Interval Observers for a Class of Nonlinear Systems Using Gaussian Process Models

Alexandre Capone, Sandra Hirche

***Abstract*— An interval observer design approach for partially unknown nonlinear systems is developed, where the unknown system component is modeled using Gaussian processes and noisy system measurements. The proposed method is applicable for bounded nonlinear systems where the system uncertainty is described by a Lipschitz continuous function. The interval observer generates a correct estimation error with high probability, and the error bound is decreased by employing new training data points.**

I. Introduction

One of the most frequently employed forms of state observer is the Luenberger observer [1], which essentially corresponds to a linear dynamical system that converges towards the actual system's state over time. Even though the Luenberger observer is generally only applicable for linear systems, its success has prompted a number of similar approaches for nonlinear systems [2], [3]. However, such observers typically require perfect model knowledge, which is often an unrealistic requirement in practice. Moreover, even though nonlinear observers often guarantee convergence of the state estimate towards the true states, they typically do not provide any information regarding the precision of the computed estimate. Interval observers aim to provide an upper and lower bound for the system's state during the state estimation procedure. This is achieved by employing two observers that serve as upper and lower bound for the state estimation. Interval observers have become a particularly popular tool when the system model contains unknown uncertainties, as they enable to determine the worst case impact of the system uncertainty on the state estimate [4]–[6].

Whenever a system's parametric structure is known, classical system identification methods can often be employed to derive good model estimates [7]. However, acquiring parametric structures for nonlinear systems is often hard, and can lead to discrepancies between model prediction and system behavior. In order to overcome this shortcoming, Gaussian processes (GPs) are increasingly employed to model nonlinear systems [8]–[10]. GPs are a form of nonlinear stochastic processes, which define a distribution over infinite sets of functions by assigning a mean and covariance to every point in the input space. GPs are employed as a nonparametric regression tool to model system dynamics using a Bayesian framework [11].

This work was supported by the was supported by the European Research Council under Grant 33765.

The authors are with the Department of Electrical Engineering, Technical University of Munich, Germany alexandre.capone@tum.de,hirche@tum.de

So far, most of the work carried out at the intersection between observers and GPs corresponds to the development of Bayesian filters for GP models, most of which employ a Kalman filter formalism [12]–[14]. Such filters aim to compute the most likely system estimate by employing a Bayesian framework in noisy settings. This is achieved by interpreting the GP model variance as the variance of the normally distributed model disturbances. However, Bayesian filters not always produce satisfying results, and the choice of filter can significantly influence the observer performance depending on the system model [13]. Moreover, no error bounds for the observer's estimate are obtained when such filters are employed.

This paper provides a general framework to design an interval observer based on Gaussian process models. The proposed approach employs the Gaussian process model's variance to determine upper and lower bounds for the system estimate. The upper and lower bounds are shown to hold with high probability.

Notation: Bold lowercase letters denote vectors, bold capital letters denote matrices. $\mathbb{R}_+$ denotes the set of positive real numbers. ∇ denotes the gradient operator. For any pair of vectors $\mathbf{w}, \mathbf{v} \in \mathbb{R}^p, \quad p \in \mathbb{N}$, the notation $\mathbf{w} \leq \mathbf{v}$ is employed to denote $u_i \leq w_i, \quad \forall i \in \{1, \cdots, p\}$.

The remainder of this paper is arranged as follows. Section II introduces the general problem setting. Section III gives a brief introduction to GP models. The general observer design approach is provided in Section IV-B. A numerical illustration of the proposed approach is given in Section V, after which come some concluding remarks, in Section VI.

II. Problem setting

We introduce the general problem setting that is analyzed in this work. Consider an n-dimensional autonomous nonlinear system of the following form.

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0, \\ y &= \mathbf{C}\mathbf{x}, \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathcal{R} \subseteq \mathbb{R}^n$ denotes the system's states, and $\mathcal{R}$ is a known and compact subset of $\mathbb{R}^n$. $y \in \mathcal{Y} \subseteq \mathbb{R}$ is the system's output. $\mathbf{C} \in \mathbb{R}_+^{1\times n}$ is a known positive row vector, the nonlinear function $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n$ denotes the known component of the system dynamics, whereas $\mathbf{g} : \mathcal{R} \mapsto \mathbb{R}^n$ describes the unknown and nonlinear component. The goal of this paper is to design an interval observer for a setting where the unknown component $\mathbf{g}(\cdot)$ is modeled using a GP and noisy system measurements. To this end, we assume that the unknown function $\mathbf{g}(\cdot)$ satisfies the following condition.

Assumption 1: The function $\mathbf{g}(\cdot)$ is bounded and Lipschitz continuous with respect to its arguments, i.e., there exist positive scalars B_g and B_{dg}, such that $||\mathbf{g}(\mathbf{x})||_2 \leq B_g$ and $||\nabla \mathbf{g}(\mathbf{x})||_2 \leq B_{dg}$ holds for all $\mathbf{x} \in \mathcal{R}$.
This is a very unrestrictive assumption due to the compactness of $\mathcal{R}$.

III. Gaussian Process Model

Given an input space $\mathcal{I} \subset \mathbb{R}^d$ with dimension $d \in \mathbb{N}$, a GP describes a distribution over functions by assigning a Gaussian distribution to every point $\xi \in \mathcal{I}$. A GP is fully defined by a *mean* function $m : \mathbb{R}^d \mapsto \mathbb{R}$ and *covariance* function $k : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$, where the latter is also called *kernel* function. A GP is then denoted as $\mathcal{GP}(m, k)$. In practice, the mean $m(\cdot)$ is often set to zero, unless prior knowledge about the mean is available. The kernel function $k(\cdot, \cdot)$ expresses the similarity between two function inputs $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$, and also reflects smoothness properties of the actual system.
GP models are used as a nonparametric regression tool to approximate nonlinear functions of the form $f : \mathbb{R}^d \mapsto \mathbb{R}$ [11]. This is achieved by computing a posterior distribution conditioned on a set of noisy system measurements of the form $y_f^{(k)} = f(\boldsymbol{\xi}^{(k)}) + \epsilon^{(k)}$, where the superscript (k) denotes the system measurement, $k \in \{1, \cdots, N\}$, and $N \in \mathbb{N}$ denotes the number of system measurements. $\boldsymbol{\xi}^{(k)} \in \mathcal{I}$ are the system inputs and $\epsilon^{(k)}$ is measurement noise. Assuming a mean $m(\cdot)$ of zero, the posterior distribution corresponding to the GP is then given by the mean and covariance

$$\mu_N(\boldsymbol{\xi}) = \mathbf{k}^T(\boldsymbol{\xi}) \left(\mathbf{K} + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{y}_N, \tag{2}$$

