Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 5127–5132

Kernel-Based Identification of Incrementally Input-to-State Stable Nonlinear Systems

Matteo Scandella * **Michelangelo Bin** *,** **Thomas Parisini** *,***,****

* *Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom.*
** *Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione, Università di Bologna, Bologna, Italy.*
*** *KIOS Research and Innovation Center of Excellence, University of Cyprus, Cyprus.*
**** *Department of Engineering and Architecture, University of Trieste, Trieste, Italy.*

Abstract: Methods based on Reproducing Kernel Hilbert Spaces (RKHS) have proven to be a valuable tool for the identification of linear time-invariant systems in both discrete- and continuous-time. In particular, unlike most other techniques, they enable to systematically confer *a priori* desirable properties, such as stability, on the estimated models. However, existing RKHS methods mainly target impulse responses and, hence, do not extend well to the context of nonlinear systems. In this work, we propose a novel RKHS-based methodology for the identification of discrete-time nonlinear systems guaranteeing that the identified system is incrementally input-to-state stable (δISS). We model the identified system using a predictor function that, given past input and output samples, yields the output prediction at the next time instant. The predictor is selected from an RKHS by solving a constrained optimization problem that guarantees its δISS properties. The proposed approach is validated via numerical simulations.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: Nonlinear system identification, Incremental input-to-state stability, Reproducing kernel Hilbert spaces, Kernel-based regularization, Gaussian process regression

1. INTRODUCTION

Methods based on *Reproducing Kernel Hilbert Spaces* (RKHS) are appealing for system identification, since they allow to systematically enforce some desired properties on the identified models by suitably "shaping" the reproducing kernel (Dinuzzo, 2015; Formentin and Chiuso, 2021). For example, in the context of impulse-response identification of linear systems, one can find conditions on the reproducing kernel ensuring that every identified model is bounded-input-bounded-output (BIBO) stable, without the need of *a priori* constraining the order of the model (Pillonetto and De Nicolao, 2010; Pillonetto et al., 2014; Scandella et al., 2022). The ability of imposing on the identified models some structural conditions, such as input-to-state (ISS) stability or smoothness, is of crucial importance in applications, as it enables to embed prior knowledge on the identified models and/or to constraint them to desirable classes.

Kernel identification of linear systems is well-developed. Instead, imposing stability guarantees on identified nonlinear models is considerably more challenging, and, essentially, it is still an open problem (Pillonetto et al., 2011; Pillonetto, 2018). The difficulty of handling the much broader framework of nonlinear models is exacerbated by the many possible notions and variations of "stability" that one may seek in nonlinear systems. Indeed, in the linear case, BIBO stability, asymptotic stability, uniform asymptotic stability, exponential stability, and incremental (input-to-state) stability are all equivalent properties. In the nonlinear case, instead, they are not. Moreover, the relationship between BIBO stability of a linear system and its impulse response does not extend to nonlinear systems. As a consequence, nonlinear identification methods cannot be based on the impulse response characterization. Instead, they must either target input-output operators (necessarily defined between infinite-dimensional spaces) such as in (van Waarde and Sepulchre, 2022), or directly prediction models (predictors for short) as in (Pillonetto et al., 2011; Bhujwalla et al., 2016; Bai et al., 2014), leading to state-space models.

In this article, we develop an RKHS-based methodology to identify incrementally input-to-state (δISS) predictors for nonlinear systems. In particular, we guarantee that the

⋆ This work has been supported by European Union's Horizon 2020 research and innovation programme under grant agreement no. 739551 (KIOS CoE) and by the Italian Ministry for Research in the framework of the 2017 Program for Research Projects of National Interest (PRIN), Grant no. 2017YKXYXJ.

2405-8963 Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2023.10.1297

identified predictor is Lipschitz by taking advantage of the conditions on the reproducing kernel recently proposed by van Waarde and Sepulchre (2022). Then, we devise a technique to select the hyperparameters in such a way to ensure that the predictor is δISS. Specifically, we propose a constrained version of the well-known marginal likelihood maximization procedure (MacKay, 1992; Mazzoleni et al., 2022).

Incremental stability is quite a restrictive stability condition within the large family of possible nonlinear stability notions. However, incremental stability represents rather a natural extension of linear stability, covering many systems of interest (Lohmiller and Slotine, 1998; Angeli, 2002; Sepulchre et al., 2022; van Waarde and Sepulchre, 2022), and incrementally stable systems share many similarities with *contractive* and *convergent* systems (Rüffer et al., 2013). It therefore represents a natural subject for a first step toward stable nonlinear predictor identification.

