
\[
\dot{e}_1 = (A + \ell C^T) e_1 + v_1(t, \theta, \lambda, \theta, \lambda) + v_2(t, \dot{\lambda}, \lambda) + v_3(t)
\]
where \( v_2(t) = -\xi(t) \) and
\[
v_1(t, \theta, \lambda, \theta, \lambda) = B(\varphi^T(t, \lambda, y(t)) \dot{\theta} - \varphi^T(t, \lambda, y(t)) \theta)
\]
\[
v_2(t, \lambda, \lambda) = g(t, \lambda, y(t), u(t)) - g(t, \lambda, y(t), u(t)). \tag{64}
\]

Next steps are based on the following lemma.

**Lemma 12**  
Consider
\[
\dot{x} = Ax + u(t) + d(t), \\
y = C^T x, \quad x(t_0) = x_0, \quad x_0 \in \mathbb{R}^n,
\]
where \( A \) and \( C \) are defined as in (1), and \( u, d : \mathbb{R} \to \mathbb{R}^n, u \in C^1, d \in C \). Let \( u, d, \) be bounded: 
\[
\max\{\|u(t)\|, \|\dot{u}(t)\|\} \leq B, \|d(t)\| \leq \Delta \xi \leq t \geq t_0.
\]

Then there are \( \kappa_1, \kappa_2 \in \mathbb{K} \):
\[
\|y(\tau)\|_{\infty, [t_0, \infty)} \leq \varepsilon \Rightarrow \exists t'(\varepsilon) \geq t_0 : \\
\|z_1(\tau) + u_1(\tau)\|_{\infty, [t', \infty)} \leq \kappa_1(\varepsilon) + \kappa_2(\Delta \xi),
\]
where \( z_1 = (1, 0, \ldots, 0)z \),
\[
\dot{z} = \Lambda z + Gu, \quad \Lambda = \begin{pmatrix} -b & I_{n-2} \\ 0 & 0 \end{pmatrix}, \tag{66}
\]
\[
G = \begin{pmatrix} -b \end{pmatrix}, \quad z(t_0) = 0,
\]
\[
\text{and } b = (b_1, \ldots, b_{n-1})^T : \text{real parts of the roots of } s^{n-1} + b_1 s^{n-2} + \cdots + b_{n-1} \text{ are negative.}
\]

Moreover, if \( d(t) \equiv 0 \), then
\[
y(t) = 0 \forall t \geq t_0 \Rightarrow \exists p \in \mathbb{R}^{n-1} : \forall t \geq t_0 \tag{67}
(1, 0, \ldots, 0)e^{\Lambda(t-t_0)}p + z_1(t) + u_1(t) = 0.
\]

The proof of Lemma 12 is provided in the Appendix. According to (8) and Assumption 3.2, \( v_1, v_2, v_3 \) and \( \dot{v}_1, \dot{v}_2 \) as functions of \( t \), are bounded. Thus assumptions of Lemma 12 are met for equations (63), (64), and hence (62) implies that there is a \( t_1(\varepsilon) \geq t_0 \) and \( \kappa_1, \kappa_2 \in \mathbb{K} \) such that \( \forall t \geq t_1(\varepsilon) \) we have:
\[
\|\varphi^T(t, \dot{\lambda}(t), y(t)) \dot{\theta}(t) - \varphi^T(t, \lambda, y(t)) \theta + v_{2,1}(t, \dot{\lambda}(t), \lambda) \| C^T \int_{t_0}^t e^{\Lambda(t-\tau)} G v_2(\tau, \dot{\lambda}(\tau), \lambda) d\tau \| \leq \kappa_1(\varepsilon) + \kappa_2(\Delta \xi),
\]
where \( G \in \mathbb{R}^{(n-1) \times n}, \quad \dot{C} \in \mathbb{R}^{n \times 1} \) are defined as in (18)–(20), and \( v_{2,1}(t) \) is the first component of \( v_2(t) \). Given that
\[
\int_{t_0}^t e^{\Lambda(t-\tau)} G v_2(\tau, \dot{\lambda}(\tau), \lambda) d\tau = \int_{t_0}^t e^{\Lambda(t-t)} G v_2(\tau, \lambda, \lambda) d\tau + \int_{t_0}^t e^{\Lambda(t-t)} G v_2(\tau, \dot{\lambda}(\tau), \lambda) d\tau - v_{2,1}(\tau, \lambda, \lambda) d\tau,
\]
noticing that \( \Lambda \) is Hurwitz and that, according to (61) \( v_2(\lambda, \lambda) \to v_2(\lambda, \lambda) \to 0 \) as \( t \to \infty \), we can conclude that there is a \( t_2(\varepsilon) \geq t_1(\varepsilon) \) such that
\[
\|\eta(t, \theta(t), \lambda^*, \theta, \lambda)\| \leq \kappa_1(\varepsilon) + \kappa_2(\Delta \xi) \text{ as defined in (19) satisfies:}
\]
\[
\forall t \geq t_2(\varepsilon) \leq \kappa_1(\varepsilon) + \kappa_2(\Delta \xi) + \varepsilon \tag{68}
\]
Recall that \( \alpha(t, (\lambda', \theta), (\lambda, \theta)) = \eta(t, \theta, \lambda', \theta, \lambda) \in wNPE(\mathbb{L}, \beta, \mathcal{E}) \). Let \( t_3(t) \) be such that \( \|C^T e_1(t)\| < 2\varepsilon \) for all \( t \geq t_3(t) \) (existence of such \( t_3(t) \) follows from (62)). Consider the sequence \( \{\tau_i\}_{i=0}^{\infty}, \tau_i = \max\{t_3, t_2(t)\} + L \). It is clear that, since the function \( \varphi(\cdot, \lambda(\cdot), y(\cdot)) \) is bounded, there is an \( M \in \mathbb{R}_{>0} \):
\[
\|\dot{\theta}(t) - \hat{\theta}(\tau_i)\|_{\infty, [t, t+1]} \leq \varepsilon 2\gamma_0 B \phi L = \varepsilon M \tag{69}
\]
for all \( t \geq t_0 \). Hence \( \forall t : \gamma(t) = \tau \in [\tau_i, \tau_{i+1}], i \in \mathbb{N}, \) we have:
\[
\|\eta(t, \theta(\tau_i), \lambda^*, \theta, \lambda)\| \leq \kappa_1(\varepsilon) + \kappa_2(\Delta \xi) + \varepsilon (MB_\varphi + 1).
\]
This, however, implies that there is an \( N \in \mathbb{N} \) such that
\[
\text{dist}\left(\begin{pmatrix} \lambda^* \\ \hat{\theta}(\tau_i) \end{pmatrix}, \mathcal{E}(\lambda, \theta)\right) \leq \beta^{-1}(\kappa_1)(\varepsilon) + \kappa_2(\Delta \xi) + \varepsilon (MB_\varphi + 1) \]

