Joint Parameter and State Estimation of Noisy Discrete-Time Nonlinear Systems: A Supervisory Multi-Observer Approach

T.J. Meijer [1,†], V.S. Dolk [2], M.S. Chong [1], R. Postoyan [3], B. de Jager [1], D. Nešić [4] and W.P.M.H. Heemels [1]

2021 60th IEEE Conference on Decision and Control (CDC) | 978-1-6654-3659-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/CDC45484.2021.9683580

***Abstract*— This paper presents two schemes to jointly estimate parameters and states of discrete-time nonlinear systems in the presence of bounded disturbances and noise. The parameters are assumed to belong to a known compact set. Both schemes are based on sampling the parameter space and designing a state observer for each sample. A supervisor selects one of these observers at each time instant to produce the parameter and state estimates. In the first scheme, the parameter and state estimates are guaranteed to converge within a certain margin of their true values in finite time, assuming that a sufficiently large number of observers is used and a persistence of excitation condition is satisfied in addition to other observer design conditions. This convergence margin is constituted by a part that can be chosen arbitrarily small by the user and a part that is determined by the noise levels. The second scheme exploits the convergence properties of the parameter estimate to perform subsequent zoom-ins on the parameter subspace to achieve stricter margins for a given number of observers. The strengths of both schemes are demonstrated using a numerical example.**

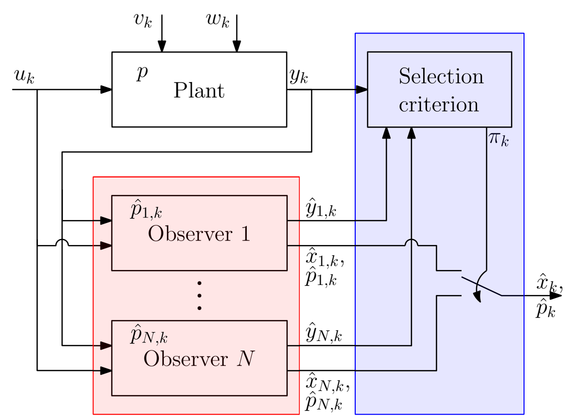

Fig. 1. Supervisory observer scheme consisting of the (□) multi-observer and (□) supervisor. Relevant notation is introduced in Section III.

I. Introduction

JOINT parameter and state estimation is a highly relevant problem in many applications, such as synchronization of digital twins with their physical counterparts, see, e.g., [1], and sensor or source localization (in distributed parameter systems), see, e.g., [2]–[4]. In many cases such combined estimation problems arise, even when the aim is to estimate only the parameters of a system, as a result of the full state either not being measurable and/or measurements being corrupted by noise. A common approach to the joint parameter and state estimation problem is to augment the state with the parameters (and add constant parameter dynamics) and formulate it as a state estimation problem [5]. The state of the resulting system is then estimated using nonlinear state estimation algorithms, such as nonlinear Kalman filters or particle filters [5], however, in general the underlying structure of the original model is lost leading to a (highly) nonlinear state estimation problem. For example, the augmented state approach turns joint estimation of an uncertain linear system with affine parameter dependencies into a bilinear state estimation problem. Following this path, it is typically difficult to provide convergence results [6]. Joint parameter and state estimation schemes that do provide analytical convergence results often apply only to specific classes of systems, see, e.g., [6]–[8].

In this paper, the joint parameter and state estimation problem for discrete-time nonlinear systems in the presence of bounded process and measurement noise is addressed in a different way. We exploit a supervisory observer framework that was recently developed in [9] for *continuous-time* systems and *without the consideration of disturbances and noise*. It is assumed that the parameters are constant and belong to a known compact set, with no restriction on its ”size”. The so-called *supervisory observer* scheme, as depicted in Fig. 1, consists of (a) the *multi-observer*, a bank of multiple state observers–each designed for a parameter value sampled from the known parameter set–and (b) the *supervisor*, which at any given time instant selects one of the observers to provide the state and parameter estimates. Such multi-observer schemes have also been proved useful for many other purposes, such as, safeguarding systems against sensor attacks, see, e.g., [10], and the context of adaptive control, see, e.g., [11]. An advantage of this sampling-based approach compared to the augmented state-space approach is that, for each parameter sample, the structure of the underlying system is preserved. This fact allows us to employ observers tailored to the specific model structure, which come with certain convergence guarantees and convenient (LMI-based) synthesis procedures, see, e.g., [12] for LPV systems or [13] for a class of nonlinear systems. The convergence properties of the individual state observers in the multi-observer are combined with a persistence of excitation (PE) condition to arrive at convergence guarantees for the supervisory observer. To be more concrete, the parameter and state estimates are guaranteed to converge within a certain margin of their true values, given that a sufficiently large number of observers is used. This sampling-based approach,

[1] Tomas Meijer, Michelle Chong, Bram de Jager and Maurice Heemels are with the Department of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. {t.j.meijer; m.s.t.chong; a.g.de.jager; m.heemels}@tue.nl
[2] Victor Dolk is with ASML, De Run 6665, 5504 DT Veldhoven, The Netherlands. victor.dolk@asml.com
[3] Romain Postoyan is with the Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France. romain.postoyan@univ-lorraine.fr
[4] Dragan Nešić is with the Department of Electrical and Electronic Engineering, the University of Melbourne, Parkville, VIC 3010, Australia. dnesic@unimelb.edu.au
† Corresponding author: T.J. Meijer.

which uses a static sampling policy, is rather simple to implement, but the number of samples (and, hence, the number of observers running in parallel) required to guarantee that the parameter error converges to within a given margin grows exponentially with the dimension of the parameter space. This inspired the development of a second scheme, which exploits the convergence result to iteratively zoom in by resampling from a shrinking subspace of the original parameter space. The resulting dynamic sampling policy is able to, for a given number of observers, guarantee tighter bounds on the parameter and state estimates. Alternatively, the dynamic scheme can be used to achieve a given margin of convergence using fewer observers than the static scheme.

The extension of the continuous-time results in [9] to discrete-time is motivated by the fact that real-time implementation of any estimation algorithm requires discretization and that measurements become available at discrete time instances. Additionally, the discrete-time formulation enables parameter and state estimation of systems in feedback interconnection with a discrete-time control architecture such as model predictive control. The inclusion of process and measurement noise in the supervisory multi-observer framework is another major contribution, which allows us to provide more realistic performance guarantees for the proposed estimator, that was not addressed in [9], [14]. However, it poses additional technical challenges including distinguishing between the effects of noise and parameter errors on our state and output estimation errors. In fact, this is only possible to some extent and, unlike in the noiseless case, the parameter error cannot be made arbitrarily small by using sufficiently many observers. Moreover, the dynamic sampling policy has to take into account the noise levels when zooming in, requiring a careful analysis. The strength of our framework is demonstrated on a numerical case study in the presence of noise.