The proposed approach is inspired by the work of van Waarde and Sepulchre (2022), where the authors consider the problem of identifying Lipschitz operators between Hilbert spaces. Unlike van Waarde and Sepulchre (2022), we directly identify a predictor for the system's output given the past outputs and inputs, which is a function between Euclidean spaces, and not an input-output model between infinite-dimensional spaces. Moreover, the proposed method guarantees that the identified predictor is δISS with respect to both the input signal and measurements noise by using a procedure that does not require manual tuning of the kernel's parameters. Finally, the proposed method is more suitable for prediction of dynamical systems, since it can cope with different initial conditions.

The article is organized as follows. Section 2 reports the problem formulation. Section 3 recalls the *Gaussian Process Regression* approach for nonlinear system identification. The proposed methodology is then explained in Section 4, and it is supported by numerical simulations in Section 5. Finally, Section 6 presents some concluding remarks.

2. PROBLEM STATEMENT

We consider a discrete-time system Ψ mapping *input* sequences $u \in \mathbb{R}^{\mathbb{N}}$ to *output* sequences $y \in \mathbb{R}^{\mathbb{N}}$. Formally, $\Psi \subset \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ and $(u, y) \in \Psi$ if and only if u is mapped to y. We assume to have at our disposal $n \in \mathbb{N}_{>0}$ measurements

$$\mathcal{D} = \Big\{ (\bar{u}_i, \bar{y}_i) \,\Big|\, 1 \le i \le n \Big\}$$

taken from an experiment executed with the system Ψ. We aim at devising an algorithm that, given the dataset $\mathcal{D}$ and an arbitrary $m \in \mathbb{N}$, provides a function $f : \mathbb{R}^m \times \mathbb{R}^{m+1} \to \mathbb{R}$ such that

$$\hat{y}_t = f\big(y_{t-m}, \dots, y_{t-1}, u_{t-m}, \dots, u_{t-1}, u_t\big) \tag{1}$$

is a "good approximation" of y_t for every $(u, y) \in \Psi$ and every $t \in \mathbb{N}_{\ge m}$. We refer to m as the *model order* and to f as the *predictor*. Moreover, we restrict our search of f to the set of functions ensuring a *stable* predictor in the sense explained hereafter. We can devise a state-space realization of (1) as follows

$$x_{t+1} = Ax_t + Gy_t + Hu_t \tag{2a}$$
$$\hat{y}_t = f(x_t, u_t) \tag{2b}$$

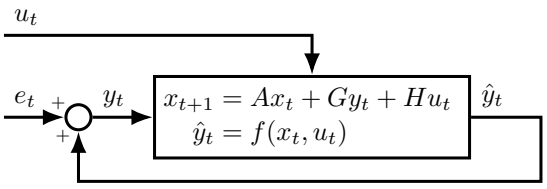

Fig. 1. Block diagram of the closed-loop system (3).

in which

$$A \coloneqq \begin{bmatrix} S & 0_{m\times m} \\ 0_{m\times m} & S \end{bmatrix}, \quad G \coloneqq \begin{bmatrix} B \\ 0_{m\times 1} \end{bmatrix}, \quad H \coloneqq \begin{bmatrix} 0_{m\times 1} \\ B \end{bmatrix},$$

and

$$S \coloneqq \begin{bmatrix} 0_{(m-1)\times 1} & I_{m-1} \\ 0 & 0_{1\times(m-1)} \end{bmatrix}, \qquad B \coloneqq \begin{bmatrix} 0_{(m-1)\times 1} \\ 1 \end{bmatrix}.$$

By letting $y_t = \hat{y}_t + e_t$ in (2a), where $e \in \mathbb{R}^{\mathbb{N}}$ is the *prediction error*, we obtain the "closed-loop" system

$$x_{t+1} = Ax_t + Gf(x_t, u_t) + Ge_t + Hu_t. \tag{3}$$

System (3), whose block-representation is shown in Fig. 1, is a system with inputs u and e and state x representing the predictor (1) operating on past measurements. In qualitative and informal terms, we say that the predictor (1) is "stable" *in some sense*, if the system (3) is stable *in the same sense*. As anticipated in the Introduction, we focus on incremental input-to-state stability, formally defined in Definition 1.

Definition 1. System (3) is said to be δISS if there exist a class-$\mathcal{KL}$ function ω and two class-$\mathcal{K}_\infty$ functions κ_{u} and κ_{e} such that every two solution triplets (x^a, u^a, e^a) and (x^b, u^b, e^b) of (3) satisfy

$$\|x_t^a - x_t^b\|_2 \le \omega\left(\|x_0^a - x_0^b\|_2, t\right) + \kappa_{\mathrm{u}}\left(\sup_{s\in\mathbb{N}_{\le t}} \left|u_s^a - u_s^b\right|\right) + \kappa_{\mathrm{e}}\left(\sup_{s\in\mathbb{N}_{\le t}} \left|e_s^a - e_s^b\right|\right)$$

for all $t \in \mathbb{N}$.

With Definition 1 in mind, the problem we consider in this article can be stated as follows.