for all \( i \geq N \). Therefore, taking (44), (69) into account, we can conclude that there is a \( t'(\varepsilon) \):
\[
\text{dist}\left(\begin{pmatrix} \lambda^* \\ \hat{\theta}(\tau_i) \end{pmatrix}, \mathcal{E}(\lambda, \theta)\right) \leq 2\varepsilon M + \beta^{-1}(\kappa_1) + \beta^{-1}(\kappa_2(\Delta \xi) + \varepsilon (MB_\varphi + 1)) \forall t \geq t'(\varepsilon).
\]
Notice that \( r_0 \) in (60) is a class \( \mathcal{K}_\infty \) function of \( \Delta \). The latter parameter, \( \Delta \), as defined in (59), is the sum: \( \Delta = P_\varphi P_\Delta P_\Delta + P_\varphi P_\Delta \). Given that the value of \( \Delta \lambda \) can be chosen arbitrarily, we can pick \( \Delta \lambda = \Delta \xi \). This renders the variable \( r_0 \) in (60) a class \( \mathcal{K}_\infty \) (and hence class \( \mathcal{K} \)) function of \( \Delta \xi \). Denote this function as \( r_1(\cdot) \), and hence \( \varepsilon > r_1(\Delta \xi) \) implies that
\[
\beta^{-1}(\kappa_1(\varepsilon)) + \kappa_2(\Delta \xi) + \varepsilon (MB_\varphi + 1) + 2\varepsilon M < \beta^{-1}(\kappa_1(\varepsilon)) + \kappa_2(\beta^{-1}(\varepsilon(\Delta \xi)) + \varepsilon (MB_\varphi + 1) + 2\varepsilon M = r_2(\varepsilon).
\]
Thus (34) holds.

Finally, if $\mathcal{E}(\lambda, \theta)$ coincides with $\mathcal{E}_1(\lambda, \theta)$, then Assumption 3.2 and (34) imply that $\|v_1(t, \hat{\lambda}, \lambda, \lambda) + v_2(t, \hat{\lambda}, \lambda)\| < M_1r_2(\varepsilon)$ for some $M_1 \in \mathbb{R}_{>0}$. Since $A + \ell \mathcal{C}^T$ in (63) is Hurwitz, statement (35) follows. □

6 Discussion and generalization

6.1 Removing passivity requirement (Assumption 3.1)

Theorem 7 requires that $A, B, C$ in (7) satisfy Assumption 3.1. Here, we illustrate the idea of filtered transformations [28], [29] to show how observer (22), (28) can be modified so that this condition could be replaced with the requirement that the pair $A, C$ is observable. Consider a generalization of (7)

$$x = Ax + \Psi(t, \lambda, y)\theta + g(t, \lambda, y, u) + \xi(t),$$

$$y = C^T x, \quad A = \begin{pmatrix} 0 & I_{n-1} \\ 0 & 0 \end{pmatrix}, \quad C = \text{col}(1, 0, \ldots, 0),$$

where $\Psi : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^{n \times m}, \Psi \in C^1$, is Lipschitz in $\lambda$, and $\Psi(\cdot, \lambda, y(\cdot)), \Psi(\cdot, \lambda, y(y(\cdot)))$ are bounded. The function $\xi$ and parameters are defined as in (7), and the function $g$ satisfies Assumption 3.2.

Let $B = \text{col}(1, b_1, \ldots, b_{n-1})$ be a vector such that the polynomial $s^{n-1} + b_1 s^{n-2} + \cdots + b_{n-1}$ is Hurwitz.

As an observer candidate for (71) we propose a system in which $S_w$ is defined as in (28), and $\hat{S}_w$ is given as follows:

$$\dot{M} = (A - BC^T A)M + (I_n - BC^T)\Psi(t, \hat{\lambda}, y),$$

$$\dot{\xi} = A\xi + \ell (C^T \xi - y) + B\varphi^T(t, \lambda, y, [\lambda, y])\theta$$

$$+ g(t, \lambda, y, u),$$

$$\dot{\theta} = -\gamma_0 (C^T \xi - y)\varphi(t, \lambda, y, [\lambda, y]), \quad \gamma_0 \in \mathbb{R}_{>0},$$

$$\dot{x} = \xi + M\theta, \quad M = \text{col}(1, 0, \ldots, 0),$$

where

$$\varphi^T(t, \lambda, y, [\lambda, y]) = C^T A M(t, [\lambda, y]) + C^T \Psi(t, \lambda, y).$$