The content of the paper is organized as follows. The problem definition is given in Section II. Section III presents the discrete-time supervisory observer using a static sampling policy. In Section IV, the supervisory observer is adapted to utilize dynamic sampling. Finally, a numerical case study and conclusions are given in Sections V and VI. All proofs can be found in the full version of this paper [15].

Notation. Let $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_{\geqslant 0} = [0, \infty)$, $\mathbb{R}_{>0} = (0, \infty)$, $\mathbb{N} = \{0, 1, \ldots\}$, $\mathbb{N}_{[n,m]} = \{n, n+1, \ldots, m\} \subset \mathbb{N}$ for $n, m \in \mathbb{N}$ and $\mathbb{N}_{\geqslant n} = \{n, n+1, \ldots\} \subseteq \mathbb{N}$ for $n \in \mathbb{N}$. Moreover, $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ with $n \in \mathbb{N}_{\geqslant 1}$ denotes an arbitrary (but the same throughout the paper) p-norm on $\mathbb{R}^n$, and we omit the subscript p in the following, i.e., $\|\cdot\| = \|\cdot\|_p$. Let $\mathbb{B}^n(\xi, r) := \{x \in \mathbb{R}^n | \|x - \xi\| \leqslant r\}$ for $n \in \mathbb{N}_{\geqslant 1}$ represent the ball centered at $\xi \in \mathbb{R}^n$ of "radius" $r \in \mathbb{R}_{\geqslant 0}$ and let $\mathbb{B}^n_r = \mathbb{B}^n(0, r)$ denote such a set centered at the origin. For a sequence $\{x_k\}_{k\in\mathbb{N}}$ with $x_k \in \mathbb{R}^n$ and $n \in \mathbb{N}$, we denote $\|\{x_k\}\| = \left\|\{x_k\}_{k\in\mathbb{N}}\right\|_p := \sup_{k\in\mathbb{N}} \|x_k\|_p$ where the subscript p is again omitted for the sake of compactness. The space of all bounded sequences taking values in $\mathbb{R}^n$ with $n \in \mathbb{N}$ is denoted $\ell^\infty := \{\{x_k\}_{k\in\mathbb{N}} \mid \|\{x_k\}\|_\infty < \infty\}$. The notation (u, v) stands for $[u^\top \, v^\top]^\top$, where $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ with $(n, m) \in \mathbb{N}^2$. A continuous function $\alpha : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is a $\mathcal{K}$-function ($\alpha \in \mathcal{K}$) if it is strictly increasing and $\alpha(0) = 0$. If, in addition, $\alpha(r) \to \infty$ as $r \to \infty$, then α is a $\mathcal{K}_\infty$-function ($\alpha \in \mathcal{K}_\infty$). A continuous function $\beta : \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is a $\mathcal{KL}$-function ($\beta \in \mathcal{KL}$) if $\beta(\cdot, s) \in \mathcal{K}$ for each $s \in \mathbb{R}_{\geqslant 0}$, $\beta(r, \cdot)$ is non-increasing and $\beta(r, s) \to 0$ as $s \to \infty$ for each $r \in \mathbb{R}_{\geqslant 0}$.

II. Problem definition

Consider the discrete-time system given by

$$x_{k+1} = f(x_k, p, u_k, v_k), \tag{1a}$$

$$y_k = h(x_k, p, u_k, w_k), \tag{1b}$$

where $x_k \in \mathbb{R}^{n_x}$, $u_k \in \mathbb{R}^{n_u}$ and $y_k \in \mathbb{R}^{n_y}$ denote the state, input and output, respectively, at time instant $k \in \mathbb{N}$. In addition, the following assumptions are adopted.

Assumption 1. *The input u_k, process noise v_k and measurement noise w_k in* (1) *are bounded, i.e., there exist constants $\Delta_u, \Delta_v, \Delta_w \in \mathbb{R}_{\geqslant 0}$ such that for all $k \in \mathbb{N}$*

$$u_k \in \mathbb{B}^{n_u}_{\Delta_u}, \quad v_k \in \mathbb{B}^{n_v}_{\Delta_v} \quad \text{and} \quad w_k \in \mathbb{B}^{n_w}_{\Delta_w}. \tag{2}$$

Assumption 2. *The parameter vector p is constant and unknown and it belongs to a given compact set $\mathbb{P}$, i.e., $p \in \mathbb{P} \subset \mathbb{R}^{n_p}$.*

Assumption 1 means that $\{u_k\}_{k\in\mathbb{N}}, \{v_k\}_{k\in\mathbb{N}}, \{w_k\}_{k\in\mathbb{N}} \in \ell^\infty$, which is a reasonable assumption in practice. It should be noted that Δ_v and Δ_w in Assumption 1 do not need to be known to implement the estimation schemes, their existence alone is sufficient. The input u_k and output y_k are known/measured, while the full state x_k, process noise v_k and measurement noise w_k are unknown. Moreover, the functions $f : \mathbb{R}^{n_x} \times \mathbb{P} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_x}$ and $h : \mathbb{R}^{n_x} \times \mathbb{P} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_w} \to \mathbb{R}^{n_y}$ are given and h is locally Lipschitz continuous. For any initial condition $x_0 \in \mathbb{R}^{n_x}$, input sequence $\{u_k\}_{k\in\mathbb{N}}$ with $u_k \in \mathbb{R}^{n_u}$, process noise sequence $\{v_k\}_{k\in\mathbb{N}}$ with $v_k \in \mathbb{R}^{n_v}$, measurement noise sequence $\{w_k\}_{k\in\mathbb{N}}$ with $w_k \in \mathbb{R}^{n_w}$ for $k \in \mathbb{N}$ and parameters $p \in \mathbb{P}$, the system (1) admits a unique solution defined for all $k \in \mathbb{N}$. Finally, the following assumption is adopted.

Assumption 3. *The solutions to* (1) *are uniformly bounded, i.e., for all $\Delta_x, \Delta_u, \Delta_v, \Delta_w \in \mathbb{R}_{\geqslant 0}$, there exists a constant $K_x = K_x(\Delta_x, \Delta_u, \Delta_v, \Delta_w) \in \mathbb{R}_{>0}$ such that for all $x_0 \in \mathbb{B}^{n_x}_{\Delta_x}$, $\{u_k\}_{k\in\mathbb{N}}$ with $u_k \in \mathbb{B}^{n_u}_{\Delta_u}$, $\{v_k\}_{k\in\mathbb{N}}$ with $v_k \in \mathbb{B}^{n_v}_{\Delta_v}$ and $\{w_k\}_{k\in\mathbb{N}}$ with $w_k \in \mathbb{B}^{n_w}_{\Delta_w}$ for any $k \in \mathbb{N}$, it holds that $x_k \in \mathbb{B}^{n_x}_{K_x}$ for all $k \in \mathbb{N}$.*

The bound K_x in Assumption 3 does not need to be known to implement the proposed estimation algorithms, only its existence has to be ensured.