Problem 1. Given the dataset $\mathcal{D}$ and the model order m, devise a procedure to find a function $f : \mathbb{R}^m \times \mathbb{R}^{m+1} \to \mathbb{R}$ such that **(i)** for every $(u, y) \in \Psi$ and every $t \in \mathbb{N}_{\ge m}$ the prediction error $e_t = y_t - \hat{y}_t$ is "sufficiently small", and **(ii)** the resulting predictor is δISS.

The term "sufficiently small" in Point **(i)** is formalized in Section 3 in terms of Bayesian inference.

3. GAUSSIAN PROCESS REGRESSION

Gaussian Process Regression (GPR) is an extension to the nonlinear case of linear regression. Also known as *Kriging*, GPR was introduced for geostatistical simulations by Matheron (1963), and it is now a well-established nonlinear regression technique (Rasmussen and Williams, 2006). GPR has also been employed for nonlinear system identification by Pillonetto et al. (2011) to estimate the predictor function.

GPR is a Bayesian methodology whose goal is to find the distribution of the predictor given the available measurements. To use GPR, it is necessary to make statistical

assumptions on the model variables. Similarly to the linear case, in this article we make the simplifying – yet convenient – assumption that the prediction error e is a 0-mean Gaussian white noise with variance $\beta^2 \in \mathbb{R}$. Namely, for all $i, j \in \mathbb{N}$ such that $i \neq j$, e_i is independent from e_j, and

$$e_t \sim \mathcal{N}(0, \beta^2), \qquad \forall t \in \mathbb{N},$$

where, given $\iota \in \mathbb{N}_{>0}$, $\mathcal{N}(\mu, \Sigma)$ is the ι-dimensional normal distribution with mean $\mu \in \mathbb{R}^\iota$ and variance $\Sigma \in \mathbb{R}^{\iota \times \iota}$. Furthermore, we assume that the predictor function f is distributed according to a Gaussian Process (Rasmussen and Williams, 2006, Def. 2.1) with 0-mean function and with a covariance function $k_\eta : \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1} \to \mathbb{R}$ parametrized by a parameter $\eta \in \mathbb{R}^{n_\eta}$, with $n_\eta \in \mathbb{N}$ to be set later on. Formally,

$$f \sim \mathcal{GP}(0, k_\eta). \tag{4}$$

Finally, as it is usually done, we assume that the function f and the prediction error e are independent.

Given the dataset $\mathcal{D}$ and a $t \in \{m+1, \dots, n\}$, we introduce the column vector

$$\bar{z}_t \coloneqq \left[\bar{y}_{t-m}, \dots, \bar{y}_{t-1}, \bar{u}_{t-m}, \dots, \bar{u}_{t-1}, \bar{u}_t\right]^\top \in \mathbb{R}^{2m+1}.$$

Additionally, we also let

$$\begin{aligned} \bar{y} &\coloneqq \left[\bar{y}_{m+1}, \dots, \bar{y}_n\right]^\top \in \mathbb{R}^r, \\ \bar{f} &\coloneqq \left[f(\bar{z}_{m+1}), \dots, f(\bar{z}_n)\right]^\top \in \mathbb{R}^r, \\ \bar{e} &\coloneqq \bar{y} - \bar{f} \in \mathbb{R}^r, \end{aligned}$$

where $r = n - m$. Then, using the aforementioned statistical assumptions and the fact that the sum of two normal variables is a normal variable, we obtain

$$\bar{e} \sim \mathcal{N}(0_{r\times 1}, \beta^2 I_r), \tag{5a}$$
$$\bar{f} \sim \mathcal{N}(0_{r\times 1}, K_\eta), \tag{5b}$$
$$\bar{y} \sim \mathcal{N}(0_{r\times 1}, K_\eta + \beta^2 I_r), \tag{5c}$$

where $K_\eta \in \mathbb{R}^{r\times r}$ is a positive semidefinite symmetric matrix, called the *kernel matrix*, whose (i,j)th element is $k_\eta(\bar{z}_{m+i}, \bar{z}_{m+j})$.

Now, consider a vector $z^* \in \mathbb{R}^{2m+1}$. Using the definition of Gaussian Process, we obtain

$$\begin{bmatrix} \bar{y} \\ f(z^*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{r\times 1} \\ 0 \end{bmatrix}, \begin{bmatrix} K_\eta + \beta^2 I_r & k_\eta^*(z^*) \\ k_\eta^*(z^*)^\top & k_\eta(z^*, z^*) \end{bmatrix}\right),$$

in which $k_\eta^*(z^*) \coloneqq \left[k_\eta(z^*, \bar{z}_{m+1}), \dots, k_\eta(z^*, \bar{z}_n)\right]^\top \in \mathbb{R}^r$. Therefore, using the properties of the Normal distribution, we obtain

$$f(z^*) \,|\, \bar{y} \sim \mathcal{N}(\mu_f(z^*), \sigma_f^2(z^*)), \tag{6}$$

where

$$\begin{aligned} \mu_f(z^*) &= k_\eta^*(z^*)^\top (K_\eta + \beta^2 I_r)^{-1} \bar{y}, \\ \sigma_f^2(z^*) &= k_\eta(z^*, z^*) - k_\eta^*(z^*)^\top (K_\eta + \beta^2 I_r)^{-1} k_\eta^*(z^*). \end{aligned}$$