The first row of $M$ is zero for all $t \geq t_0$, and $\dot{y} = C^T \dot{x} = C^T \dot{\xi}$. Moreover, since $\Psi(\cdot, \lambda(y(\cdot)), y(\cdot))$ is bounded and Lipschitz in $\lambda$, $M(\cdot, [\lambda, y]), M(\cdot, [\lambda, y])$ are globally bounded for all $t \geq t_0$, and $M(t, [\lambda, y])$ is Lipschitz in $\lambda$ for $\lambda = \text{const}$. Let $\xi = x - M\theta$, then using (71)–(73) we can write

$$\dot{\xi} = A\xi + B\varphi(t, \lambda, y, y, [\lambda, y])\theta + (\Psi(t, \lambda, y, u)$$

$$- \Psi(t, \lambda, y, u))\theta + g(t, \lambda, y, u) + \xi(t).$$

Dynamics of (71), (72) in the coordinates $e_1 = \dot{\xi} - x + M\theta, e_2 = \theta - \theta$ is

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} A + \ell C^T & B\varphi^T(t, \hat{\lambda}, y, [\hat{\lambda}, y]) \\ -\gamma_0 \varphi(t, \hat{\lambda}, y, [\hat{\lambda}, y])C^T & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} + \begin{pmatrix} \tilde{v}(t, \hat{\lambda}, y, u) \\ 0 \end{pmatrix}$$

where $\tilde{v}(t, \hat{\lambda}, y, u) = (\Psi(t, \hat{\lambda}, y) - \Psi(t, \lambda, y))\theta + g(t, \lambda, y, u) - g(t, \lambda, y, u) - \xi(t)$.

Finally, consider the function $\eta_1$

$$\eta_1(t, \lambda, \theta, \lambda', \theta') = \varphi^T(t, \lambda', y(t), [\lambda', y])\theta' - \theta + C^T(\Psi(t, \lambda', y(t)) - \Psi(t, \lambda, y(t))\theta + g(t, \lambda', y(t), u(t)) - g(t, \lambda, y(t), u(t))) + q(t, \lambda', \theta),$$

where $q(t, \lambda', \theta, \lambda, \theta) = z \tilde{z} = \lambda z + G((\Psi(t, \lambda', y(t)) - \Psi(t, \lambda, y(t)))\theta + g(t, \lambda', y(t), u(t)) - g(t, \lambda, y(t), u(t))),$ $z(t_0) = 0,$ and $\tilde{C}, \tilde{\lambda}, \tilde{G}$ are defined as in (20). The following is now immediate

Theorem 13 Consider (71), (72), (28)–(30). Suppose that condition A1 of Assumption 4.1 holds for the function $\alpha_1 : \mathbb{R}_{>0} \times \Omega \to \mathbb{R}^m, \alpha_1(\lambda) = \varphi(t, \lambda, y(t), [\lambda, y]),$ where $\varphi$ is defined as in (72). Furthermore, let the function $\alpha_4 : \mathbb{R} \times \mathbb{R}^{p+m} \times \mathbb{R}^{p+m} \to \mathbb{R},$ $\alpha_4(t, \lambda, \theta, \lambda', \theta') = \eta_1(t, \lambda, \theta, \lambda', \theta')$ be weakly nonlinearly persistently exciting in $(\lambda, \theta)$ w.r.t. to the map $E_1:

$$E_1(\lambda, \theta) = \{(\lambda', \theta'), \lambda' \in \mathbb{R}^p, \theta' \in \mathbb{R}^m | B(\theta' - \theta)^T \varphi(t, \lambda, y(t), [\lambda, y]) + (\Psi(t, \lambda', y(t)) - \Psi(t, \lambda, y(t)))\theta$$

$$+ g(t, \lambda, y(t), u(t)) - g(t, \lambda, y(t), u(t)) = 0, \forall t \geq t_0 \}.$$

Then there exist a constant $\tilde{\gamma} \in \mathbb{R}_{>0}$ and functions $r_1, r_2, r_3 \in \mathbb{K}$ such that if $\gamma, \varepsilon$ are the corresponding parameters of (28), and $\gamma \in (0, \tilde{\gamma}), \varepsilon > r_1(\Delta \xi),$ then (35), (34) hold (with $E$ replaced by $E_1$) for the interconnection (71), (72), and (28).

The proof is largely identical to that of Theorem 7 (a sketch is presented in the Appendix). According to Remarks 6, 9 one can replace the requirement that $\alpha_4(t, \lambda, \theta, \lambda', \theta') \in \text{wNEPE}(L, \beta, E_1)$ with that of the $\lambda'$-uniform persistency of excitation of the function $\alpha_5$:

$$\alpha_5(t, \lambda') = (\varphi(t, \lambda', y(t), [\lambda', y]), R_1(t, \lambda, \lambda', \theta, \theta)),$$

where $R_1(t, \lambda, \lambda', \theta) = \int_0^1 \frac{\partial}{\partial \lambda} r_1(t, s(\xi, \lambda', \lambda), \theta) d\xi,$ $s(\xi, \lambda, \lambda') = \lambda' \xi + (1 - \xi) \lambda,$ and $r_1(t, \lambda, \theta) = C^T$.
\[\Psi(t, \lambda, y(t))\theta + g_1(t, \lambda, y(t), u(t))+\tilde{C}^T\int_{t_0}^t e^{\lambda(t-\tau)} G(\Psi(\tau, \lambda, y(\tau)))\theta + g(\tau, \lambda, y(\tau), u(\tau))d\tau.\]

Let us now consider systems (9). Since \(A, C\) is observable, there is a coordinate transformation \(x \rightarrow T(A)x\) bringing system (9) into the form (71), albeit with the functions \(\Psi, g\) and vector \(\theta\) defined differently. An example illustrating the viability of this approach is provided in Section 7. Notice also that observability of \(A, C\) implies that the system \(\dot{x} = Ax + \tilde{P}(t, \lambda, x)\theta + \tilde{g}(t, \lambda, x, u) + \xi(t), y = CTx\), in which the functions \(\tilde{P}, \tilde{g}\) are bounded and Lipschitz in \(x\) can be brought into the form (71) by using an auxiliary high-gain observer (cf part 3 of the proof).