$$k_N(\boldsymbol{\xi}, \boldsymbol{\xi}') = k(\boldsymbol{\xi}, \boldsymbol{\xi}') - \mathbf{k}^T(\boldsymbol{\xi}) \left(\mathbf{K} + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{k}(\boldsymbol{\xi}'), \tag{3}$$

where $\mathbf{y}_N = (y_f^{(1)}, ..., y_f^{(N)})^{\mathrm{T}} \in \mathbb{R}^N$ contains the noisy output measurements, and the entries of the function $\mathbf{k} : \mathbb{R}^d \mapsto \mathbb{R}^N$ and matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ are given by $k_i(\boldsymbol{\xi}) = k(\boldsymbol{\xi}^{(i)}, \boldsymbol{\xi})$ and $K_{ij} = k(\boldsymbol{\xi}^{(i)}, \boldsymbol{\xi}^{(j)})$, $i, j = 1, \cdots, N$, respectively.
A GP kernel defines a Hilbert space, denoted *reproducing kernel Hilbert space* (RKHS). The induced norm $\| \cdot \|_k$ measures smoothness with respect to the kernel $k(\cdot, \cdot)$. For a detailed discussion on RKHSs, see [15]. The following theorem is due to [16].

Theorem 1: Suppose that $\|f\|_k^2 \leq B_f$, $\mathbf{y}_N$ is a vector of N noisy system measurements, where the noise ϵ is additive and zero-mean σ-sub-Gaussian. Choose $\beta_N = B_f + 4\sigma\sqrt{\gamma_N + 1 + \ln(\frac{1}{\delta})}$, where $\delta \in (0, 1)$ and γ_N is the maximal information that can be gained about the GP prior using N noisy data samples, i.e.,

$$\gamma_N = \max_{\tilde{\boldsymbol{\xi}}^{(1)}, \cdots, \tilde{\boldsymbol{\xi}}^{(N)} \in \mathcal{I}} I(\tilde{\mathbf{y}}_N, \tilde{\mathbf{f}}_N), \tag{4}$$

where $\tilde{\mathbf{f}}_N = (f(\tilde{\boldsymbol{\xi}}^{(1)}), ..., f(\tilde{\boldsymbol{\xi}}^{(N)}))^{\mathrm{T}}$ contains noiseless measurements and $\tilde{\mathbf{y}}_N = (f(\tilde{\boldsymbol{\xi}}^{(1)}) + \epsilon^{(1)}, ..., f(\tilde{\boldsymbol{\xi}}^{(N)}) + \epsilon^{(N)})^{\mathrm{T}}$ contains noisy measurements. Then, for all $N \geq 1$ and $\boldsymbol{\xi} \in \mathcal{I}$, the following holds with probability of at least $1 - \delta$.

$$|f(\boldsymbol{\xi}) - \mu_{N-1}(\boldsymbol{\xi})| \leq \beta_N \sigma_{N-1}(\boldsymbol{\xi}). \tag{5}$$

If the set $\mathcal{I}$ is compact, then γ_N has sublinear dependence on N for a multitude of kernels and can efficiently be approximated up to a constant [17]. This implies that the GP model error bound can be decreased with high probability despite β_N growing with N. If the bound B_f is not available a priori, one can be obtained by employing guess-and-doubling strategies [17].
Each entry of the unknown function $\mathbf{g}(\cdot)$ is approximated using a GP and corresponding noisy measurements. The training data consists of pairs of system inputs $\mathbf{x}$ and noisy measurements, i.e., $\mathcal{D} = \left\{ g_i(\mathbf{x}^{(j)}) + \epsilon_i^{(j)} \right\},\ i = 1, \cdots, n,\ j = 1, \cdots, N.$ In order to obtain the necessary noisy measurements, the known component of the signal, i.e., $\mathbf{f}(\mathbf{x})$, is subtracted from noisy measurements of the system's time derivative $\dot{\mathbf{x}}$. The data points are collected prior to the observer design, and should be chosen such that the model error bound (5) is reduced. The variance term $\sigma_{N-1}(\boldsymbol{\xi})$ is small close to data points [11], hence the data points should cover the region of interest efficiently. For a detailed discussion on how data and kernel choice affect the bound (5), the reader is referred to [17]. The computed GP mean functions are then employed to approximate the system dynamics corresponding to the entries of the functions $\mathbf{g}$, i.e.,

$$\hat{\mathbf{g}}(\mathbf{x}) = \begin{pmatrix} \hat{g}_1(\mathbf{x}) & \hat{g}_2(\mathbf{x}) & \cdots & \hat{g}_n(\mathbf{x}) \end{pmatrix}^{\mathrm{T}}, \tag{6}$$

where $\hat{\mathbf{g}} : \mathbb{R}^n \mapsto \mathbb{R}^n$ denotes the estimated model, and $\hat{g}_i(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R},\ i = 1, \cdots, n$ are GP means. Even though a multivariate formulation exists for GPs [11], which enables the function $\mathbf{g}(\cdot)$ to be approximated by a single GP, such a formulation is very cumbersome and corresponds to a high computational cost, as opposed to employing multiple scalar GPs. The following assumptions are assumed to hold.

Assumption 2: The kernel $k(\cdot, \cdot)$ used for the GP model is Lipschitz continuous and bounded.

Assumption 3: The components of the function $\mathbf{g}$ have bounded RKHS norm with respect to the kernel $k(\cdot, \cdot)$, i.e., $||g_i||_k < \infty, \quad \forall i \in \{1, \cdots, n\}$.
Assumption 3 implies that the kernel $k(\cdot, \cdot)$ is adequate to approximate the function $\mathbf{g}(\cdot)$. Choosing an appropriate kernel comprises a problem-dependent task, which strongly depends on the system characteristics. In practice, universal kernels are often employed [11], since they can arbitrarily approximate continuous functions on compact sets. Moreover, Assumption 2 together with Assumption 3 implies that the entries of the functions $\mathbf{g}(\cdot)$ are smooth and bounded. This applies to a number of physical systems.
A probabilistic model error bound for the i-th entry of $\hat{\mathbf{g}}(\cdot)$ at a given point $\mathbf{x} \in \mathbb{R}^n$ is given by $\beta_N^i \sigma_N^i(\mathbf{x})$, where β_N^i is chosen as in Theorem 1 for a fixed δ. For simplicity of exposition, the same δ is assumed to have been chosen to compute β_N^i for all i.

IV. Observer Design

This section provides a brief introduction to interval observers and describes how such an observer is designed for a system of the form given by Equation (1).

A. Interval Observers

In this section, a brief introduction to interval observers is given, together with related concepts.