A common choice for the estimator is the expected value of (6), referred to as $f_{GP}(z)$, and defined as

$$f_{GP}(z) \coloneqq \mu_f(z) = k_\eta^*(z)^\top c = \sum_{i=m+1}^{n} c_i k_\eta(z, \bar{z}_i), \tag{7}$$

in which $c \coloneqq (K_\eta + \beta^2 I_r)^{-1} \bar{y} = \left[c_{m+1}, \dots, c_n\right]^\top \in \mathbb{R}^r$.

In order to use Predictor (7), it is necessary to select the values of **(i)** the variance of the prediction error β^2, and **(ii)** the kernel parameter η. A common way to select them is by maximizing the likelihood to obtain the measurements $\bar{y}$ given a certain value of (β^2, η) (MacKay, 1992; Rasmussen and Williams, 2006). In particular, from (5c), we notice that

$$\bar{y} \,|\, (\beta^2, \eta) \sim \mathcal{N}(0, K_\eta + \beta^2 I_r).$$

Therefore, (β^2, η) can be obtained by solving the optimization problem

$$\operatorname*{argmin}_{\beta^2 > 0,\, \eta \in \mathbb{R}^{n_\eta}} \quad \bar{y}^\top c + \log\det(K_\eta + \beta^2 I_r), \tag{8}$$

where $\det(\cdot)$ is the determinant operator and, for computational convenience, we equivalently minimize the negative log-likelihood instead of maximizing the likelihood.

The GPR methodology applies for every symmetric positive semidefinite kernel k_η. Nevertheless, in the literature it is possible to find kernels specialized for the task of identifying predictors for nonlinear dynamical systems; see, e.g., (Pillonetto et al., 2011; Mazzoleni et al., 2020). However, it is worth noticing that none of these works are able to guarantee stability of the identified predictor. In the next section, we will tackle this problem by guaranteeing stability of the predictor according to Definition 1.

4. SOLUTION METHODOLOGY

In this section, we provide an estimation method for the predictor function f guaranteeing that all the identified predictors are δISS. The proposed method builds on the GPR estimation methodology described in Section 3. In particular, we first derive sufficient conditions on the function f guaranteeing that system (3) is δISS (Proposition 1 hereafter). Then, we translate such conditions into properties of the estimated predictor (7). Finally, we exploit these conditions within the GPR procedure.

Proposition 1. Suppose that there exist a class-$\mathcal{K}$ function ϱ and a $\mu \in (0, 1)$ such that, for all $x^a, x^b \in \mathbb{R}^{2m}$ and $u^a, u^b \in \mathbb{R}$, the following bound holds

$$\left|f(x^a, u^a) - f(x^b, u^b)\right|^2 \leq \frac{\mu}{m}\|x^a - x^b\|_2^2 + \varrho(|u^a - u^b|).$$

Then, system (3) is δISS. In particular, Definition 1 applies with

$$\begin{aligned} \omega(s,t) &\coloneqq \sqrt{m}\alpha^{\frac{t}{2}} s, \\ \kappa_{\mathrm{u}}(s) &\coloneqq \sqrt{\frac{m(2-\mu)\varrho(s) + ms^2}{1-\alpha}}, \\ \kappa_{\mathrm{e}}(s) &\coloneqq s\sqrt{\frac{m(2-\mu)}{(1-\alpha)(1-\mu)}}, \end{aligned}$$

where

$$\alpha = 1 - \frac{(\mu - 1)^2}{m}.$$

Notice that the covariance function k_η of the Gaussian Process (4) is a symmetric positive semidefinite function. Hence, there exists an RKHS $\mathcal{H}(k_\eta)$ with k_η as a reproducing kernel (see Aronszajn (1950)) such that the estimator f_{GP} defined in (7) is an element of $\mathcal{H}(k_\eta)$. Therefore, we can enforce δISS of the estimator by selecting a kernel k_η guaranteeing that the functions inside $\mathcal{H}(k_\eta)$ satisfy the assumptions of Proposition 1. In the following, for simplicity, we restrict the focus on the case where $\varrho(s) = \mu m^{-1} s^2$. In particular, we aim to enforce the following stronger Lipschitz condition

$$|f(a)-f(b)|^2 \le \lambda \|a-b\|_2^2, \qquad \forall a,b \in \mathbb{R}^{2m+1}, \tag{9}$$

with $\lambda := \mu m^{-1} < m^{-1}$. Clearly, the only RKHS whose elements all satisfy (9) is the one that contains only the zero function. Indeed, if a non-zero function $f \in \mathcal{H}(k_\eta)$ satisfies (9), then there exists $h \in \mathbb{R}$ such that $hf \in \mathcal{H}(k_\eta)$ does not satisfy (9). Therefore, by following van Waarde and Sepulchre (2022), we proceed as follows: **(i)** first, we define a class of RKHS containing only functions that satisfy (9) with a generic $\lambda \in \mathbb{R}_{>0}$; **(ii)** then, we derive a condition on the kernel parameter η and the noise variance β^2 guaranteeing that $\lambda < m^{-1}$ holds for all identified functions.