6.2 Presence of measurement noise

Suppose now that observations of the unmodeled dynamics and is supposed to be unavailable for direct observation.

Let the task be to infer the values of \(\alpha, \lambda, y\) from the measurements of \(y\) over time. System (76) belongs to the class of equations described by (7) with \(\varphi(t, \lambda, y) = 1 \forall t, \lambda, y, \text{ and } g(t, \lambda, y, u) = B(\sin(\lambda \cos(t)) + e^{\lambda s\sin(t)}).\) Moreover, it satisfies Assumption 3.1 with

\[P = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 6 & -3 \\ -3 & 2 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\]

and Assumption 3.2 with \(D_{\theta} = 0, D_y = B_y = M_y = \sqrt{\gamma(1+e)}, d_{\gamma} = 1\) and \(M_{\theta} = 0\). According to (22)–(28) the observer candidate is:

\[\dot{x} = Ax + B\theta + B(\sin(\lambda \cos(t)) + e^{\lambda s\sin(t)})\]

\[\dot{\theta} = -\gamma_\theta(C^T\dot{x} - y)\]

\[\dot{s_{1,1}} = \gamma \tanh(|C^T\dot{x} - y|)(s_{1,1} - s_{2,1}) - s_{1,1}(s_{2,1}^2 + s_{2,2}^2)\]

\[\dot{s_{2,1}} = \gamma \tanh(|C^T\dot{x} - y|)(s_{1,1} + s_{2,1}) - s_{2,1}(s_{2,1}^2 + s_{2,2}^2)\]

\[\lambda = 0.1 + 0.45(s_{1,1} + 1)\]

Parameters \(\gamma, \gamma_\theta\), and \(\varepsilon\) will be specified in a later stage.

It is clear that the function \(\varphi(t, \lambda, y) = 1\) is a UPE. This assures that condition A1 in Assumption 4.1 holds. Finally, notice that the function \(\eta\) as defined in (19), in this case becomes:

\[\eta(t, \lambda, \theta, \lambda', \theta') = \theta - \theta' + \sin(\lambda \cos(t)) - e^{\lambda s\sin(t)}\]

One can also check that the corresponding function \(\alpha(t, \lambda, \theta, \lambda', \theta')\) is weakly nonlinearly persistently exciting wrt \(\theta, \lambda\), and thus A2 is satisfied. According to Theorem 7, system (76) is an adaptive observer for the unmodeled dynamics and is supposed to be unavailable for direct observation. Let us now show how this procedure works in the example.

According to the theorem, parameter \(\gamma_\theta\) is an arbitrary positive number; here, for simplicity, we set \(\gamma_\theta = 1\). With regards to parameters \(\gamma, \varepsilon\), they are to satisfy condition (59), (60). The choice of parameter \(\gamma\) is essentially subjected to two constraints. The first constraint is \(\gamma \in (0, \gamma^*]\), where \(\gamma^*\) is specified in (45). This constraint ensures that the function \(\varphi(t, \lambda(t), y(t))\) is persistently exciting. In our case, since the variable \(\varphi(t, \lambda(t), y(t))\) is independent of \(\lambda(t)\), this property holds for any \(\gamma > 0\). The second constraint is \(\gamma < \frac{\varepsilon\lambda}{\kappa} \left|\ln\left(D_0 \frac{s}{\delta}\right)\right|^{-1} \frac{\rho}{e(1+\kappa D_0 s)(1-d)}\), \(\kappa > 0\).

7 Examples

Consider the following system:

\[\dot{x} = Ax + B\theta + B(\sin(\lambda \cos(t)) + e^{\lambda s\sin(t)}) + \xi(t), y = CTx,\]

where \(\theta \in [0,1] = \Omega_0, \lambda \in [0,1] = \Omega_\lambda\) are unknown parameters, and \(x_0\) is only partially known. The function \(\xi : R \rightarrow R^2, \xi(t) = 0.001\cos(t) \cos(t)\), stands for the unmodeled dynamics and is supposed to be unavailable for direct observation.

\[A = \begin{pmatrix} -2 & 1 \\ -1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad C = \text{col}(1,0,\ldots,0),\]

for the unmodeled dynamics and is supposed to be unavailable for direct observation.
and \( \rho = 0.5 \), \( D_\rho = 4.242 \). Picking \( d = 0.2 \), \( \kappa = 2 \) results in \( \gamma = 0.00286 \). Since the values of \( d, \kappa \) are now defined, we can set the value of the parameter \( \varepsilon \). Taking (59), (60) into account and noticing that \( \Delta_\lambda = 0 \) (this is because trajectories of (78) in which the term \( \tanh(\|C^T x - y\|_\varepsilon) \) is replaced with 1 will pass through every point of the set \( \Omega_\lambda = [0.1, 1] \)), we arrive at \( \varepsilon \geq \frac{D_\Delta}{\Delta_\varepsilon} \), where \( \Delta_\varepsilon : \| \xi(t) \| \leq \Delta_\varepsilon \). Given that \( \Delta_\varepsilon = 0.001 \) leads to \( \varepsilon \geq 0.018 \).