Our objective is to jointly estimate the parameter vector p and the state of the system (1) (within certain margins) subject to bounded process noise v_k and measurement noise w_k, given the input u_k and the measured output y_k.

III. Supervisory observer: static sampling policy

The parameter and state estimation schemes presented in this paper consist of two subsystems, as shown in Fig. 1. The first subsystem is the so-called multi-observer, which is a collection of observers that operate in parallel, where each observer is designed for a different parameter vector sampled from the parameter space. The second subsystem is a supervisor. The outputs of the observers are fed to the supervisor, which selects one of the observers based on a selection criterion and outputs its state estimate and corresponding parameter sample as the estimates produced by the overall estimation scheme. In this section, the parameter samples are obtained using a *static* sampling policy meaning that these samples are fixed for all times. Later, in Section IV, we consider a *dynamic* sampling policy, which aims to reduce the computational complexity of the estimation scheme.

A. Multi-observer

The parameter space $\mathbb{P}$ is sampled to produce N parameter samples $\hat{p}_i \in \mathbb{P}$ for $i \in \mathcal{N} := \mathbb{N}_{[1,N]}$. This sampling is performed in such a way that the maximum distance of the true parameter to the nearest sample tends to zero as N tends to infinity, i.e.,

$$\lim_{N\to\infty} \max_{p\in\mathbb{P}} \min_{i\in\mathcal{N}} \|\hat{p}_i - p\| = 0. \tag{3}$$

This can be ensured, for instance, by employing a uniform sampling of the parameter space. For each $\hat{p}_i$, $i \in \mathcal{N}$, a state observer is designed, given by

$$\hat{x}_{i,k+1} = \hat{f}\left(\hat{x}_{i,k}, \hat{p}_i, u_k, y_k\right), \tag{4a}$$

$$\hat{y}_{i,k} = h\left(\hat{x}_{i,k}, \hat{p}_i, u_k, 0\right), \tag{4b}$$

where $\hat{x}_{i,k} \in \mathbb{R}^{n_x}$ and $\hat{y}_{i,k} \in \mathbb{R}^{n_y}$ denote, respectively, the state and output estimate of the i-th observer at time $k \in \mathbb{N}$. The function $\hat{f} : \mathbb{R}^{n_x} \times \{\hat{p}_i\}_{i\in\mathcal{N}} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}^{n_x}$ is well-designed such that the solutions to (4) are defined for all time $k \in \mathbb{N}$, any initial condition $\hat{x}_{i,0} \in \mathbb{R}^{n_x}$, input sequence $\{u_k\}_{k\in\mathbb{N}}$, output sequence $\{y_k\}_{k\in\mathbb{N}}$ and parameter sample $\hat{p}_i \in \mathbb{P}$, $i \in \mathcal{N}$.

Let $\tilde{x}_{i,k} := \hat{x}_{i,k} - x_k$ denote the state estimation error, $\tilde{y}_{i,k} := \hat{y}_{i,k} - y_k$ the output estimation error and $\tilde{p}_i := \hat{p}_i - p$ the parameter estimation error of the i-th observer. Since $\mathbb{P}$ is compact, there exists a compact set $\mathbb{D} \subset \mathbb{R}^{n_p}$ such that $\tilde{p}_i \in \mathbb{D}$ for any $p, \hat{p}_i \in \mathbb{P}$. The state and output estimation errors are governed by

$$\tilde{x}_{i,k+1} = F\left(\tilde{x}_{i,k}, x_k, \tilde{p}_i, p, u_k, v_k, w_k\right), \tag{5a}$$

$$\tilde{y}_{i,k} = H\left(\tilde{x}_{i,k}, x_k, \tilde{p}_i, p, u_k, w_k\right), \tag{5b}$$

where the functions $F : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{D} \times \mathbb{P} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_w} \to \mathbb{R}^{n_x}$ and $H : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{D} \times \mathbb{P} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_w} \to \mathbb{R}^{n_y}$ are given by $F\left(\tilde{x}, x, \tilde{p}, p, u, v, w\right) = \hat{f}\left(\tilde{x} + x, \tilde{p} + p, u, h\left(x, p, u, w\right)\right) - f\left(x, p, u, v\right)$ and $H\left(\tilde{x}, x, \tilde{p}, p, u, w\right) = h\left(\tilde{x} + x, \tilde{p} + p, u, 0\right) - h\left(x, p, u, w\right)$. The observers (4) are assumed to be robust with respect to the parameter error and noise in the following sense.

Assumption 4. *There exist functions $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{K}_\infty$ and a continuous non-negative function $\sigma : \mathbb{D} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_w} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}_{\geqslant 0}$ with $\sigma\left(0, 0, 0, x, u\right) = 0$ for all $x \in \mathbb{R}^{n_x}$ and $u \in \mathbb{R}^{n_u}$ such that there exists a function $V : \mathbb{P} \times \mathbb{R}^{n_x} \to \mathbb{R}_{\geqslant 0}$, which satisfies, for all $\tilde{x}, x \in \mathbb{R}^{n_x}$, $p, \hat{p} \in \mathbb{P}$, $u \in \mathbb{R}^{n_u}$, $v \in \mathbb{R}^{n_v}$ and $w \in \mathbb{R}^{n_w}$, that $V(\hat{p}, \cdot)$ is continuous and*

$$\alpha_1\left(\|\tilde{x}\|\right) \leqslant V\left(\hat{p}, \tilde{x}\right) \leqslant \alpha_2\left(\|\tilde{x}\|\right), \tag{6a}$$

$$V\left(\hat{p}, \tilde{x}^+\right) \leqslant V\left(\hat{p}, \tilde{x}\right) - \alpha_3\left(\|\tilde{x}\|\right) + \sigma\left(\tilde{p}, v, w, x, u\right), \tag{6b}$$

for $\tilde{x}^+ = F\left(\tilde{x}, x, \tilde{p}, p, u, v, w\right)$.

Assumption 4 implies that the error systems (5) corresponding to the observers in (4) are locally input-to-state stable (ISS) with respect to $\tilde{p}_i$, v_k and w_k [16]. For linear uncertain systems, Luenberger observers satisfy Assumption 4 and, in Section V, it is shown that a class of circle-criterion-based nonlinear observers also satisfies this assumption.