Definition 1: Consider the system given by Equation (1). A system of the form

$$\Sigma = \begin{cases} \dot{\mathbf{x}}^+ = \mathbf{h}^+(\mathbf{x}^+, \mathbf{x}^-, y), & \mathbf{x}^+(0) = \mathbf{x}_0^+ \\ \dot{\mathbf{x}}^- = \mathbf{h}^-(\mathbf{x}^+, \mathbf{x}^-, y), & \mathbf{x}^-(0) = \mathbf{x}_0^- \end{cases} \tag{7}$$

is called a framer for Equation (1) if $\mathbf{x}^-(t) \leq \mathbf{x}(t) \leq \mathbf{x}^+(t)$ for all $t \geq 0$.

If a framer produces estimation errors $\mathbf{x}^+ - \mathbf{x}$ and $\mathbf{x}^- - \dot{\mathbf{x}}$ that are globally uniformly ultimately bounded, then it is called an interval observer. Hence, interval observers consist of dynamical systems that act as upper and lower bounds for the system's actual state.

B. Design Approach

In order to avoid unnecessarily conservative observer bounds, the bound of the known set $\mathcal{R}$ are taken into account. This is carried out by defining the bounds $\mathbf{x}^{\max} \in \mathbb{R}^n$ and $\mathbf{x}^{\min} \in \mathbb{R}^n$, whose entries are given by $x_i^{\max} = \max_{\mathbf{x} \in \mathcal{R}} x_i$ and $x_i^{\min} = \min_{\mathbf{x} \in \mathcal{R}} x_i$, respectively.

Even though an interval observer can be trivially obtained by employing the bounds $\mathbf{x}^{\max}$ and $\mathbf{x}^{\min}$, such a bound is generally too conservative. Hence, in this subsection, an interval observer that generates tight probabilistic bounds is designed.

Consider the dynamical system obtained by setting $\mathbf{g}(\cdot) = \mathbf{0}$, i.e., the reduced system

$$\begin{aligned} \dot{\tilde{\mathbf{x}}} &= \mathbf{f}(\tilde{\mathbf{x}}), \quad \tilde{\mathbf{x}}(0) = \tilde{\mathbf{x}}_0, \\ \tilde{y} &= \mathbf{C}\tilde{\mathbf{x}}, \end{aligned} \tag{8}$$

where $\tilde{\mathbf{x}} \in \mathbb{R}^n$ is the system's state and $\tilde{\mathbf{x}}_0 \in \mathbb{R}^n$ is the system's initial condition. The following assumption is made.

Assumption 4: Let $\tilde{\mathbf{x}}_0^-, \tilde{\mathbf{x}}_0^+ \in \mathbb{R}^n$, such that $\tilde{\mathbf{x}}_0^- \leq \tilde{\mathbf{x}}_0 \leq \tilde{\mathbf{x}}_0^+$. There exist known smooth functions $\mathbf{f}^+ : \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{Y} \mapsto \mathcal{R}$ and $\mathbf{f}^- : \mathcal{R} \times \mathcal{R} \times \mathcal{Y} \mapsto \mathcal{R}$, such that the system

$$\Sigma_{\text{red}} = \begin{cases} \dot{\tilde{\mathbf{x}}}^+ &= \mathbf{f}^+(\tilde{\mathbf{x}}^+, \tilde{\mathbf{x}}^-, \tilde{y}), \quad \tilde{\mathbf{x}}^+(0) = \tilde{\mathbf{x}}_0^+ \\ \dot{\tilde{\mathbf{x}}}^- &= \mathbf{f}^-(\tilde{\mathbf{x}}^+, \tilde{\mathbf{x}}^-, \tilde{y}), \quad \tilde{\mathbf{x}}^-(0) = \tilde{\mathbf{x}}_0^- \end{cases} \tag{9}$$

is an interval observer for Equation (8), and $f_i^+(\tilde{\mathbf{x}}^+, \tilde{\mathbf{x}}^-, \tilde{y}) \geq f_i(\tilde{\mathbf{x}}) \geq f_i^-(\tilde{\mathbf{x}}^+, \tilde{\mathbf{x}}^-, \tilde{y})$ holds for all $\tilde{\mathbf{x}}^+, \tilde{\mathbf{x}}^-, \tilde{\mathbf{x}} \in \mathbb{R}^n$ with $\tilde{\mathbf{x}}^+ \geq \tilde{\mathbf{x}} \geq \tilde{\mathbf{x}}^-$ and $\tilde{x}_i^+ = \tilde{x}_i = \tilde{x}_i^-$. Moreover, there exists a Hurwitz matrix $\mathbf{A} \in \mathbb{R}^{2n \times 2n}$, such that the following holds

$$\mathbf{A}\tilde{\boldsymbol{e}} \geq \begin{pmatrix} \mathbf{f}^+(\tilde{\mathbf{x}}^+, \tilde{\mathbf{x}}^-, \tilde{y}) - \mathbf{f}(\tilde{\mathbf{x}}) \\ \mathbf{f}(\tilde{\mathbf{x}}) - \mathbf{f}^-(\tilde{\mathbf{x}}^+, \tilde{\mathbf{x}}^-, y) \end{pmatrix} = \dot{\tilde{\boldsymbol{e}}}, \tag{10}$$

where $\tilde{\boldsymbol{e}} := \begin{pmatrix} \tilde{\mathbf{x}}^+ - \tilde{\mathbf{x}} & \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^- \end{pmatrix}^{\mathrm{T}}$ concatenates the errors of the interval observer for the reduced system.

This means that the error of the interval observer for the reduced system is bounded by an exponentially stable linear system. Such an observer can be obtained using a series of different approaches [5], [6], [18].

Proposition 1: Let $\hat{\mathbf{g}}(\cdot)$ denote the GP estimate of the function $\mathbf{g}(\cdot)$, as shown in Equation (6). There exist Lipschitz continuous and differentiable functions $\hat{\mathbf{g}}^+ : \mathcal{R} \mapsto \mathbb{R}^n$ and $\hat{\mathbf{g}}^- : \mathcal{R} \mapsto \mathbb{R}^n$, such that $\nabla \hat{\mathbf{g}}^+(\mathbf{x}) \geq \mathbf{0}$, $\nabla \hat{\mathbf{g}}^-(\mathbf{x}) \geq \mathbf{0}$, and $\hat{\mathbf{g}}^+(\mathbf{x}) - \hat{\mathbf{g}}^-(\mathbf{x}) = \hat{\mathbf{g}}(\mathbf{x})$ holds for all $\mathbf{x} \in \mathcal{R}$.