Results of computer simulations of the combined system (76)–(78) with parameters \( \theta = 0.2, \lambda = 0.7, \gamma = 0.0028 \), \( \gamma_\theta = -1 \), \( \varepsilon = 0.018 \) are summarized in Fig. 2. As can be observed, the system has two weakly attracting sets (marked as white circles). These sets correspond to the true parameter values of \( \theta \) and \( \lambda \). Even though trajectories of the system are converging to the attracting sets asymptotically, small neighborhoods of these sets are not forward-invariant. This implies that the sets themselves are not stable in the Lyapunov sense, albeit clearly attracting. Middle panel depicts typical trajectories of \( \lambda \) and \( \theta \). To show how the proposed observer behaves in presence of measurement noise we simulated the model-observer system in which signal \( y(t) = x_1(t) \) in the observer subsystem was replaced with \( y_\varepsilon(t) = \sin(\sin(2t) \varepsilon) \). The value of \( \varepsilon \) was changed to 0.068 to account for this perturbation. It can be seen that the observer retained functionality, albeit, as expected, with lower precision of estimation.

In order to illustrate what the behavior of the system would look like in case of multiple equivalent parameterizations, we simulated a modified version of the combined system (76)–(78), in which the nonlinearly parameterized terms, i.e. \( B(\sin(\lambda \cos(t)) + e^{\lambda \sin(t)}) \) in (76) and \( B(\sin(\lambda \cos(t)) + e^{\lambda \sin(t)}) \) in (77), are replaced with \( B(\sin((\lambda - 0.45)^2 \cos(t)) + e^{(\lambda - 0.45)^2 \sin(t)}) \) and \( B(\sin((\lambda - 0.45)^2 \cos(t)) + e^{(\lambda - 0.45)^2 \sin(t)}) \) respectively. Assumptions 3.1, 3.2 and A1 in Assumption 4.1 still hold for the modified system (with the same values of parameters). Yet, system (76) with the modified \( g(t, \lambda) = B(\sin((\lambda - 0.45)^2 \cos(t)) + e^{(\lambda - 0.45)^2 \sin(t)}) \) is no longer uniquely identifiable since \( g(t, 0.7) = g(t, 0.2) \)

\[ \Phi(t, t_0) = e^{A_1(t-t_0)}, \quad A_1 = \begin{pmatrix} -2 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix}, \]

for all \( t \). Simulation results of the modified system are presented in Fig. 2, right panel. Instead of two attractors as in the previous configuration, the modified system has four weakly attracting sets corresponding to two equivalent parameterizations \( \theta = 0.2, \lambda = 0.7 \) (true) and \( \theta = 0.2, \lambda = 0.2 \) (spurious). Parameters estimates converge to small vicinities of these alternative parameterizations. Note that the estimates do not jump between neighborhoods of \( \theta = 0.2, \lambda = 0.7 \) and \( \theta = 0.2, \lambda = 0.2 \), which is consistent with Remark 8.

Finally, we illustrate applicability of our approach to models (9). For this purpose consider the third example from Table 1 with the nominal values of parameters as follows: \( \tau_m = 0.1666, \tau_\varepsilon = 5, A_f = 1, \sigma_f = 2, \sigma_s = 0.8 \). Suppose that true values of these parameters are unknown, but it is known that they are within \( \pm 25\% \) of their nominal values. Since the pair \( A, C \) is observable, there is a parameter-dependent coordinate transformation \( x \mapsto Tx, T = \begin{pmatrix} 1 & 0 & \xi_{m,1} \\ 0 & 0 & \xi_{m,2} \\ \xi_{m,3} & \xi_{m,4} & \xi_{m,5} \end{pmatrix} \), \( \xi_m \rightarrow T^{-1}(\xi_m) \) rendering the original equations into (71) with \( \Psi(t, \lambda, \gamma) = \begin{pmatrix} y \tanh(\lambda y) & 0 & 0 \\ 0 & y \tanh(\lambda y) \end{pmatrix} \), \( g = 0 \) and \( \theta = ( -1, -1, -1, -1, -1, -1 )^T \). Note that \( A, \Psi \) and \( \theta \) differ from those in the original parametrization. Let \( B = (1, 1)^T \) and consider \( M(t, [\lambda, y]) = (m_{1j}(t, [\lambda, y]), m_{2j}(t, [\lambda, y]))^T \), \( i = 1, 2, j = 1, \ldots, 4 \) in (72). It is clear that the polynomial \( s + 1 \) formed by the coefficients of \( B \) is Hurwitz, \( m_{1j}(t, [\lambda, y]) = 0, m_{2j}(t, [\lambda, y]) \) are defined as

\[ \dot{\nu}_{m,1} = -\dot{\nu}_{m,1} - y, \quad \dot{\nu}_{m,2} = -\dot{\nu}_{m,2} - \tanh(\lambda y), \]

\[ \dot{\nu}_{m,3} = -\dot{\nu}_{m,3} + y, \quad \dot{\nu}_{m,4} = -\dot{\nu}_{m,4} + \tanh(\lambda y), \]
of systems with general nonlinear parametrization. This class can be viewed as a natural extension of the adaptive observer canonical forms [5], [27]. In contrast to earlier approaches addressing the problem of nonlinear parametrization in the problem of adaptive observer design [12], [17], [16], [18], [24], [42], the class of parameterizations for which the reconstruction is guaranteed is not limited to convex/concave or one-to-one functions. We showed that reconstruction can be achieved, subject to additional conditions of linear/nonlinear persistency of excitation, even if nonlinearly parameterized functions in the model are bounded, differentiable and Lipschitz.

In order to ensure convergence of the parameter estimates for this class of models, we have gone beyond the usual requirement that the error dynamics in the model-observer system be stable in the sense of Lyapunov. The set to which the estimates converge is not guaranteed to be Lyapunov-stable. Yet the set is attracting in a weaker, Milnor sense, cf. [34]. The time required for convergence in our approach depends heavily on the dimension of $\lambda$; it depends, however, crucially on the dimension of $\theta$. This renders the method more efficient than exhaustive search; the smaller the dimension of $\lambda$ the more advantageous our method becomes. In this respect a related question arises: is there a “best” parametrization for a given physical model in the class of systems (7) or (9)? The answer is likely to require quantitative assessment of the performance of various observers for all admissible parametrizations. We do not answer this question here, but hope to be able to address it in future.