B. Supervisor

At every time $k \in \mathbb{N}$, the supervisor selects one observer from the multi-observer. To be able to assess the accuracy of the different observers, the supervisor computes a monitoring signal for each observer, which, for $i \in \mathcal{N}$, is given by

$$\mu_{i,k} = \sum_{j=0}^{k-1} \lambda^{k-1-j} \left\|\tilde{y}_{i,j}\right\|^2, \quad k \in \mathbb{N}, \tag{7}$$

where $\lambda \in [0, 1)$ is a design parameter. The i-th monitoring signal (7) can be implemented using the difference equation

$$\mu_{i,k+1} = \lambda\mu_{i,k} + \left\|\tilde{y}_{i,k}\right\|^2, \quad k \in \mathbb{N}, \tag{8}$$

with the initial condition $\mu_{i,0} = 0$. The output errors of the state observers are assumed to satisfy the following PE condition.

Assumption 5. *For any $\Delta_{\tilde{x}}, \Delta_x, \Delta_u, \Delta_v, \Delta_w \geqslant 0$, there exist a function $\alpha_{\tilde{y}} \in \mathcal{K}_\infty$ and an integer $N_{\mathrm{pe}} = N_{\mathrm{pe}}\left(\Delta_{\tilde{x}}, \Delta_x, \Delta_u, \Delta_v, \Delta_w\right) \in \mathbb{N}_{\geqslant 1}$ such that for all $\tilde{p}_i \in \mathbb{D}$, $i \in \mathcal{N}$, $\tilde{x}_{i,0} \in \mathbb{B}^{n_x}_{\Delta_{\tilde{x}}}$, $p \in \mathbb{P}$, $x_0 \in \mathbb{B}^{n_x}_{\Delta_x}$, $\{v_k\}_{k\in\mathbb{N}}$ with $v_k \in \mathbb{B}^{n_v}_{\Delta_v}$ and $\{w_k\}_{k\in\mathbb{N}}$ with $w_k \in \mathbb{B}^{n_w}_{\Delta_w}$ and for some input sequence $\{u_k\}_{k\in\mathbb{N}}$ with $u_k \in \mathbb{B}^{n_u}_{\Delta_u}$ for $k \in \mathbb{N}$, the solutions to* (1) *and* (5) *satisfy*

$$\sum_{j=k-N_{\mathrm{pe}}}^{k-1} \left\|\tilde{y}_{i,j}\right\|^2 \geqslant \alpha_{\tilde{y}}\left(\|\tilde{p}_i\|\right), \quad k \in \mathbb{N}_{\geqslant N_{\mathrm{pe}}}. \tag{9}$$

Assumption 5 differs from the classical PE condition, see, e.g., [17], in that it considers solutions to (5b) parametrized by $\tilde{p}_i$ and requires the sum in (9) to grow with the norm of the parameter error. This ensures that the supervisor is able to infer quantitative information about the parameter estimation error of each state observer based on its monitoring signal.

At every time instant $k \in \mathbb{N}$, the supervisor selects (one of) the observer(s) with the smallest monitoring signal to obtain the estimates of p and x_k. In the event that $\min_{i\in\mathcal{N}} \mu_{i,k}$ is not unique any observer from this subset can be chosen, resulting in a selection criterion where the index of the selected observer $\pi_k : \mathbb{N} \to \mathcal{N}$ satisfies

$$\pi_k \in \arg\min_{i\in\mathcal{N}} \mu_{i,k}, \quad k \in \mathbb{N}. \tag{10}$$

The resulting parameter estimate, state estimate and state estimation error at time $k \in \mathbb{N}$, denoted $\hat{p}_k$, $\hat{x}_k$ and $\tilde{x}_k$, respectively, are defined using π_k as

$$\hat{p}_k := \hat{p}_{\pi_k}, \quad \hat{x}_k := \hat{x}_{\pi_k,k} \quad \text{and} \quad \tilde{x}_k := \tilde{x}_{\pi_k,k}. \tag{11}$$

C. Convergence guarantees

The parameter and state estimates (11) converge to within certain margins of their true values p and x_k as stated in the following theorem.

Theorem 1. *Consider the system* (1)*, the multi-observer* (4)*, the monitoring signals* (7)*, the selection criterion* (10)*, the parameter estimate, state estimate and state estimation error in* (11)*. Suppose Assumptions 1-5 hold. For any*

$\Delta_{\tilde{x}}, \Delta_x, \Delta_u, \Delta_v, \Delta_w \in \mathbb{R}_{\geqslant 0}$ *and any margins* $\nu_{\tilde{p}}, \nu_{\tilde{x}} \in \mathbb{R}_{>0}$, *there exist functions* $\upsilon_{\tilde{p}}, \upsilon_{\tilde{x}}, \omega_{\tilde{p}}, \omega_{\tilde{x}} \in \mathcal{K}_\infty$, *constant* $K_{\tilde{x}} \in \mathbb{R}_{>0}$ *and sufficiently large integers* $N^\star \in \mathbb{N}_{\geqslant 1}$ *and* $M \in \mathbb{N}_{\geqslant 1}$ *such that for any* $N \in \mathbb{N}_{\geqslant N^\star}$, *it holds for any* $\tilde{x}_{i,0} \in \mathbb{B}^{n_x}_{\Delta_{\tilde{x}}}$, $i \in \mathcal{N}$, $p \in \mathbb{P}$, $x_0 \in \mathbb{B}^{n_x}_{\Delta_x}$, $\{v_k\}_{k\in\mathbb{N}}$ *with* $v_k \in \mathbb{B}^{n_v}_{\Delta_v}$, $\{w_k\}_{k\in\mathbb{N}}$ *with* $w_k \in \mathbb{B}^{n_w}_{\Delta_w}$ *and for some input sequence* $\{u_k\}_{k\in\mathbb{N}}$ *with* $u_k \in \mathbb{B}^{n_u}_{\Delta_u}$ *for* $k \in \mathbb{N}$, *which satisfies Assumption 5, that* $\tilde{x}_{i,k} \in \mathbb{B}^{n_x}_{K_{\tilde{x}}}$ *for all* $k \in \mathbb{N}$, $i \in \mathcal{N}$, *and*

$$\|\tilde{p}_{\pi_k}\| \leqslant \nu_{\tilde{p}} + \upsilon_{\tilde{p}}(\|\{v_j\}\|) + \omega_{\tilde{p}}(\|\{w_j\}\|), \quad \forall k \in \mathbb{N}_{\geqslant M},$$
$$\limsup_{k\to\infty} \|\tilde{x}_k\| \leqslant \nu_{\tilde{x}} + \upsilon_{\tilde{x}}(\|\{v_j\}\|) + \omega_{\tilde{x}}(\|\{w_j\}\|). \quad (12)$$

In the noiseless case, i.e., $v_k = w_k = 0$, the convergence margins can be made arbitrarily small since $\nu_{\tilde{p}}$ and $\nu_{\tilde{x}}$ can be made arbitrarily small by using sufficiently many observers. However, this is impossible in the presence of noise due to the terms in (12) depending on $\|\{v_k\}\|$ and $\|\{w_k\}\|$.