Proof: Due to Assumption 2, the function $\hat{\mathbf{g}}(\cdot)$ is Lipschitz continuous in $\mathbb{R}^n$. The proof then follows straightforwardly from [18, Property 3]. ∎

There are infinitely many functions that satisfy such requirements [18], and the choice of $\hat{\mathbf{g}}^+(\cdot)$ and $\hat{\mathbf{g}}^-(\cdot)$ can strongly affect the performance of the interval observer. How to choose such functions is out of the scope of this paper.

In order to obtain an interval observer, the model uncertainty needs to be taken into account. In order to do so, define the entries of the vector valued function $\boldsymbol{\zeta}_{\mathbf{x}} : \mathcal{R} \times \mathcal{R} \mapsto \mathbb{R}^n$ as $\zeta_{\mathbf{x},i}(\mathbf{x}_1, \mathbf{x}_2) = \max_{\mathbf{x}_1 \leq \mathbf{x} \leq \mathbf{x}_2} \beta_N^i \sigma_N^i(\mathbf{x})$. Note that due to Assumption 2, the function $\boldsymbol{\zeta}_{\mathbf{x}}(\cdot, \cdot)$ is bounded. Denote the corresponding bound by $\boldsymbol{\zeta}_{\mathbf{x},\max}$, where $\zeta_{\mathbf{x},\max,i} = \max_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{R}} \zeta_{\mathbf{x},i}(\mathbf{x}_1, \mathbf{x}_2)$.

Lemma 1: Let $\mathbf{z}, \mathbf{z}^+, \mathbf{z}^- \in \mathcal{R}$, such that $\mathbf{z}^- \leq \mathbf{z} \leq \mathbf{z}^+$. Choose $\hat{\mathbf{g}}^+(\cdot)$ and $\hat{\mathbf{g}}^-(\cdot)$ as discussed in Proposition 1. Then the following holds with probability at least $1 - n\delta$

$$\begin{aligned} &\hat{g}_i^+(\mathbf{z}^-) - \hat{g}_i^-(\mathbf{z}^+) - \zeta_{\mathbf{x},i}(\mathbf{z}^-, \mathbf{z}^+) \leq g_i(\mathbf{z}) \\ \leq &\hat{g}_i^+(\mathbf{z}^+) - \hat{g}_i^-(\mathbf{z}^-) + \zeta_{\mathbf{x},i}(\mathbf{z}^-, \mathbf{z}^+). \end{aligned} \tag{11}$$

Proof: By design,

$$\begin{aligned} \hat{g}_i(\mathbf{z}) - \zeta_{\mathbf{x},i}(\mathbf{z}^-, \mathbf{z}^+) &\leq \hat{g}_i(\mathbf{z}) - \beta_N^i \sigma_N^i(\mathbf{z}), \\ \hat{g}_i(\mathbf{z}) + \beta_N^i \sigma_N^i(\mathbf{z}) &\leq \hat{g}_i(\mathbf{z}) + \zeta_{\mathbf{x},i}(\mathbf{z}^-, \mathbf{z}^+). \end{aligned} \tag{12}$$

holds. Moreover, due to Theorem 1, the following holds with probability at least $1 - \delta$.

$$\hat{g}_i(\mathbf{z}) - \beta_N^i \sigma_N^i(\mathbf{z}) \leq g_i(\mathbf{z}) \leq \hat{g}_i(\mathbf{z}) + \beta_N^i \sigma_N^i(\mathbf{z}). \tag{13}$$

Applying the union bound yields the same result for all $i = 1, \cdots, n$ with probability $1 - n\delta$. Moreover, due to the monotonicity of $\hat{\mathbf{g}}^+(\cdot)$ and $\hat{\mathbf{g}}^-(\cdot)$, as discussed in Proposition 1, together with $\mathbf{z}^- \leq \mathbf{z} \leq \mathbf{z}^-$,

$$\hat{g}_i^+(\mathbf{z}^-) - \hat{g}_i^-(\mathbf{z}^+) \leq \hat{g}_i(\mathbf{z}) \leq \hat{g}_i^+(\mathbf{z}^+) - \hat{g}_i^-(\mathbf{z}^-) \tag{14}$$

holds. Putting together Equations (12) to (14) yields the desired result. ∎

In order to design the interval observer, two nonlinear functions $\mathbf{b}^+ : \mathcal{R} \times \mathcal{R} \times \mathbb{R} \mapsto \mathbb{R}^n$, $\mathbf{b}^- : \mathcal{R} \times \mathcal{R} \times \mathbb{R} \mapsto \mathbb{R}^n$ are introduced, which are given by

$$\begin{aligned} \mathbf{b}^+(\mathbf{z}, \mathbf{z}', a) =& \mathbf{f}^+(\mathbf{z}, \mathbf{z}', a) + \hat{\mathbf{g}}^+(\mathbf{z}) - \hat{\mathbf{g}}^-(\mathbf{z}') \\ &+ \boldsymbol{\zeta}_{\mathbf{z}}(\mathbf{z}', \mathbf{z}) + \mathbf{F}(\mathbf{z} - \mathbf{z}') \\ &+ \boldsymbol{\theta}^+(\mathbf{C}\mathbf{z} - a), \\ \mathbf{b}^-(\mathbf{z}, \mathbf{z}', a) =& \mathbf{f}^-(\mathbf{z}, \mathbf{z}', a) + \hat{\mathbf{g}}^+(\mathbf{x}') - \hat{\mathbf{g}}^-(\mathbf{x}) \\ &- \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}', \mathbf{x}) - \mathbf{F}'(\mathbf{x} - \mathbf{x}') \\ &+ \boldsymbol{\theta}^-(\mathbf{C}\mathbf{x}' - a), \end{aligned} \tag{15}$$

where $\mathbf{F}^+, \mathbf{F}^- \in \mathbb{R}^{n \times n}$ are gain matrices and $\boldsymbol{\theta}^-, \boldsymbol{\theta}^+ \in \mathbb{R}^n$ are gain vectors. Choose $\mathbf{x}_0^-, \mathbf{x}_0^+ \in \mathbb{R}^n$ such that