References

[1] H.D.I. Abarbanel, D. Creveling, R. Farisian, and M. Kostuk. Dynamical state and parameter estimation. SIAM J. Applied Dynamical Systems, 8(4):1341–1381, 2009.
[2] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. Automatica, 44:1753–1765, 2008.
[3] V.I. Arnold. Mathematical Methods in Classical Mechanics. Springer-Verlag, 1978.
[4] G. Bastin and D. Dochain. On-line Estimation and Adaptive Control of Bioreactors. Elsevier, 1990.
[5] G. Bastin and M. Gevers. Stable adaptive observers for nonlinear time-varying systems. IEEE Trans. on Automatic Control, 33(7):650–658, 1988.
[6] G. Besancon. Remarks on nonlinear adaptive observer design. Systems and Control Letters, 41:271–280, 2000.
[7] J.D. Boskovic. Stable adaptive control of a class of first-order nonlinearly parameterized plants. IEEE Trans. on Automatic Control, 40(2):347–350, 1995.
[8] C. Cao, A.M. Annaswamy, and A. Kojic. Parameter convergence in nonlinearly parameterized systems. IEEE Trans. Automatic Control, 48(3):397–411, 2003.
[9] M. Chapell. Structural identifiability of models characterizing saturable binding: Comparison of pseudo-steady-state and non-pseudo-steady state model formulations. Mathematical Biosciences, 133:1–20, 1996.
Proof of Lemma 14. Noticing that $y(t)$ for $t \geq t_0 + T$, $T > 0$, can be expressed as:

$$y(t) = y(t - T) e^{kT} + \int_{t-T}^{t} e^{k(t-\tau)} (u(\tau) + d(\tau)) d\tau$$

and using the Mean-value theorem we obtain:

$$y(t) - y(t - T) e^{kT} = T e^{k(t-T')} (u(\tau') + d(\tau'))$$

for $\tau' \in [t - T, t]$. Hence

$$\varepsilon (1 + e^{kT}) \geq T e^{k(t-T')} |u(\tau) - TB - \Delta_\xi|$$

and

$$\Delta_\xi + TB + \frac{\varepsilon (1 + e^{kT})}{T \min \{1, e^{kT} \}} \geq \Delta_\xi + TB + \frac{\varepsilon (1 + e^{kT})}{T \min \{1, e^{k(t-T')} \}} \geq |u(t)| \forall t \geq t_0 + T.$$
Given that \( T \) can be chosen arbitrarily we let \( T = \sqrt{e} \), and thus \( |u(t)| \leq \sqrt{e(1 + e^{k\sqrt{e}})} \max\{1, e^{-k\sqrt{e}}\} + B\sqrt{e} + \Delta \leq \sqrt{e(1 + e^{k\sqrt{e}} + B)} + \Delta \forall t \geq t_0 \geq \sqrt{e}.

**Proof of Lemma 12.** Let us rewrite (65) as

\[
\dot{y} = a_1 y + \hat{C}x + u_1(t) + d_1(t)
\]

\[
\ddot{x} = \hat{A}x + \hat{a}y + b_{x1} + Gu(t) + d(t),
\]

where \( \hat{a} = \) col\((a_1, \ldots, a_\alpha)\), \( \hat{C} = \) col\((1, 0, \ldots, 0)\), \( \hat{d}(t) = \) col\((d_2(t), \ldots, d_n(t))\), and

\[
G = \begin{pmatrix} -b & I_{n-1} \end{pmatrix}, \quad \hat{A} = \begin{pmatrix} 0 & I_{n-2} \\ 0 & 0 \end{pmatrix}.
\]

Let \( \|u(t)\|_{\infty, [t_0, \infty]} \leq \varepsilon \) and denote \( e(t) = \hat{C}^T \hat{x} + u_1(t) \).

According to Lemma 14, there is a \( t_1(t) > t_0 \) and \( t_1(t) > t_2 \leq \varepsilon (1 + e^{k\sqrt{e}} + B) \). Using the notation above we obtain:

\[
\hat{x} = (\hat{A} - b\hat{C}^T)\hat{x} + \hat{a}y(t) + Gu(t) + be(t) + d(t).
\]

Matrix \( \hat{A} - b\hat{C}^T = \hat{A} \) is Hurwitz, and hence there are \( D, k \in \mathbb{R}_{>0} \) such that \( \|e^{\hat{A}(t-t_0)}\| \leq De^{-k(t-t_0)}. \)

Therefore

\[
\|\hat{C}^T \hat{x}(t) - \hat{C}^T \hat{x}(t_0)\| \leq \|\hat{C}^T \hat{x}(t_0)\| + \frac{D}{k}(\|e\| + \|b\|)(1 + e^{k\sqrt{e}} + B) + \Delta_\varepsilon.
\]

Noticing that \( z_1 = \hat{C}^T \int_{t_0}^t e^{\hat{A}(t-t')} \hat{C}u(t') \, dt' \)

and denoting \( \kappa_\varepsilon(\tau) = 2\frac{D}{k}(\|e\| + \|b\|)(1 + e^{k\sqrt{e}} + B) + \Delta_\varepsilon \), we can conclude that there is a \( t_1(\tau) > t_1(\varepsilon) \) such that

\[
\|z_1(\tau) + u_1(\tau)\|_{\infty, [t_2, \infty]} \leq \kappa_\varepsilon(\tau) + \kappa_\varepsilon(\Delta_\varepsilon) \forall t \geq t_1(\tau),
\]

Noticing that \( y(t) \equiv 0 \Rightarrow e(t) \equiv 0 \) ensures that (67) holds too. \( \square \)