IV. Supervisory observer: dynamic sampling policy

In this section we develop a dynamic sampling policy for joint parameter and state estimation of (1). As stated in Theorem 1, when using a sufficiently large number of observers N, the parameter estimate converges to a given margin within a finite time. We exploit this result in the dynamic sampling policy to iteratively zoom in on the parameter subspace defined by the aforementioned margins through resampling. As a result, stricter bounds on the parameter and state estimates can be guaranteed compared to the static sampling policy for a given number of observers.

A. Dynamic sampling policy

Since the parameter set $\mathbb{P}$ is compact, there exist $p_c \in \mathbb{R}^{n_p}$ and $\Delta_0 \in \mathbb{R}_{>0}$ such that

$$\mathbb{P} \subseteq \mathbb{B}^{n_p}(p_c, \Delta_0). \quad (13)$$

Let $\nu \in \mathbb{R}_{>0}$ denote the desired bound on the parameter error, which either represents the required bound on the parameter error to guarantee asymptotic convergence of the state estimation error to within a desired margin or a desired bound imposed directly on the parameter estimation error. We also introduce a design parameter $\alpha \in (0,1)$, the so-called zooming factor, which determines the rate at which the considered parameter set shrinks. The dynamic sampling policy is initialized at $k = k_0 = 0$ by sampling $\mathbb{P}_0 := \mathbb{P}$, using a sampling scheme, which satisfies (3), to obtain $N \in \mathbb{N}_{\geqslant 1}$ parameter samples $\hat{p}_{i,0}$, $i \in \mathcal{N}$. Here, N is chosen sufficiently large such that, by Theorem 1, it holds for sufficiently large $M \in \mathbb{N}_{\geqslant 1}$ that $\|\hat{p}_{\pi_k,0} - p\| \leqslant C$ for all $k \in \mathbb{N}_{\geqslant M}$ with

$$C := \max\{\nu, \alpha\Delta_0\} + \upsilon_{\tilde{p}}(\|\{v_j\}\|) + \omega_{\tilde{p}}(\|\{w_j\}\|). \quad (14)$$

As a consequence, for $k \in \mathbb{N}_{\geqslant M}$, either the desired margin is achieved or $p \in \mathbb{P}_1 := \mathbb{B}^{n_p}(\hat{p}_{\pi_k,0}, \alpha\Delta_0 + \upsilon_{\tilde{p}}(\|\{v_j\}\|) + \omega_{\tilde{p}}(\|\{w_j\}\|)) \cap \mathbb{P}_0$. Both cases cannot be distinguished online since the true parameter is unknown. Therefore, at $k = k_1$ with $k_1 \in \mathbb{N}_{\geqslant M}$, even if the desired margin has already been achieved, the set $\mathbb{P}_1$ is sampled to obtain N new samples $\{\hat{p}_{i,1}\}_{i\in\mathcal{N}}$. This procedure is performed iteratively at every k_m, $m \in \mathbb{N}$, with

$$M_d := k_{m+1} - k_m, \quad (15)$$

where $M_d \in \mathbb{N}_{\geqslant \max\{1,M\}}$ denotes the number of time steps between subsequent zoom-ins. The shrinking parameter subset $\mathbb{P}_m$, $m \in \mathbb{N}$, is defined recursively by

$$\mathbb{P}_{m+1} := \mathbb{D}_m \cap \mathbb{P}_m \quad (16)$$

with $\mathbb{P}_0 = \mathbb{P}$, $\Delta_m := \alpha\Delta_{m-1} = \alpha^m\Delta_0$ and $\mathbb{D}_m = \mathbb{B}^{n_p}\left(\hat{p}_{\pi_{k_{m+1}},m}, \Delta_{m+1} + \upsilon_{\tilde{p}}(\|\{v_j\}\|) + \omega_{\tilde{p}}(\|\{w_j\}\|)\right)$. The spaces $\mathbb{P}_m$, $m \in \mathbb{N}$, are sampled in such a way that

$$\max_{p\in\mathbb{P}_m} \min_{i\in\mathcal{N}} \|\hat{p}_{i,m} - p\| \leqslant \rho(\Delta_m, N) \quad (17)$$

with $\rho \in \mathcal{KL}$ and where $\{\hat{p}_{i,m}\}_{i\in\mathcal{N}}$ denote the obtained samples. The corresponding parameter errors are denoted $\tilde{p}_{i,m} = \hat{p}_{i,m} - p$, $i \in \mathcal{N}$ and $m \in \mathbb{N}$. It is worth mentioning that once the desired margin is achieved the algorithm still keeps zooming in and it can occur that, after zooming in a certain number of times, the subset that is being sampled no longer contains the true parameter. Regardless, the true parameter still lies within the desired margin of the selected parameter estimate and the convergence guarantees provided in this section remain valid.

The dynamic sampling policy is incorporated into the multi-observer by designing state observers for each parameter sample $\hat{p}_{i,m}$, $i \in \mathcal{N}$, for the time instance $k \in \mathbb{N}_{[k_m,k_{m+1}-1]}$, $m \in \mathbb{N}$. The i-th state observer is given by

$$\hat{x}_{i,k+1} = \hat{f}(\hat{x}_{i,k}, \hat{p}_{i,m}, u_k, y_k), \quad (18a)$$
$$\hat{y}_{i,k} = h(\hat{x}_{i,k}, \hat{p}_{i,m}, u_k, 0), \quad (18b)$$

for $k \in \mathbb{N}_{[k_m,k_{m+1}-1]}$, $m \in \mathbb{N}$. Here, $\hat{f} : \mathbb{R}^{n_x} \times \{\hat{p}_{i,m}\}_{i\in\mathcal{N},m\in\mathbb{N}} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}^{n_x}$ is well-designed such that the solutions to (18) are defined for all $k \in \mathbb{N}$ and any initial condition $\hat{x}_{i,0} \in \mathbb{R}^{n_x}$, input sequence $\{u_k\}_{k\in\mathbb{N}}$, output sequence $\{y_k\}_{k\in\mathbb{N}}$ and parameters $\hat{p}_{i,m} \in \mathbb{P}$, $i \in \mathcal{N}$ and $m \in \mathbb{N}$. We assume these observers satisfy Assumption 4.