The first step is obtained with the following proposition by adapting the work of van Waarde and Sepulchre (2022, Thm. 8) to the setting considered in this paper.

Proposition 2. Let $k_\eta : \mathbb{R}^{2m+1} \times \mathbb{R}^{2m+1} \to \mathbb{R}$ be a symmetric positive semidefinite kernel such that

$$k_\eta(a,a) - 2k_\eta(a,b) + k_\eta(b,b) \le \|a-b\|_2^2, \quad \forall a,b \in \mathbb{R}^{2m+1}. \tag{10}$$

Let $\mathcal{H}(k_\eta)$ be the RKHS defined by k_η, and let $\|\cdot\|_{k_\eta}$ be the norm on $\mathcal{H}(k_\eta)$. Then, every function $f \in \mathcal{H}(k_\eta)$ satisfies condition (9) with $\lambda = \|f\|_{k_\eta}^2$.

In view of Proposition 2, a function $f \in \mathcal{H}(k_\eta)$ satisfies (9) if k_η satisfies (10) and $\|f\|_{k_\eta}^2 < m^{-1}$. Therefore, to guarantee that an identified predictor f_{GP} of the form (7) is δISS, we need to select a parameter η ensuring that k_η satisfies (10) and, at the same time, that the squared norm of f_{GP} is smaller than m^{-1}. Once a set of feasible parameters η ensuring that k_η satisfies (10) is found, we can guarantee $\|f\|_{k_\eta}^2 < m^{-1}$ by suitably constraining the optimization problem (8). To this aim, notice that

$$\|f_{GP}\|_{k_\eta}^2 = \sum_{i=m+1}^{n} \sum_{j=m+1}^{n} c_i c_j \langle r_{\bar z_i}, r_{\bar z_j} \rangle_{k_\eta} = c^\top K_\eta c,$$

where $\langle \cdot,\cdot \rangle_{k_\eta}$ is the inner product of $\mathcal{H}(k_\eta)$, $r_a = k_\eta(a,\cdot) \in \mathcal{H}(k_\eta)$ is the *representer* function of $a \in \mathbb{R}^{2m+1}$, and where we used the definition of K_η and the fact that $k_\eta(a,b) = \langle r_a, r_b \rangle_{k_\eta}$ by construction. Then, the optimization problem (8) can be modified as follows

$$\begin{cases} \underset{\beta^2>0,\ \eta \in \mathbb{R}^{n_\eta}}{\operatorname{argmin}} \ \bar y^\top c + \log\det(K_\eta + \beta^2 I_r), \\ \text{subj. to} \ \ \beta^2 \ge \varepsilon,\ \eta \in \Omega_k,\ mc^\top K_\eta c \le \psi, \end{cases} \tag{11}$$

where $\Omega_k \subseteq \mathbb{R}^{n_\eta}$ is the set of all the values of η such that the condition (10) is satisfied, while $\varepsilon \in (0,\infty)$ and $\psi \in (0,1)$ are arbitrary constants introduced to make sure the feasibility set is closed. Ideally, ε is as small as possible, and ψ as large as possible. In particular, smaller values of ψ lead to more contractive estimated predictors but, at the same time, narrow the search space.

Now, we focus on Condition (10) and, in particular, on how to select a kernel guaranteeing that $\Omega_k \neq \varnothing$. A kernel ensuring that $\Omega_k \neq \varnothing$ is called *viable*. Following the proofs reported by van Waarde and Sepulchre (2022), in Proposition 3 below we show that the widely used Gaussian kernel is viable. More interestingly, in Proposition 4, we show that the kernel proposed in Pillonetto et al. (2011) for the identification of nonlinear predictors is viable as well.

Proposition 3. Let $\eta = (\tau,\gamma) \in \mathbb{R}^2_{>0}$ and k_η the Gaussian kernel defined by

$$k_\eta(a,b) = \tau \exp\big(-\gamma \|a-b\|_2^2\big), \tag{12}$$

for all $a,b \in \mathbb{R}^{2m+1}$. Then $\Omega_k = \{(\tau,\gamma) \in \mathbb{R}^2_{>0} \,|\, 2\tau\gamma \le 1\}$.