**Proof of Lemma 10.** Consider \( J(\lambda, t) = \int_{t}^{t^*} \alpha(\tau, \lambda) \alpha^T(\tau, \lambda) d\tau \), where \( \alpha \) is an arbitrary non-zero vector from \( \mathbb{R}^{n+m} \). Given that \( \alpha(\lambda, t) \in \lambda UP(T, \mu) \) the following estimate holds:

\[
J(\lambda, t) \geq \mu|z|^2 \quad \forall z \in \Omega_\lambda.
\]

Suppose that \( \hat{A} : \mathbb{R} \rightarrow \Omega_\lambda \) is a differentiable function, and consider \( J(\hat{A}(t), t) - \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau = \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau - \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau = \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau + \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 - \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau = \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau + \int_{t}^{t^*} T^*(\tau, \hat{A}(t)) \alpha^T(\tau, \hat{A}(t)) - \alpha^T(\tau, \hat{A}(t)) \alpha^T(\tau, \hat{A}(t)) d\tau \). Applying the Cauchy-Schwarz inequality to the last equality, and invoking C4 and C3 we obtain:

\[
J(\hat{A}(t), t) - \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau \leq \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau + \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau \leq 2MT \|z\|^2 \leq 2MT \|z\|^2 \max_{\tau \in [t, t+T]} \|\dot{z}(\tau)\|. \]

Thus (40), (41) ensure that \( \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau \geq J(\hat{A}(t), t) - r\|z\|^2 \). The latter inequality, in accordance with C2, guarantees that \( \int_{t}^{t^*} \|z^2\alpha(\tau, \hat{A}(t))\|^2 d\tau \geq (1 - r)\|z\|^2 \). This implies that the function \( \alpha(t, \hat{A}(t)) \) is persistently exciting in the sense of Definition 2. Note that the value of \( (1 - r)\|z\|^2 \) does not depend on the choice of \( \alpha \) as long as (41), (40) hold. Finally, noticing that C4 together with (40) guarantee boundedness of \( \alpha(t, \hat{A}(t)) \) and its derivative wrt \( t \): \( \max \|\alpha(t, \hat{A}(t))\|_{\infty, \|\alpha(t, \hat{A}(t))\|} \leq M + M M_\lambda = M + \frac{\mu}{\text{MT}} \). Taking into account C1 and Theorem 3 we can conclude that the lemma’s statement follows.

\( \square \)

**Proof of Lemma 11.** According to condition (58) we can conclude that that \( h(t_0) \geq 0 \). Let us introduce a strictly decreasing sequence: \( \{\sigma_i\}, i = 0, 1, \ldots, \sigma_i = (1/\nu)^i, \) \( \nu \in (1, \infty) \). Further, let \( \{t_i\}, i = 1, \ldots, t_i < t_2 < \cdots < t_{n} < \cdots \) be an ordered infinite sequence of time instants such that

\[
h(t_i) = \sigma_i h(t_0), \quad i = 0, 1, \ldots.
\]

We wish to show that if (57), (58), and (A.1) then

\[
h(t) \rightarrow 0 \quad t \rightarrow \infty.
\]

In order to do so consider the time differences \( t_i = t_i - t_{i-1}. \) It is clear from (56) that

\[
T_i D_{s_{\lambda}} \max_{\tau \in [t_i, t_{i+1}]} \|x(\tau) + d(\tau)\| \geq h(t_0)(\sigma_{i+1} - \sigma_i).
\]

In addition, \( \max_{\tau \in [t_i, t_{i+1}]} \|x(\tau) + d(\tau)\| = \|x(\tau) + d(\tau)\|_{\infty, \|x(\tau) + d(\tau)\|} \geq \epsilon \) if \( \|x(\tau) + d(\tau)\|_{\infty, \|x(\tau) + d(\tau)\|} > \epsilon \), and \( \max_{\tau \in [t_i, t_{i+1}]} \|x(\tau) + d(\tau)\| = 0 \) otherwise, we can see from (A.3) that

\[
T_i \geq \frac{1}{D_{t_i}} \min_{\tau \in [t_i, t_{i+1}]} \frac{1}{\|x(\tau) + d(\tau)\|_{\infty, \|x(\tau) + d(\tau)\|}} \|x(\tau) + d(\tau)\|_{\infty, \|x(\tau) + d(\tau)\|}^{-\tau} \quad \|x(\tau) + d(\tau)\|_{\infty, \|x(\tau) + d(\tau)\|} \geq \epsilon.
\]

Consider the case when \( \|x(\tau) + d(\tau)\|_{\infty, \|x(\tau) + d(\tau)\|} \leq \epsilon < 0 \) for all \( i = 1, 2, \ldots \). Let us pick

\[
\tau^* = \rho^{-1}(d/\nu), \quad d \in (0, 1),
\]

and select the value of \( D_{t_i} \) such that (58) holds. Given that \( \|x(\tau) + d(\tau)\|_{\infty, [t_0, t_1]} - \epsilon \leq g(0)\|x(t_0)\| + ch(t_0) + \epsilon \)
\[
\Delta + \Delta_d - \varepsilon, \text{ conditions (58), (57), and (A.5) imply } D_i \leq \frac{\kappa^{-1}}{g(0)\|x(t_0) + h(0)\|} \leq \frac{\kappa^{-1}}{g(0)\|x(t_0) + h(0)\|} \leq \frac{\kappa^{-1}}{g(0)\|x(t_0) + h(0)\|} + \Delta
\]

This, as follows from (A.4), guarantees that \( T_1 \geq \tau^* \).