The dynamic sampling policy requires the monitoring signals used by the supervisor to be redefined. The redefined monitoring signals are reset upon resampling, i.e.,

$$\mu_{i,k+1} = \begin{cases} \|\tilde{y}_{i,k}\|^2, & k \in \{k_m\}_{m\in\mathbb{N}}, \\ \lambda\mu_{i,k} + \|\tilde{y}_{i,k}\|^2, & k \in \mathbb{N} \setminus \{k_m\}_{m\in\mathbb{N}}, \end{cases} \quad (19)$$

for $i \in \mathcal{N}$, with $\lambda \in [0,1)$. As before, the supervisor selects an observer from the multi-observer (18) using the signal π_k as defined in (10). The definition of the state estimate $\hat{x}_k$ and corresponding error $\tilde{x}_k$ in (11) are unchanged, however, the parameter estimate and corresponding error are redefined as

$$\hat{p}_k := \hat{p}_{\pi_k,m} \text{ and } \tilde{p}_k := \tilde{p}_{\pi_k,m}, \text{ for } k \in \mathbb{N}_{[k_m+1,k_{m+1}]}. \quad (20)$$

B. Convergence guarantees

The parameter and state estimates produced by the supervisory observer using a dynamic sampling scheme satisfy similar convergence guarantees as in the static sampling case. This is stated in the following theorem.

Theorem 2. *Consider the system* (1)*, the multi-observer* (18)*, the monitoring signals* (19)*, the selection criterion* (10)*, the parameter estimate and corresponding error* (20)*, the state estimate and corresponding error in* (11) *and the dynamic sampling policy* (15)-(16)*. Suppose Assumptions 1-5 hold. For any* $\Delta_{\tilde{x}}, \Delta_x, \Delta_u, \Delta_v, \Delta_w \in \mathbb{R}_{\geqslant 0}$, *any margins* $\nu_{\tilde{p}}, \nu_{\tilde{x}} \in \mathbb{R}_{>0}$ *and zooming factor* $\alpha \in (0,1)$, *there exist functions $\upsilon_{\tilde{p}}, \upsilon_{\tilde{x}}, \omega_{\tilde{p}}, \omega_{\tilde{x}} \in \mathcal{K}_\infty$, scalar $K_{\tilde{x}} \in \mathbb{R}_{>0}$ and sufficiently large integers $M^\star \in \mathbb{N}_{\geqslant 1}$, $\bar{M} \in \mathbb{N}_{\geqslant 1}$ and $N^\star \in \mathbb{N}_{\geqslant 1}$ such that for any $N \in \mathbb{N}_{\geqslant N^\star}$ and $M_d \in \mathbb{N}_{\geqslant M^\star}$, it holds for any $\tilde{x}_{i,0} \in \mathbb{B}^{n_x}_{\Delta_{\tilde{x}}}$, $i \in \mathcal{N}$, $p \in \mathbb{P}$, $x_0 \in \mathbb{B}^{n_x}_{\Delta_x}$, $\{v_k\}_{k\in\mathbb{N}}$ with $v_k \in \mathbb{B}^{n_v}_{\Delta_v}$, $\{w_k\}_{k\in\mathbb{N}}$ with $w_k \in \mathbb{B}^{n_w}_{\Delta_w}$ and for some input sequence $\{u_k\}_{k\in\mathbb{N}}$ with $u_k \in \mathbb{B}^{n_u}_{\Delta_u}$ for $k \in \mathbb{N}$, which satisfies Assumption 5, that $\tilde{x}_{i,k} \in \mathbb{B}^{n_x}_{K_{\tilde{x}}}$ for all $k \in \mathbb{N}$, $i \in \mathcal{N}$, and*

$$\begin{aligned} &\|\tilde{p}_k\| \leqslant \nu_{\tilde{p}} + \upsilon_{\tilde{p}}\left(\|\{v_j\}\|\right) + \omega_{\tilde{p}}\left(\|\{w_j\}\|\right), \quad \forall k \in \mathbb{N}_{\geqslant \bar{M}}, \\ &\limsup_{k\to\infty} \|\tilde{x}_k\| \leqslant \nu_{\tilde{x}} + \upsilon_{\tilde{x}}\left(\|\{v_j\}\|\right) + \omega_{\tilde{x}}\left(\|\{w_j\}\|\right). \end{aligned} \tag{21}$$

Theorem 2 ensures the same guarantees as in Theorem 1, but it typically requires less observers to do so using the dynamic sampling policy, as will be illustrated in Section V.

Remark 1. *The dynamic sampling policy in this paper uses a fixed number of samples, however, an alternative policy using the DIRECT algorithm which adds samples on-line is proposed for the continuous-time setting in [14]. This eliminates the need to estimate the required number of observers a-priori, which can be challenging.*

V. Case study

In this section, we apply the results of Theorems 1 and 2 to estimate the parameters and states of an example within the class of nonlinear systems given by

$$\begin{aligned} x_{k+1} &= A\left(p\right)x_k + G\left(p\right)\phi\left(Hx_k\right) + B\left(p\right)\left(u_k + v_k\right), \\ y_k &= Cx_k + w_k, \end{aligned} \tag{22}$$

where $x_k \in \mathbb{R}^{n_x}$, $u_k \in \mathbb{R}^{n_u}$, $v_k \in \mathbb{R}^{n_v}$, $w_k \in \mathbb{R}^{n_w}$ and $p \in \mathbb{R}^{n_p}$. Suppose Assumptions 1-3 hold and $A(p)$, $B(p)$ and $G(p)$ are continuous in p on $\mathbb{P}$. The nonlinearity $\phi : \mathbb{R}^{n_\phi} \to \mathbb{R}^{n_\phi}$ is such that $\phi(v) = (\phi_1(v_1), \dots, \phi_{n_\phi}(v_{n_\phi}))$ for $v = (v_1, \dots, v_{n_\phi}) \in \mathbb{R}^{n_\phi}$ and there exist constants $\ell_i \in \mathbb{R}_{>0}$, $i \in \mathbb{N}_{[1,n_\phi]}$, such that, for all $v \in \mathbb{R}$, we have

$$0 \leqslant \frac{\partial \phi_i(v)}{\partial v} \leqslant \ell_i. \tag{23}$$

For $\hat{p} \in \mathbb{P}$, a state observer of the form [13], [18]

$$\begin{aligned} \hat{x}_{k+1} = A(\hat{p})\hat{x}_k + G(\hat{p})\phi(H\hat{x}_k + K(\hat{p})(C\hat{x}_k - y_k)) \\ + B(\hat{p})u_k + L(\hat{p})(C\hat{x}_k - y_k), \end{aligned} \tag{24a}$$

$$\hat{y}_k = C\hat{x}_k, \tag{24b}$$

is designed by synthesizing observer matrices $K(\hat{p})$ and $L(\hat{p})$ such that the following proposition applies.