$\mathbf{x}_0^- \leq \mathbf{x}_0 \leq \mathbf{x}_0^+$ holds. The candidate interval observer is then given by

$$\Sigma_{\mathcal{GP}} = \begin{cases} \dot{\mathbf{x}}^+ = & \mathbf{o}^+(\mathbf{x}^+, \mathbf{x}^-, y), & \mathbf{x}^+(0) = \mathbf{x}_0^+ \\ \dot{\mathbf{x}}^- = & \mathbf{o}^-(\mathbf{x}^+, \mathbf{x}^-, y), & \mathbf{x}^-(0) = \mathbf{x}_0^- \end{cases} \tag{16}$$

where $\mathbf{x}^+, \mathbf{x}^- \in \mathcal{R}$ are the estimated state bounds and $\mathbf{o}^+ : \mathcal{R} \times \mathcal{R} \times \mathbb{R} \mapsto \mathbb{R}^n$, $\mathbf{o}^- : \mathcal{R} \times \mathcal{R} \times \mathbb{R} \mapsto \mathbb{R}^n$ are nonlinear functions, whose entries are given by

$$o_i^+(\mathbf{x}^+, \mathbf{x}^-, y) = \begin{cases} 0, & \text{if } x_i^+ \geq x_{\max,i}^+ \text{ and} \\ & b_i^+(\mathbf{x}^+, \mathbf{x}^-, y) \geq 0 \\ b_i^+(\mathbf{x}^+, \mathbf{x}^-, y), & \text{otherwise} \end{cases} \tag{17}$$

and

$$o_i^-(\mathbf{x}^+, \mathbf{x}^-, y) = \begin{cases} 0, & \text{if } x_i^- \leq x_{\min,i}^+ \text{ and} \\ & b_i^-(\mathbf{x}^+, \mathbf{x}^-, y) \leq 0 \\ b_i^-(\mathbf{x}^+, \mathbf{x}^-, y), & \text{otherwise} \end{cases} \tag{18}$$

respectively.

Lemma 2: Let Assumptions 2 to 4 hold. Then the system given by Equation (16) corresponds to a framer for Equation (1) with probability $1 - n\delta$.

Proof: Consider the estimation error $\mathbf{e}^+ = \mathbf{x}^+ - \mathbf{x}$. Since $\mathbf{e}^+(0) \geq \mathbf{0}$ holds by design, and $\mathbf{e}^+(t)$ is continuous in t, it suffices to prove that $\dot{e}_i \geq 0$ holds if $e_i = 0$ for all $i \in \{1, \cdots, n\}$. Hence, assume $e_i = 0$ holds for an $i \in \{1, \cdots, n\}$. The two cases corresponding to Equation (17) need to be analyzed separately. If $x_i^+ \geq x_{\max,i}^+$ and $b_i^+(\mathbf{x}^+, \mathbf{x}^-, y) \geq 0$ holds, then $e_i = 0$ implies $x_i^+ = x_{\max,i}^+$, which in turn means that $\dot{x}_i \leq 0$ must hold. Hence,

$$\begin{aligned} \dot{e}_i^+ =& o_i^+(\mathbf{x}^+, \mathbf{x}^-, y) - f_i(\mathbf{x}) - g_i(\mathbf{x}) \\ =& 0 - f_i(\mathbf{x}) - g_i(\mathbf{x}) = -\dot{x}_i \geq 0 \end{aligned} \tag{19}$$

holds. On the other hand, if either $x_i^+ < x_{\max,i}^+$ or $b_i^+(\mathbf{x}^+, \mathbf{x}^-, y) < 0$ holds, then the following holds with probability at least $1 - n\delta$.

$$\begin{aligned} \dot{e}_i^+ =& f_i^+(\mathbf{x}^+, \mathbf{x}^-, y) - f_i(\mathbf{x}) + \hat{g}_i^+(\mathbf{x}^+) \\ & - \hat{g}_i^-(\mathbf{x}^-) - g_i(\mathbf{x}) + \zeta_{\mathbf{x},i}(\mathbf{x}^-, \mathbf{x}^+) \\ \geq& f_i^+(\mathbf{x}^+, \mathbf{x}^-, y) - f_i(\mathbf{x}) \geq 0, \end{aligned} \tag{20}$$

where the first and second inequalities are due to Lemma 1 and Assumption 4, respectively. Using the analogous procedure, the same can be shown for the negative error components. ∎

Lemma 3: Choose $\hat{\mathbf{g}}^+(\cdot)$ and $\hat{\mathbf{g}}^-(\cdot)$ as in Proposition 1. Then there exist matrices $\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \mathbf{N}_4 \in \mathbb{R}^{n \times n}$, such that for all $\mathbf{x}^+, \mathbf{x}, \mathbf{x}^- \in \mathbb{R}^n$ with $\mathbf{x}^- \leq \mathbf{x} \leq \mathbf{x}^+$,

$$\hat{\mathbf{g}}^+(\mathbf{x}^+) - \hat{\mathbf{g}}^-(\mathbf{x}^-) - \hat{\mathbf{g}}(\mathbf{x}) \leq \mathbf{N}_1\mathbf{e}^+ + \mathbf{N}_2\mathbf{e}^-, \tag{21}$$

$$\hat{\mathbf{g}}(\mathbf{x}) - \hat{\mathbf{g}}^+(\mathbf{x}^-) + \hat{\mathbf{g}}^-(\mathbf{x}^+) \leq \mathbf{N}_3\mathbf{e}^+ + \mathbf{N}_4\mathbf{e}^-. \tag{22}$$

Proof: See [18, Property 5]. ∎

Theorem 2: Consider the matrix $\mathbf{A}$, as discussed in Assumption 4 and the corresponding partition

$$\mathbf{A} := \begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{pmatrix}, \tag{23}$$

where $\mathbf{A}^{11}, \mathbf{A}^{12}, \mathbf{A}^{21}, \mathbf{A}^{22} \in \mathbb{R}^{n \times n}$. Moreover, choose the matrices $\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \mathbf{N}_4$ as in Lemma 3. Define the concatenated error dynamics $\boldsymbol{e} := \begin{pmatrix} \mathbf{e}^+ & \mathbf{e}^- \end{pmatrix}^{\mathrm{T}}$. If Assumptions 2 to 4 hold and the matrix

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}^{11} + \mathbf{N}_1 + \mathbf{F}^+ + \boldsymbol{\theta}^+\mathbf{C} & \mathbf{A}^{12} + \mathbf{N}_2 + \mathbf{F}^+ \\ \mathbf{A}^{21} + \mathbf{N}_3 + \mathbf{F}^- & \mathbf{A}^{22} + \mathbf{N}_4 + \mathbf{F}^- + \boldsymbol{\theta}^-\mathbf{C} \end{pmatrix} \tag{24}$$

is Hurwitz, then with probability of at least $1 - n\delta$,

$$\lim_{t \to \infty} \boldsymbol{e}(t) \leq -2\tilde{\mathbf{A}}^{-1} \begin{pmatrix} \boldsymbol{\zeta}_{\mathbf{x},\max} \\ \boldsymbol{\zeta}_{\mathbf{x},\max} \end{pmatrix}. \tag{25}$$

holds.