Without loss of generality suppose that there is an \( i \geq 2 \) such that \( T_j \geq \tau^* \) for all \( 1 \leq j \leq i - 1 \). We will now show that the following implication holds:

\[ T_{i-1} \geq \tau^* \Rightarrow T_i \geq \tau^* \]

This will ensure that (A.2) is satisfied and, consequently, that the lemma hold. Consider \( \|x(\tau)\|_{\infty, [t_{i-1}, t_i]} \); (55) and (A.1) imply that:

\[ \|x(\tau)\|_{\infty, [t_{i-1}, t_i]} \leq g(0)\|x(t_{i-1})\| + \sigma_{i-1}h(t_0) + \Delta. \]

Estimating \( \|x(t_{i-1})\| \) from above, according to (55), results in

\[
\|x(\tau)\|_{\infty, [t_{i-1}, t_i]} \leq g(0)\|x(t_{i-1})\| + \sigma_{i-1}h(t_0) + g(0)\Delta + \Delta.
\]

Invoking (55) in order to express an upper bound for \( \|x(t_{i-1})\| \) in terms of \( \|x(t_{i-3})\| \) leads to:

\[
\|x(\tau)\|_{\infty, [t_{i-1}, t_i]} \leq g(0)\|x(t_{i-1})\| + P_i,
\]

where \( P_i = g(0)\sigma_{i-2}h(t_0) + g(0)\Delta + \Delta. \)

Noticing that \( \tau^* \) is chosen in accordance with (A.5) one can therefore obtain:

\[
\sigma_{i-1}^{-1}(\|x(\tau) + d(\tau)\|_{\infty, [t_{i-1}, t_i]} - \varepsilon) \leq g(0)\|x(t_0)\| + \chi(t_0) + \chi(t_0)g(0)\varepsilon_{\delta} \sum_{j=0}^d + k_{\tau}^{-1}\left[\Delta \left(\varepsilon_{\delta} \sum_{j=0}^d + 1\right) + \Delta_d - \varepsilon\right]
\]

Condition (57) implies that \( \Delta \left(\varepsilon_{\delta} \sum_{j=0}^d + 1\right) + \Delta_d - \varepsilon \leq g(0)\|x(t_0)\| + \chi(t_0)\left(1 + \varepsilon_{\delta} \sum_{j=0}^d\right). \)

Substituting the latter estimate into (A.7) and using (58) yields \( T_i \geq \tau^* \) as required. Therefore \( T_i \geq \tau^* \) for all \( i \geq 1 \), and (A.2) holds. Thus the trajectory \( h(t) \) is bounded from above and below, and hence so is the trajectory \( x(t) \).

**Proof of Theorem 13.** The proof is largely identical to that of Theorem 7, and hence we outline just its sketch. Let \( A_0 = A - BC^T A \), \( G_0 = I_0 - BC^T \), and consider the function \( \tilde{\psi}(t, \lambda(t), T_1) = C^T A I_{[t_{i-1}, t_i]} \psi_0(t, \lambda(t), y(t)) \psi_0(t, \lambda(t), y(t)) \). It is clear that for any \( \varepsilon > 0 \) there are \( T_1, t_1 \) sufficiently large and \( \gamma_1 \) sufficiently small such that

\[
\|\tilde{\varphi}(t, \lambda(t), T_1) - \varphi(t, \lambda(t), y(t), \lambda, y)\| < \varepsilon_1 \quad \text{(A.8)}
\]

for all \( \gamma \in (0, \gamma_1) \) and \( t \geq t_1, t \geq t_0, \). Indeed, consider \( \delta(T_1, t) = C^T A I_{[-T_1, t]} e^{\lambda(t_{i-1} - T_1)} G_0 \psi_0(t, \lambda(t), y(t)) dt \), \( \delta_1(T_1, t) = C^T A I_{[-T_1, t]} e^{\lambda(t_{i-1} - T_1)} G_0 \psi_0(t, \lambda(t), y(t)) dt \), \( \delta_2(T_1, t) = \psi_0(t, \lambda(t), y(t)) dt \), \( \Psi(t, \lambda(t), y(t)) \). Pick \( \varepsilon_1 > 0 \), and let \( T_1 \) be so large that \( \delta(T_1, t) < \varepsilon/2 \) for all \( t \geq t_0 + T_1 \). Let \( \gamma_1 \in (0, \gamma_1) \) be so small that \( \delta_1(T_1, t) \leq \varepsilon_1/2 \) for all \( t \geq t_0 + T_1 \) (such a choice is always possible due to that \( \Psi \) is Lipschitz in \( \lambda \)). Noticing that \( \varphi(t, \lambda(t), y(t), \lambda, y) = \delta_1(T_1, t) + \delta_2(T_1, t) + \varphi(t, \lambda(t), T_1) \) we can conclude that (A.8) holds.

Given that \( \varphi(t, \lambda, y(t), \lambda, y) \in \text{LUMP}(T, \mu) \) is a \( \varepsilon_1 \) in (A.8) such that \( \varphi(t, \lambda, T_1) \in \text{LUMP}(T, \mu - \varepsilon) \), \( \varepsilon \in (0, \mu/3) \). On the other hand (see the first part of the proof of Lemma 10), there is a \( \gamma_2 \) such that \( \varphi(t, \lambda(t), T_1) \) is persistently exciting with parameters \( T_1, \mu - \varepsilon \) for all \( \gamma \in (0, \gamma_2) \). Choosing \( \gamma \in (0, \min(\{\gamma_1, \gamma_2\}) \) and taking (A.8) into account we conclude that \( \varphi(t, \lambda(t), y(t), \lambda, y) \) is persistently exciting for \( t \geq t_0 > t_1 > t_0 \) provided that \( \varepsilon_1 \) is small enough and \( t_1 \) is sufficiently large. Thus, invoking the argument presented in Part 2 of the proof of Theorem 7 we can conclude that (62) and (61) hold for the combined system. Convergence of state and parameter estimates can now be shown similarly to the 3-d part of the proof of Theorem 7.