Proposition 4. Let $\eta = (\tau,\xi,\gamma,p) \in \mathbb{R}^3_{>0} \times \mathbb{N}_{\le m}$ and k_η the kernel such that, given

$$z^a = \big[y^a_1,\dots,y^a_m,u^a_1,\dots,u^a_m,u^a_{m+1}\big]^\top \in \mathbb{R}^{2m+1},$$
$$z^b = \big[y^b_1,\dots,y^b_m,u^b_1,\dots,u^b_m,u^b_{m+1}\big]^\top \in \mathbb{R}^{2m+1},$$

we have

$$k_\eta(z^a,z^b) = \tau \sum_{t=1}^{m-p+1} \exp(-\xi t - \gamma \|z^a_t - z^b_t\|_2^2), \tag{13}$$

where

$$z^a_t = \big[y^a_t,\dots,y^a_{t+p-1},u^a_t,\dots,u^a_{t+p-1}\big]^\top \in \mathbb{R}^{2p},$$
$$z^b_t = \big[y^b_t,\dots,y^b_{t+p-1},u^b_t,\dots,u^b_{t+p-1}\big]^\top \in \mathbb{R}^{2p}.$$

Then

$$\Omega_k \supseteq \big\{(\tau,\xi,\gamma,p) \in \mathbb{R}^3_{>0} \times \mathbb{N}_{\le m} \,\big|\, 2\tau\gamma \le 1, \ 2e^{-\xi} - e^{-\xi(m-p+2)} \le 1\big\}.$$

Kernel (13) is designed so as older measurements weight less than more recent ones in the computation of the predicted values. Hence, it models systems with fading memory. In particular, the parameter ξ plays the role of a forgetting factor. Proposition 4 validates this intuition by stating that, in order to guarantee the predictor δISS, ξ should be larger than a threshold that depends on m. In particular, it is easy to see that, regardless of the model order, every $\xi \ge \log 2$ is automatically fine for what concerns δISS, although smaller values may be feasible in general.

To summarize, the proposed identification algorithm reads as follows:

1. Select an arbitrary order $m \in \mathbb{N}$ and a viable kernel k_η, e.g., those analyzed in Propositions 3 and 4;
2. Select (β^2,η) by solving the optimization problem (11);
3. The identified predictor is then given by (7), and it is δISS as defined in Definition 1.

5. NUMERICAL RESULTS

In this section, we consider the problem of identifying a discrete-time predictor of the output y generated by the following two deterministic δISS continuous-time systems

$$\mathsf{A}: \begin{cases} \dot x = -x - x^3 + \varsigma(u^2) \\ y = x \end{cases} \qquad \mathsf{B}: \begin{cases} \dot x_1 = -\frac{8}{3}x_1 + \varsigma(x_2)\varsigma(x_3) \\ \dot x_2 = 10(x_3 - x_2) \\ \dot x_3 = -x_3 + u \\ y = 28x_2 - x_1x_2 \end{cases}$$

where u is the input and $\varsigma(\cdot) = \max(\min(\cdot,3),-3)$ is a piecewise linear saturation. System B is analyzed by Angeli (2002). In particular, we consider an experiment $\mathcal{D}_j$ (with sampling time 0.1) generated by system $j \in \{\mathsf{A},\mathsf{B}\}$ obtained by sampling the stochastic variables

$$\bar x_0 \sim \mathcal{N}(0_{d_j\times 1}, I_{d_j}), \tag{14a}$$
$$\bar u_t \sim \mathcal{N}(u(0.1(t-1)), \sigma_u^2), \qquad \forall t \in \{1,\dots,n\}, \tag{14b}$$
$$\bar y_t \sim \mathcal{N}(v_j(0.1(t-1), u, \bar x_0), \sigma_y^2), \quad \forall t \in \{1,\dots,n\}, \tag{14c}$$

where $n = 100$, $\sigma_u = \sigma_y = 0.1$, $d_{\mathsf{A}} = 1$, $d_{\mathsf{B}} = 3$, $u : \mathbb{R} \to \mathbb{R}$ is the input, and $v_j(h, u, x_0)$ is the forced output of system j at time h given the initial condition x_0. Hence, $\bar{u}$ and $\bar{y}$ are noisy measurements of u and y, respectively. The input at time $t \in \mathbb{R}$ is chosen as the random variable

$$u(t) = \sum_{i=1}^{20} A_i \sin(2\pi\nu_i t + \varphi_i), \tag{14d}$$

where, for every $i \in \{1, \dots, 20\}$,

$$A_i \sim \mathcal{U}(0.1, 1), \quad \nu_i \sim \mathcal{U}(0.1, 1), \quad \varphi_i \sim \mathcal{U}(0, 2\pi), \tag{14e}$$