Proof: The following holds with probability $1 - n\delta$.

$$\begin{aligned} \dot{\boldsymbol{e}} =& \begin{pmatrix} \mathbf{o}^+(\mathbf{x}^+, \mathbf{x}^-, y) - \mathbf{f}(\mathbf{x}) - \mathbf{g}(\mathbf{x}) \\ \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}) - \mathbf{o}^-(\mathbf{x}^+, \mathbf{x}^-, y) \end{pmatrix} \\ \leq& \begin{pmatrix} \mathbf{b}^+(\mathbf{x}^+, \mathbf{x}^-, y) - \mathbf{f}(\mathbf{x}) - \mathbf{g}(\mathbf{x}) \\ \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}) - \mathbf{b}^-(\mathbf{x}^+, \mathbf{x}^-, y) \end{pmatrix} \\ =& \begin{pmatrix} \mathbf{f}^+(\mathbf{x}^+, \mathbf{x}^-, y) - \mathbf{f}(\mathbf{x}) + \hat{\mathbf{g}}^+(\mathbf{x}^+) - \hat{\mathbf{g}}^-(\mathbf{x}^-) - \mathbf{g}(\mathbf{x}) \\ \mathbf{f}(\mathbf{x}) - \mathbf{f}^-(\mathbf{x}^+, \mathbf{x}^-, y) + \mathbf{g}(\mathbf{x}) - \hat{\mathbf{g}}^+(\mathbf{x}^-) + \hat{\mathbf{g}}^-(\mathbf{x}^+) \\ \quad + \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}^-, \mathbf{x}^+) + \mathbf{F}^+(\mathbf{x}^+ - \mathbf{x}^-) + \boldsymbol{\theta}(y^+ - y) \\ \quad + \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}^-, \mathbf{x}^+) + \mathbf{F}^-(\mathbf{x}^+ - \mathbf{x}^-) - \boldsymbol{\theta}(y^- - y) \end{pmatrix} \\ \leq& \tilde{\mathbf{A}}\boldsymbol{e} + \begin{pmatrix} \mathbf{g}(\mathbf{x}) - \hat{\mathbf{g}}(\mathbf{x}) + \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}^-, \mathbf{x}^+) \\ \hat{\mathbf{g}}(\mathbf{x}) - \mathbf{g}(\mathbf{x}) + \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}^-, \mathbf{x}^+) \end{pmatrix} \\ \leq& \tilde{\mathbf{A}}\boldsymbol{e} + \begin{pmatrix} \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}, \mathbf{x}) + \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}^-, \mathbf{x}^+) \\ \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}, \mathbf{x}) + \boldsymbol{\zeta}_{\mathbf{x}}(\mathbf{x}^-, \mathbf{x}^+) \end{pmatrix} \\ \leq& \tilde{\mathbf{A}}\boldsymbol{e} + 2\begin{pmatrix} \boldsymbol{\zeta}_{\mathbf{x},\max} \\ \boldsymbol{\zeta}_{\mathbf{x},\max} \end{pmatrix}, \end{aligned} \tag{26}$$

where the first inequality results from $\mathbf{o}^+(\mathbf{x}^+, \mathbf{x}^-, y) \leq \mathbf{b}^+(\mathbf{x}^+, \mathbf{x}^-, y)$ together with $\mathbf{o}^-(\mathbf{x}^+, \mathbf{x}^-, y) \geq \mathbf{b}^-(\mathbf{x}^+, \mathbf{x}^-, y)$, and the second inequality is due to Lemma 3. The third inequality is due to Lemma 1. Since $\tilde{\mathbf{A}}$ is Hurwitz, the autonomous linear system

$$\dot{\mathbf{z}} = \tilde{\mathbf{A}}\mathbf{z} + 2\begin{pmatrix} \boldsymbol{\zeta}_{\mathbf{x},\max} \\ \boldsymbol{\zeta}_{\mathbf{x},\max} \end{pmatrix}, \qquad \mathbf{z}(0) = \mathbf{z}_0 \tag{27}$$

is stable for all $\mathbf{z}_0 \in \mathcal{R}$, and admits a single equilibrium point, which is given by $\mathbf{z}_e = -2\tilde{\mathbf{A}}^{-1}\begin{pmatrix} \boldsymbol{\zeta}_{\mathbf{x},\max} \\ \boldsymbol{\zeta}_{\mathbf{x},\max} \end{pmatrix}$. This in turn implies the desired result. ∎

Since the entries of $\boldsymbol{\zeta}_{\mathbf{x},\max}$ depend on the variance of the GPs used to model $\mathbf{g}(\cdot)$, this implies that the observer's ultimate error bound is made smaller by incorporating system measurements such that the variance of the GP models is decreased, as discussed in Section III. Moreover, the choice of $\theta^-, \theta^+, \mathbf{F}^-, \mathbf{F}^+$ also affects the error bound, as they influence the bounding matrix $\tilde{\mathbf{A}}$. This in turn can be employed to robustify the proposed approach in settings where the data points are not sufficient to achieve a low error bound $\boldsymbol{\zeta}_{\mathbf{x},\max}$.

V. Numerical Simulation

The observer design approach proposed in Section IV-B is employed to estimate the state of a dynamical system in a numerical simulation. The system employed is based on versions of Chua's circuit found in [19] and [20]. Chua's circuit consists of a three-dimensional nonlinear electric circuit that exhibits chaotic behavior, and is described using the following set of dimensionless differential equations.

$$\dot{x}_1 = \alpha(x_2 - h(x_1)), \tag{28}$$

$$\dot{x}_2 = x_1 - x_2 + x_3, \tag{29}$$

$$\dot{x}_3 = -\kappa x_2 - \eta x_3, \tag{30}$$

where $\alpha, \kappa, \eta, a, b \in \mathbb{R}$ are scalar parameters, and $h : \mathbb{R} \mapsto \mathbb{R}$ is the following nonlinear function.

$$h(s) = \frac{1}{6}s^3 - \frac{1}{6}s. \tag{31}$$

The system output is

$$y = x_1 + x_2 + x_3. \tag{32}$$

The model parameters are set to $\alpha = 11.85$, $\kappa = 14.9$, $\eta = 0.29$. We assume to have the following erroneous model of the system dynamics.