Proposition 1. *Consider the system* (22) *and state observer* (24)*. Suppose there exist $P = P^\top \in \mathbb{R}^{n_x \times n_x}$, $M = \operatorname{diag}(m_1, \dots, m_{n_\phi})$ with $m_i \in \mathbb{R}_{>0}$, $i \in \mathbb{N}_{[1,n_\phi]}$, and $\kappa_{\tilde{x}}, \kappa_v, \kappa_w \in \mathbb{R}_{>0}$, such that $P \succ 0$ and, for all $p, \hat{p} \in \mathbb{P}$,*

$$\begin{bmatrix} -P & \star & \star & \star & \star \\ -\boldsymbol{A}^\top(\hat{p})P & \frac{\kappa_{\tilde{x}}}{2}I - \frac{1}{2}P & \star & \star & \star \\ -G^\top(\hat{p})P & \frac{1}{2}M\boldsymbol{H}(\hat{p}) & -M\Lambda^{-1} & \star & \star \\ B^\top(p)P & 0 & 0 & -\frac{\kappa_v}{2}I & \star \\ L^\top(\hat{p})P & 0 & -\frac{1}{2}K^\top(\hat{p})M & 0 & -\frac{\kappa_w}{2}I \end{bmatrix} \preccurlyeq 0, \tag{25}$$

where $\boldsymbol{A}(\hat{p}) = A(\hat{p}) + L(\hat{p})C$, $\boldsymbol{H}(\hat{p}) = H + K(\hat{p})C$ and $\Lambda = \operatorname{diag}(\ell_1, \dots, \ell_{n_\phi})$, then Assumption 4 is satisfied.

The condition in (25) represents infinitely many linear matrix inequalities (LMIs) in P, $PL(\hat{p})$, M, $MK(\hat{p})$, $\kappa_{\tilde{x}}$, κ_v and κ_w, due to its dependence on $\hat{p}$ and p. In order to solve (25), it either needs to be discretized or, as we will see in our case study, sometimes structure can be exploited to reduce (25) to a finite number of LMIs. If Proposition 1 and Assumption 5 apply, then Theorems 1 or 2 hold, respectively, when the static or dynamic sampling policy is used.

Consider the system (22) with the following matrices

$$\begin{aligned} A(p) &= \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} - p\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{bmatrix} T_s,\ G(p) = p\begin{bmatrix} \frac{T_s}{2} \\ T_s \end{bmatrix}, \\ H^\top &= \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ B(p) = \begin{bmatrix} T_s \\ T_s \end{bmatrix} + p\begin{bmatrix} T_s \\ -T_s \end{bmatrix} \text{ and } C = I, \end{aligned} \tag{26}$$

which is obtained by discretizing a continuous-time system, see [13], with sampling time $T_s = 0.01$. The nonlinearity in (22) is given by

$$\phi(v) = v + \sin(v), \tag{27}$$

which satisfies (23) with Lipschitz constant $\ell_1 = 2$. Moreover, the parameter p belongs to $\mathbb{P} := [1, 50]$. This example is a variation on [13, Example 1] and [18, Example 1] where we included process and measurement noise and an additional parameter dependency in $B(p)$. Notice that the system matrices (26) all depend affinely on the unknown parameter. If we restrict the observer matrices $L(\hat{p})$ and $K(\hat{p})$ to also be affine in p, i.e.,

$$L(\hat{p}) = L_0 + \hat{p}L_1 \quad \text{and} \quad K(\hat{p}) = K_0 + \hat{p}K_1, \tag{28}$$

with $L_i \in \mathbb{R}^{n_x \times n_y}$ and $K_i \in \mathbb{R}^{n_\phi \times n_y}$, $i = 0, 1$, the LMI in (25) becomes affine in $(p, \hat{p}) \in \mathbb{P} \times \mathbb{P}$. Since $\mathbb{P} \times \mathbb{P}$ is convex, the condition (25) is satisfied for all $p, \hat{p} \in \mathbb{P}$ if and only if it is satisfied at each of the $(2n_p)^2 = 4$ vertices [19]. We set $\kappa_{\tilde{x}} = 0.1$ and minimize $\kappa_v + 5\kappa_w$ subject to (25), for all $p, \hat{p} \in \{1, 50\}$, by means of the MATLAB toolbox YALMIP [20] together with the external solver MOSEK [21]. Restricting ourselves to a Lyapunov function, which is independent of $\hat{p}$ and to affine observer matrices (28) introduces conservatism compared to, for instance, sampling the parameter space and then solving the LMIs. However, it has the advantage that resampling in the dynamic sampling policy is computationally efficient as it only requires evaluating (28) without solving LMIs on-line.

Both the static and dynamic sampling policy are implemented using $N = 10$ equidistant parameter samples, i.e., $\hat{p}_i = \underline{p} + \frac{\bar{p} - \underline{p}}{2N} + (i-1)\frac{\bar{p} - \underline{p}}{N}$, where $\underline{p}$ and $\bar{p}$ denote the extrema of the set that is currently being sampled (for the dynamic scheme $\underline{p}$ and $\bar{p}$ will move closer together over time). For this sampling scheme, we can guarantee (17) with $\rho(\Delta_m, N) = \frac{\Delta_m}{N}$, as the distance between the true parameter and the nearest sample never exceeds half the distance between neighbouring samples. This also guarantees (3) for the static sampling since $\rho(\Delta_0, N) \to 0$ as $N \to \infty$. We simulate both the static and the dynamic schemes with design parameters $\lambda = 0.995$, $N_{pe} = M_d = 1 \cdot 10^3$ (which corresponds to 10 seconds) and $\alpha = 0.8$. The resulting parameter estimate and norm of the state error are shown in Fig. 2 together with the shrinking parameter set and estimated noise level $\upsilon_{\tilde{p}}(\Delta_v) + \omega_{\tilde{p}}(\Delta_w)$ with $\Delta_v = \Delta_w = 0.01$.