with $\mathcal{U}(a, b)$ the uniform distribution on $[a, b] \subset \mathbb{R}$. The variables $\{A_i, \nu_i, \phi_i\}_{i=1}^{20}$ are all mutually independent. For the identification procedure, we selected $m = 1$ and the Gaussian kernel (12). The optimization problem (11) is solved using $\varepsilon = 10^{-10}$ and $\psi = 0.99$. Given a dataset $\mathcal{D} = \{(\bar{u}_i, \bar{y}_i) | 1 \le i \le n\}$, we define the performance index

$$q(\mathcal{D}) = 1 - \sqrt{\frac{\sum_{t=m+1}^{n} |\bar{y}_t - f_{GP}(\bar{z}_t)|^2}{\sum_{t=m+1}^{n} |\bar{y}_t - \sum_{i=m+1}^{n} \bar{y}_i|^2}}.$$

For each $j \in \{\mathsf{A}, \mathsf{B}\}$, the performance of the identified predictor is assessed on a validation dataset $\mathcal{D}_j^v \neq \mathcal{D}_j$ distributed according to (14) with $n = 5 \cdot 10^4$ and $\sigma_u = \sigma_y = 0$. After selecting the hyperparameters, we employ the procedure presented by Scandella et al. (2021) to reduce the computational complexity of the identified model. The proposed method is analyzed using Monte Carlo simulations with 10^3 runs.

Figures 2 and 3 show the box plots of the performance index $q(\mathcal{D}_j^v)$ for the two cases $j = \mathsf{A}$ and $j = \mathsf{B}$, respectively. Here, we can note that, by enforcing stability using the proposed method, we obtain a slight decrease in performance, i.e. in the capability of the identified predictor to explain future measurements. However, in both cases, the variance of the performance decreases significantly. Therefore, the proposed procedure is more reliable and less susceptible to the specific datasets used. These observations can be explained by the fact that the proposed procedure select (β^2, η) in a smaller set of possible parameters due to the additional constraints of (11). Furthermore, the conditions that we impose are only sufficient, and they can be too restrictive in some cases.

The main advantage of the proposed procedure is that the identified predictor is *guaranteed* to be δISS. As stated in Section 4, the predictor is δISS if its norm squared is smaller than $m^{-1} = 1$. In Figure 4, it is possible to see that this condition is satisfied for all the identified predictors. Furthermore, in Figure 5, we show the output of one of the estimated systems from system B to 25 different inputs that converge to 1 asymptotically starting from random initial conditions. Here, we can see that the outputs always converge to the same value as expected from a δISS system.

6. DISCUSSION & CONCLUSIONS

We introduced a novel identification procedure based on Gaussian Process Regression of nonlinear systems guaranteeing that the estimated predictor is incrementally input-to-state stable. In particular, we devised a sufficient condition on the kernel parameters and the regularization term to enforce such stability guarantees on the identified predictor. We also showed that two commonly used kernels in system identification satisfy the proposed condition and, hence, can be employed for this application.

The main limitation of the proposed approach lies in the restrictiveness of the imposed conditions ensuring δISS (Proposition 1) when $m > 1$. Future research will aim to weaken such conditions to increase the performance of the identified model without sacrificing stability.

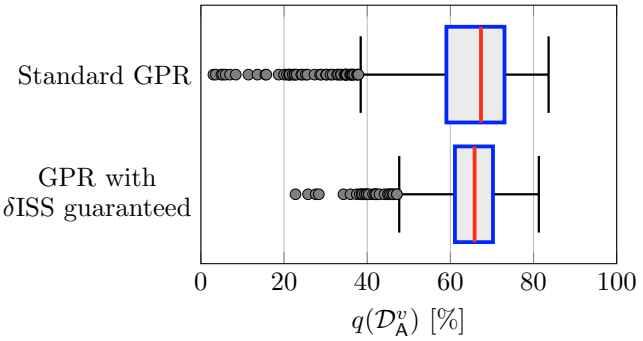

Fig. 2. Box plots of the performance index on the validation dataset for System A.

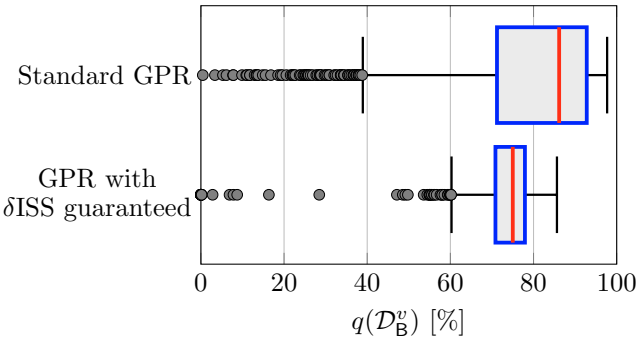

Fig. 3. Box plots of the performance index on the validation dataset for System B.