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \alpha\,(0.2x_1 + 1.05x_2) \\ x_1 - 0.8x_2 + x_3 \\ -0.9\kappa x_2 - 0.8\eta x_3 \end{pmatrix}, \tag{33}$$

hence the unknown component is

$$\mathbf{g}(\mathbf{x}) = \begin{pmatrix} \alpha(-0.05x_2 - h(x_1)) \\ -0.2x_2 \\ -0.1\kappa x_2 - 0.\eta x_3 \end{pmatrix}. \tag{34}$$

The function $\mathbf{g}(\cdot)$ satisfies Assumption 1, hence the interval observer design approach discussed in Section IV-B is employed. For the known component $\mathbf{f}(\cdot)$, the following observer is employed, is derived using the approach given by [18].

$$\Sigma_{\text{red}} = \begin{cases} \dot{\mathbf{x}}^+ = \mathbf{f}(\mathbf{x}^+) + \tilde{\boldsymbol{\theta}}^+(y^+ - y) + \tilde{\mathbf{F}}^+(\mathbf{x}^+ - \mathbf{x}^-) \\ \dot{\mathbf{x}}^- = \mathbf{f}(\mathbf{x}^-) - \tilde{\boldsymbol{\theta}}^-(y - y^-) - \tilde{\mathbf{F}}^+(\mathbf{x}^+ - \mathbf{x}^-), \end{cases}, \tag{35}$$

where

$$\tilde{\boldsymbol{\theta}}^+ = \tilde{\boldsymbol{\theta}}^- = \begin{pmatrix} -\alpha & -1 & 0 \end{pmatrix}^{\mathrm{T}}, \tag{36}$$

$$\tilde{\mathbf{F}}^+ = \tilde{\mathbf{F}}^- = \begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & 0 \\ 0 & 0.9\kappa & 0 \end{pmatrix}. \tag{37}$$

Using the procedure described in [18], it can be shown that the error of the reduced system interval observer Σ_{red} corresponds to an asymptotically stable system. Each entry of the unknown function $\mathbf{g}(\cdot)$ is modeled using a GP, as discussed in Section III. However, we do not employ inputs from the full state space to model each estimate $\hat{g}_i(\mathbf{x})$. Instead, for an $i \in \{1, 2, 3\}$, we only employ the components x_1, x_2, x_3 of the state space that directly influence $g_i(\mathbf{x})$. For each GP model, $N = 64$ noisy measurements are taken from a simulation of Chua's circuit. The corresponding measurement noise ϵ is randomly sampled from a uniform distribution, which is bounded by $\epsilon_{\max} = 0.01$. For the GP models, squared-exponential kernels are employed, which encode infinitely differentiable functions. The corresponding parameters are obtained by optimizing the respective marginal likelihoods [11]. The values suggested by Theorem 1 are often too conservative in practice, and can be divided by a constant factor [17]. Here we first estimate the multiplicative factors for the model uncertainty $\beta_N^1,\ \beta_N^2,\ \beta_N^3$ via a guess-and-doubling approach using the training data. This roughly corresponds to estimating B_{g_i}, where $\|g_i\|_{k_i} < B_{g_i}$, and setting $\beta_N^i = B_{g_i}$. We then scale down the resulting β_N^i by a factor of 10. The bounding vectors are set to $\mathbf{x}^{\max} = \begin{pmatrix} 2.5 & 0.5 & 5 \end{pmatrix}^{\mathrm{T}}$ and $\mathbf{x}^{\min} = \begin{pmatrix} -2.5 & -0.5 & -5 \end{pmatrix}^{\mathrm{T}}$. If $\mathbf{g}(\cdot)$ and $\hat{\mathbf{g}}(\cdot)$ were identical, then, since $\nabla \mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, holds, a partition $\hat{\mathbf{g}}^+(\cdot)$ and $\hat{\mathbf{g}}^-(\cdot)$ could be obtained by choosing $\hat{g}_1^+ = 0,\ \hat{g}_2^+ = 0,\ \hat{g}_3^+ = 0,$ $\hat{g}_1^- = -\hat{g}_1,\ \hat{g}_2^- = -\hat{g}_2,$ and $\hat{g}_3^- = -\hat{g}_3$. However, the GP models only reproduce $\mathbf{g}(\cdot)$ approximately, and do not necessarily exhibit monotonous behavior. In order to compensate this, the scalar values

$$L_{1,i} = \max\left\{0, \max_{\mathbf{z}\in\mathcal{R}} \frac{\partial g_1}{\partial x_i}(\mathbf{z})\right\},\ i = 1, 2, 3 \tag{38}$$

$$L_{2,i} = \max\left\{0, \max_{\mathbf{z}\in\mathcal{R}} \frac{\partial g_2}{\partial x_i}(\mathbf{z})\right\},\ i = 1, 2, 3 \tag{39}$$

$$L_{3,i} = \max\left\{0, \max_{\mathbf{z}\in\mathcal{R}} \frac{\partial g_3}{\partial x_i}(\mathbf{z})\right\},\ i = 1, 2, 3 \tag{40}$$

are computed, which are employed to design the partitions

$$\hat{g}_j^+(\mathbf{z}) = \sum_{i=1}^{3} L_{j,i}\mathbf{z},\ j = 1, 3$$

$$\hat{g}_j^-(\mathbf{z}) = \hat{g}_j(\mathbf{z}) + \sum_{i=1}^{3} L_{j,i}\mathbf{z},\ j = 1, 3$$

It is easy to show that $\hat{\mathbf{g}}^+(\cdot)$ and $\hat{\mathbf{g}}^-(\cdot)$ are monotonously increasing and that $\hat{\mathbf{g}}^+(\cdot) - \hat{\mathbf{g}}^-(\cdot) = \hat{\mathbf{g}}(\cdot)$ holds. The gain matrices are set to

$$\boldsymbol{\theta}^+ = \boldsymbol{\theta}^- = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}^{\mathrm{T}}, \tag{41}$$

$$\mathbf{F}^+ = \mathbf{F}^- = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0.1\kappa & 0 \end{pmatrix}. \tag{42}$$

The corrective vectors $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^-$ are set to zero in the present case, since the interval observer for the reduced system already contains a corrective term. The system initial conditions are $\mathbf{x}_0 = \begin{pmatrix} 0.1 & 0.2 & 0.05 \end{pmatrix}^{\mathrm{T}}$, whereas the observer initial conditions are set to $\mathbf{x}_0^+ = \begin{pmatrix} 0.5 & 0.5 & 0.5 \end{pmatrix}^{\mathrm{T}}$ and $\mathbf{x}_0^- = \begin{pmatrix} -0.5 & -0.5 & -0.5 \end{pmatrix}^{\mathrm{T}}$. The simulation results and respective errors are presented in Figure 1. The interval observer provides a boundary for all the state's during the whole simulation. In order to illustrate the stochastic nature of the interval observer, a Monte Carlo simulation consisting of 50 runs is carried out, where each each run is identical to the simulation described above except for the partition