Figure 2 shows that for both schemes the parameter estimates as well as the state estimate converge within a certain margin of their true value. As can be seen in Fig. 2,

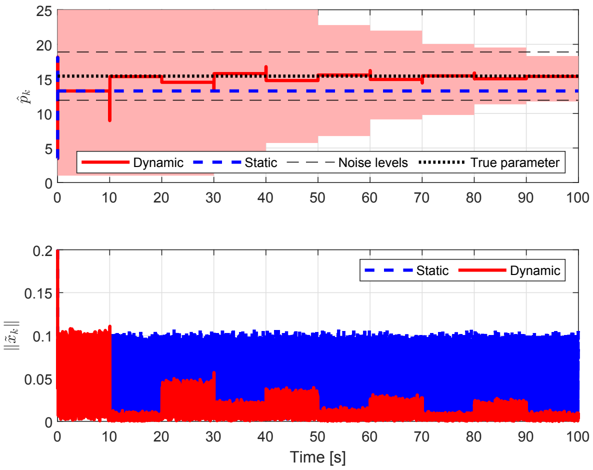

Fig. 2. Parameter estimate (top) and norm of the state estimation error (bottom) using the static (dashed) and dynamic sampling policy (solid). The transparent regions indicate the set being sampled by the dynamic sampling policy. The dotted black line indicates the true parameter and the dashed black lines are the noise levels.

the first resampling occurs after 10 seconds, which causes the parameter estimate to jump. The spikes in the parameter estimation error at the switching instances are a result of the monitoring signals being reset, which may cause the supervisor to select a "suboptimal" observer temporarily. Figure 2 also shows that the estimates do not necessarily become more accurate after individual zoom-ins, which is explained by the fact that if one parameter sample happens to be very accurate, it is not necessarily preserved during the resampling. It should be noted that the number of observers $N = 10$ used here is significantly less than the theoretical estimates. To be more specific, our estimates dictate that at least $N^\star = 357$ observers are required in the static sampling policy to guarantee that the parameter converges to within $\|\tilde{p}_k\| \leqslant 3 + \upsilon_{\tilde{p}}(\Delta_v) + \omega_{\tilde{p}}(\Delta_w)$. However, the simulations show that this estimate is conservative and that the margin is already achieved for $N = 10$. For the dynamic sampling policy, the estimated required number of observers decreases to $N^\star = 94$, which is still conservative, however, it confirms that the dynamic scheme requires fewer observers to guarantee similar accuracy.

VI. Conclusions

In this paper, we presented two schemes to jointly estimate parameters and states of discrete-time nonlinear systems in the presence of bounded noise. The first scheme utilizes a static sampling policy and the second scheme uses a dynamic sampling policy. For both schemes, convergence guarantees are provided, which also show that the dynamic scheme typically requires lower computational effort. These results were illustrated by means of a numerical example.

Future work is directed towards obtaining, in an easy and non-conservative manner, the estimates for the required number of observers, time until convergence and minimum time between subsequent zoom-ins on the parameter space needed to get the guarantees as provided in our main theorems. One concrete direction could be the DIRECT algorithm proposed in [14], which eliminates the need to estimate the number of observers a-priori and can be extended to the noisy case to overcome one of these drawbacks. Obtaining non-conservative estimates of the contribution of the noise on the convergence margins is another important research topic. Extending the framework to stochastic noise assumptions may improve performance at the cost of not having worst-case convergence guarantees. Finally, allowing for slowly time-varying parameters, such as in [22], in our estimation framework is an interesting future research direction.

References

[1] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21 980–22 012, 2020.

[2] M. L. G. Boerlage, "Rejection of disturbances in multivariable motion systems," Ph.D. dissertation, Technische Universiteit Eindhoven, 2008.

[3] N. Atanasov, R. Tron, V. M. Preciado, and G. J. Pappas, "Joint estimation and localization in sensor networks," in *53rd IEEE Conf. Decis. Control*, 2014, pp. 6875–6882.

[4] F. Sawo, "Nonlinear state and parameter estimation of spatially distributed systems," Ph.D. dissertation, Karlsruher Institut für Technologie, 2009.

[5] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[6] G. Besançon, J. De León-Morales, and O. Huerta-Guevara, "On adaptive observers for state affine systems," *Int. J. Control*, vol. 79, no. 6, pp. 581–591, 2006.

[7] M. Farza, M. M'Saad, T. Maatoug, and M. Kamoun, "Adaptive observers for nonlinearly parameterized class of nonlinear systems," *Automatica*, vol. 45, no. 10, pp. 2292–2299, 2009.

[8] E. Alvaro-Mendoza, J. De León-Morales, and O. Salas-Peña, "State and parameter estimation for a class of nonlinear systems based on sliding mode approach," *ISA Trans.*, vol. 112, pp. 99–107, 2020.

[9] M. S. Chong, D. Nešić, R. Postoyan, and L. Kuhlmann, "Parameter and state estimation for nonlinear systems," *IEEE Trans. Autom. Control*, vol. 60, no. 9, pp. 2336–2349, 2015.

[10] M. S. Chong, H. Sandberg, and J. P. Hespanha, "A secure state estimation algorithm for nonlinear systems under sensor attacks," in *59th IEEE Conf. Decis. Control*, 2020, pp. 5743–5748.

[11] J. Hespanha, D. Liberzon, A. S. Morse, B. D. O. Anderson, T. S. Brinsmead, and F. De Bruyne, "Multiple model adaptive control. Part 2: Switching," *Int. J. Robust Nonlinear Control*, vol. 11, no. 5, pp. 479–496, 2001.

[12] W. P. M. H. Heemels, J. Daafouz, and G. Millerioux, "Observer-based control of discrete-time LPV systems with uncertain parameters," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2130–2135, 2010.

[13] S. Ibrir, "Circle-criterion approach to discrete-time nonlinear observer design," *Automatica*, vol. 43, no. 8, pp. 1432–1441, 2007.

[14] M. S. Chong, R. Postoyan, S. Z. Khong, and D. Nešić, "Supervisory observer for parameter and state estimation of nonlinear systems using the DIRECT algorithm," in *56th IEEE Conf. Decis. Control*, 2017, pp. 2089–2094.

[15] T. J. Meijer, V. S. Dolk, M. S. Chong, R. Postoyan, B. de Jager, D. Nešić, and W. P. M. H. Heemels, "Joint parameter and state estimation of noisy discrete-time nonlinear systems: A supervisory multi-observer approach," 2021, preprint: https://arxiv.org/abs/2109.12450.

[16] Z.-P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857–869, 2001.

[17] R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 183–191, 1984.

[18] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "A robust circle-criterion observer-based estimator for discrete-time nonlinear systems in the presence of sensor attacks," in *57th IEEE Conf. Decis. Control*, 2018, pp. 571–576.

[19] C. Scherer and S. Weiland, "Linear matrix inequalities in control," in *The Control Systems Handbook, Second Edition: Control Systems Advanced Methods*, W. S. Levine, Ed. CRC Press, 2011, pp. 24/1–24/30.

[20] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *2004 IEEE Int. Symp. Comput. Aided Control Syst. Des.*, 2004, pp. 285–289.

[21] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019.

[22] L. Cuevas, "A general estimation framework for nonlinear singularly perturbed systems," Ph.D. dissertation, The University of Melbourne, 2019.