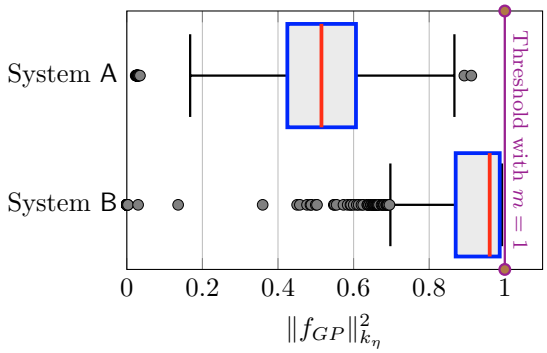

Fig. 4. Box plots of the squared norm of the identified predictor of the two systems.

REFERENCES

Angeli, D. (2002). A Lyapunov approach to incremental stability properties. *IEEE Transactions on Automatic Control*, 47(3), 410–421.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*,

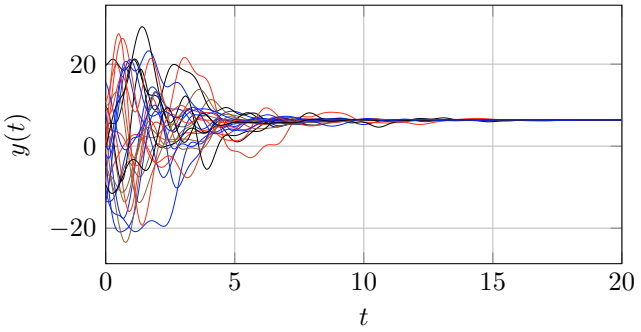

Fig. 5. Output of one of the identified systems using data from system B in response to 25 different inputs that converge to the same value.

68(3), 337–404.

Bai, E.W., Li, K., Zhao, W., and Xu, W. (2014). Kernel based approaches to local nonlinear non-parametric variable selection. *Automatica*, 50(1), 100–113.

Bhujwalla, Y., Laurain, V., and Gilson, M. (2016). An RKHS approach to systematic kernel selection in nonlinear system identification. In *Proceedings of the IEEE 55th Conference on Decision and Control (CDC)*, 3898–3903.

Dinuzzo, F. (2015). Kernels for Linear Time Invariant System Identification. *SIAM Journal on Control and Optimization*, 53(5), 3299–3317.

Formentin, S. and Chiuso, A. (2021). Control-oriented regularization for linear system identification. *Automatica*, 127, 109539.

Lohmiller, W. and Slotine, J.J.E. (1998). On Contraction Analysis for Non-linear Systems. *Automatica*, 34(6), 683–696.

MacKay, D.J.C. (1992). Bayesian Interpolation. *Neural Computation*, 4(3), 415–447.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246–1266.

Mazzoleni, M., Chiuso, A., Scandella, M., Formentin, S., and Previdi, F. (2022). Kernel-based system identification with manifold regularization: A bayesian perspective. *Automatica*, 142, 110419.

Mazzoleni, M., Scandella, M., Formentin, S., and Previdi, F. (2020). Enhanced kernels for nonparametric identification of a class of nonlinear systems. In *Proceedings of the 2020 European Control Conference (ECC)*, 540–545.

Pillonetto, G. (2018). System identification using kernel-based regularization: New insights on stability and consistency issues. *Automatica*, 93, 321–332.

Pillonetto, G. and De Nicolao, G. (2010). A new kernel-based approach for linear system identification. *Automatica*, 46(1), 81–93.

Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., and Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3), 657–682.

Pillonetto, G., Quang, M.H., and Chiuso, A. (2011). A New Kernel-Based Approach for Nonlinear System Identification. *IEEE Transactions on Automatic Control*, 56(12), 2825–2840.

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. MIT Press Ltd.

Rüffer, B.S., van de Wouw, N., and Mueller, M. (2013). Convergent systems vs. incremental stability. *Systems & Control Letters*, 62(3), 277–285.

Scandella, M., Mazzoleni, M., Formentin, S., and Previdi, F. (2021). A Note on the Numerical Solutions of Kernel-Based Learning Problems. *IEEE Transactions on Automatic Control*, 66(2), 940–947.

Scandella, M., Mazzoleni, M., Formentin, S., and Previdi, F. (2022). Kernel-based identification of asymptotically stable continuous-time linear dynamical systems. *International Journal of Control*, 95(6), 1668–1681.

Sepulchre, R., Chaffey, T., and Forni, F. (2022). On the incremental form of dissipativity. In *Proceedings of the 25th International Symposium on Mathematical Theory of Networks and Systems MTNS 2022*, volume 55, 290–294.

van Waarde, H.J. and Sepulchre, R. (2022). Training Lipschitz Continuous Operators Using Reproducing Kernels. In *Proceedings of the 4th Annual Learning for Dynamics and Control Conference*, volume 168 of *Proceedings of Machine Learning Research*, 221–233.