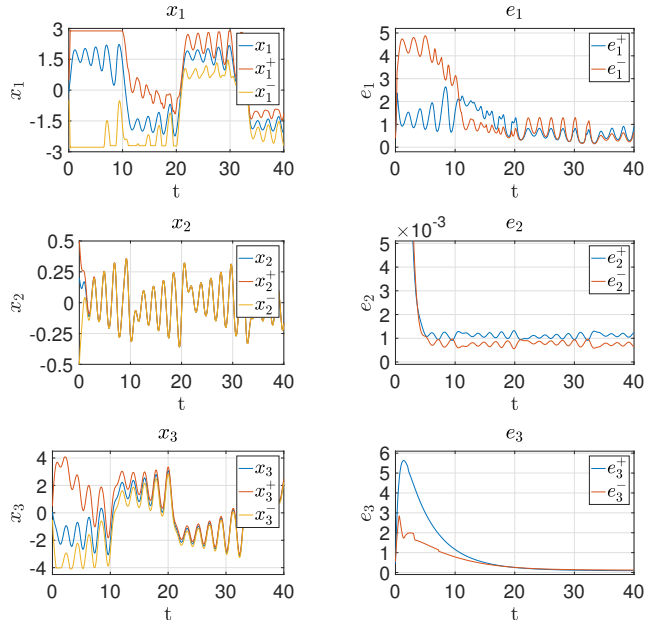

Fig. 1. Simulation results of dimensionless Chua's system using the proposed interval observer.

constants given by Equations (38) to (40), and the simulation time, which is set to 80. The measurement noise is sampled from the same distribution as in the first case. The interval observer errors resulting from all 50 runs are presented in Figure 2. All interval observers generate different error

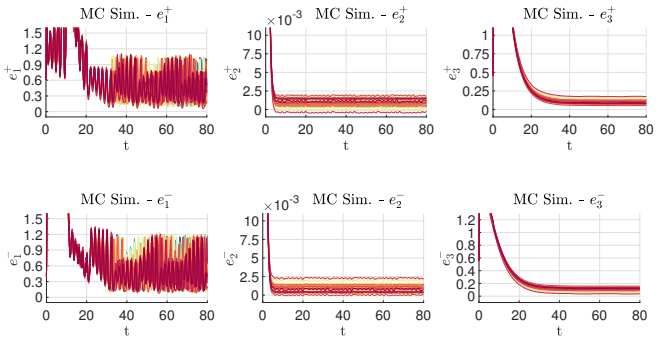

Fig. 2. Interval observer errors of Monte Carlo simulation consisting of 50 runs.

dynamics, which is due to the probabilistic nature of the measurement noise used to compute the GP models. All observers yield positive errors at all times, except for five observers, which generate errors e_2^- and e_2^+ that are at times negative. The corresponding lower bound for the incorrect interval observer errors is $e_2^-, e_2^+ \geq -5^{-4}$ for all $t \in [0, 80]$. This reflects the result of Theorem 2, which indicates that the bounds provided by each interval observer are correct with a predetermined probability.

VI. Conclusion

An interval observer for Gaussian process models is presented. The interval observer is designed by employing an existing observer for the known component of the system model, which is extended by and observer for the unknown component. The proposed interval observer employs a decomposition of the Gaussian process model into two monotonous components, as well as the Gaussian process model uncertainty to generate upper and lower bounds for the state estimation. The interval observer is shown to provide correct error bounds with high probability. The corresponding failure probability is made smaller by increasing the number of data points employed.

References

[1] E. D. Sontag, *Mathematical control theory: deterministic finite dimensional systems*. Springer, 2013, vol. 6.

[2] M. Zeitz, "The extended Luenberger observer for nonlinear systems," *Systems & Control Letters*, vol. 9, no. 2, pp. 149–156, 1987.

[3] N. Kazantzis and C. Kravaris, "Nonlinear observer design using Lyapunov's auxiliary theorem," in *Proceedings of the 36th IEEE Conference on Decision and Control*, vol. 5, 1997, pp. 4802–4807.

[4] J.-L. Gouzé, A. Rapaport, and M. Z. Hadj-Sadok, "Interval observers for uncertain biological systems," *Ecological modelling*, vol. 133, no. 1-2, pp. 45–56, 2000.

[5] F. Mazenc, S.-I. Niculescu, and O. Bernard, "Exponentially stable interval observers for linear systems with delay," *SIAM Journal on Control and Optimization*, vol. 50, no. 1, pp. 286–305, 2012.

[6] T. Raïssi, D. Efimov, and A. Zolghadri, "Interval state estimation for a class of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 260–265, 2012.

[7] L. Ljung, "System identification," in *Signal analysis and prediction*. Springer, 1998, pp. 163–173.

[8] J. Umlauft, T. Beckers, M. Kimmel, and S. Hirche, "Feedback linearization using Gaussian processes," in *IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 5249–5255.

[9] T. Beckers, J. Umlauft, and S. Hirche, "Stable model-based control with Gaussian process regression for robot manipulators," *World Congress of the International Federation of Automatic Control (IFAC)*, vol. 50, no. 1, pp. 3877–3884, 2017.

[10] A. Capone and S. Hirche, "Backstepping for partially unknown nonlinear systems using Gaussian processes," *IEEE Control Systems Letters*, vol. 3, no. 2, pp. 416–421, April 2019.

[11] C. E. Rasmussen and C. K. Williams, "Gaussian processes for machine learning. 2006," *The MIT Press, Cambridge, MA, USA*, vol. 38, pp. 715–719, 2006.

[12] M. P. Deisenroth, M. F. Huber, and U. D. Hanebeck, "Analytic moment-based Gaussian process filtering," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 225–232.

[13] J. Ko and D. Fox, "Gp-bayesfilters: Bayesian filtering using Gaussian process prediction and observation models," *Autonomous Robots*, vol. 27, no. 1, pp. 75–90, 2009.

[14] M. P. Deisenroth, R. D. Turner, M. F. Huber, U. D. Hanebeck, and C. E. Rasmussen, "Robust filtering and smoothing with Gaussian processes," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1865–1871, 2012.

[15] G. Wahba, *Spline models for observational data*. Society for Industrial and Applied Mathematics, 1990, vol. 59.

[16] S. R. Chowdhury and A. Gopalan, "On kernelized multi-armed bandits," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2017, pp. 844–853.

[17] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.

[18] M. Moisan and O. Bernard, "Robust interval observers for global Lipschitz uncertain chaotic systems," *Systems & Control Letters*, vol. 59, no. 11, pp. 687–694, 2010.

[19] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential equations, dynamical systems, and an introduction to chaos*. Academic press, 2012.

[20] G. Leonov, D. Ponomarenko, V. Smirnova, and L. Chua, "Global stability and instability of canonical Chua's circuits," in *Chua's circuit: a paradigm for chaos*. World Scientific, 1993, pp. 725–739.